Predicting Regulatory Drug Approvals for the U.S. Pharmaceutical Industry

Gillian Maciver

BA (Hons), MSc

Submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

Adam Smith Business School

College of Social Sciences

University of Glasgow

March 2020

ABSTRACT

This thesis intends to predict outcomes from the drug discovery process. These predictions are then utilised to develop a potentially profitable investment strategy. It explores whether certain patent and firm-level publishing characteristics can prove informative to this predictive process. This study hypothesises and investigates (empirically tests), for the first time, the links between drug approval granted by the U.S. Food and Drug Administration (FDA) and a range of firm-level patent and publishing characteristics. Using hand-collected data for 290 drugs approved by the FDA and their corresponding 632 patents I analyse how certain selected characteristics influence the likelihood of FDA approval. Findings suggest that predicting the probability that drug-related patents result in FDA drug approvals or project failures are enhanced by employing five patent-based and two firm-publication indicators. At the time of patent publication, inventor team size, citations to scientific journal articles, independent patent claims and dependent patent claims all have a positive and significant role in the prediction process. Independent patent claims are especially informative at the time of patent publication, whilst patent citations received are found to be the strongest indicator of approval success, three years after patent publication. Firm publication and university co-authoring also possess a significant, although negative, influence on the likelihood of drug approval.

The model's success in predicting approval-linked patents, correctly predicting 19.75% and 17.82% of these patents for in-sample and out-of-sample analyses, is matched by its success as the basis of a profitable investment strategy. The high levels of correctly predicted approval-linked patents enable the investment portfolios, constructed using the model's predictions, to generate significant positive long run abnormal returns. I find that, contrary to the market efficiency hypothesis, my portfolios are able to generate in-sample approval date buy-and-hold abnormal returns of 60.40%.

With the high levels of information, asymmetry and uncertainty that exists within the drug discovery process these results will prove beneficial to practitioners, academic researchers and investors.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would firstly like to thank my supervisors, Dr. Antonios Siganos and Dr. Mark Aleksanyan, for their support, helpful advice and constructive feedback, which have been invaluable to the writing of this thesis.

I am also thankful to the Accounting and Finance subject group (University of Glasgow) for its support throughout this process.

Finally but most importantly, I would like to thank my family and friends. Especially my husband, Mark, and my children, Eilidh and Alasdair, whose love and encouragement supported me during what has been a challenging time.

Author’s Declaration

“I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.”

Printed Name: Gillian Maciver

Signature:

Chapter 1 Introduction

1.1 Introduction

The aim of the present chapter is to introduce the thesis and provide the context for this research.

Drug[1] discovery and development are complex knowledge-intensive processes, requiring novel ideas and continual innovation to create and maintain corporate success within the pharmaceutical industry. The complexity of the discovery process can make it difficult to understand the key process drivers and to predict which drug candidates will eventually become marketed products.

There are several reasons why understanding this process is important to many stakeholders. Firm-level research managers want to ensure that research efforts are focussed on projects with the greatest likelihood of scientific and commercial success. Understanding why certain research projects result in successful, marketed drug products, whilst others do not, is a dilemma for many managers when allocating scarce resources amongst competing projects. Policy makers also want to ensure that resources are directed efficiently to potentially successful public-private collaborative projects, which can produce potentially life-saving drug products. A better understanding of this process may also reduce the number of failed drug research projects, potentially reducing development timescales and improving access to medicines. It has also been demonstrated that shareholders experience significant increases in wealth from successful drug innovation outcomes (see, for example, Hamill, McIlkenny and Opong, 2013 and Hamill, Hutchison, Nguyen and Mulcahy, 2018). An improved ability to identify potentially successful drug candidates may provide a foundation for a profitable investment strategy.

This study hypothesises that the likelihood of a candidate becoming a marketed drug product is positively associated with certain patent[2] characteristic cited in its regulatory approval[3]

[1] A medicine or other substance which has a physiological effect when introduced into the body.
[2] A patent is a property right granted by the United States Government to an inventor conferring a right or title for a set period, giving the patent holder the sole right to exclude others from making, using, or selling an invention. A patent, in this study, is a utility patent. This is a patent that covers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement (See glossary of terms in Appendix 2 for further detail).
[3] This the receipt of United States Food and Drug Administration (FDA) regulatory drug approval at the end of a formalised review process (See glossary of terms in Appendix 2 for further detail).

application and the publishing behaviour of the firm that holds these patents. Some patent characteristics have been identified as both patent quality and value-indicators in prior empirical studies. Prior studies found that higher numbers of backward patent citations[4] indicate higher quality (see for example Lanjouw and Schankerman, 2001 and 2004). In particular, greater numbers of non-patent references (NPR)[5] were found to indicate a stronger link to prior scientific knowledge and therefore were more innovative in nature (Deng, Lev and Narin, 1999; Harhoff, Scherer and Vopel, 2003; Hall, Jaffe and Trajtenberg, 2005 and Silverberg and Verspagen, 2007). Increased numbers of forward citations[6] (citations received from subsequent patents) indicated a patent with a greater impact on the technological field. These patents were perceived as being more valuable both technologically and economically (Trajtenberg, 1990; Harhoff, Scherer and Vopel, 2003 and Hall et al., 2005). However, this study differs from prior studies by investigating potential patent value-indicators and also the relationship between this early stage of the research and development (R&D) process and the final commercialisation of an invention[7] (receipt of United States Food and Drug Administration (FDA) regulatory drug approval). This relationship is utilised to predict the likelihood of drug approval success. There is still little understanding as to which, and to what extent, patent characteristics convey information regarding an invention's commercialisation potential. An understanding of which characteristics influence the commercialisation phase of a product is essential to interpret the predicted influence of these early stage indicators. This study provides insights into whether selected patent and firm-level publishing characteristics can be utilised to more fully understand the outcomes of a drug candidate's R&D process.

In so doing, these characteristics can then be utilised to improve value estimates for intangible assets, held in the form of patents, and assist in more informed strategic research-

[4] All prior art that is patent-related referenced in a patent application (See glossary of terms in Appendix 2 for further detail).

[5] Non-patent references (NPR) result from the search for prior art, that is the total body of knowledge, which relates directly to an invention. These include references to various non-patent documents, such as scientific articles, technical papers, conference proceedings, etc. The knowledge contained within NPR is not usually as formalised or codified as the knowledge contained within patents (See glossary of terms in Appendix 2 for further detail).

[6] For a particular patent, a forward citation is another (subsequent) patent document's citation back to that particular earlier patent (from the perspective of the earlier patent) (See glossary of terms in Appendix 2 for further detail).

[7] Any art or process (way of doing or making things), machine, manufacture, design, or composition of matter, or any new and useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States.

project selections. This research builds upon previous studies by examining the potential use of scientific and technical knowledge to increase productivity by reducing the number of unsuccessful research projects undertaken by firms.

I examine the role that patents play in the drug discovery process and how they can inform our understanding of this process. The focus of this research is on the variation in drug patent characteristics, and how these may be employed as predictors of future regulatory approval success. I also investigate how these characteristics, along with firm-level publishing behaviour, may be utilised to inform firm-level decisions on asset valuation and research project selection.

One of the key objectives of this study is to investigate whether a relationship exists between a firm's patent quality indicators, firm-level publishing characteristics and the likelihood of a drug product receiving regulatory approval. If such a relationship does exist, can it be predicted to any degree?

I demonstrate that multiple factors influence a successful drug application. Firstly, the innovative process does not just suddenly happen as patents and their characteristics can be influenced by firm-level factors. The success of its innovative process can be determined by how the firm creates in-house knowledge and absorbs knowledge from external sources. This study models the effects on drug development outcomes by using knowledge generated internally and externally, from the date of patent application. It also identifies which factors differentiate potentially successful candidates from those that fail. As discussed further in Section 1.2, previous patenting literature recognised that valuable patents differ, in certain aspects, from those of lesser or no value (Allison, Lemley, Moore and Trunkey, 2004). I highlight how these differences also affect the future commercialisation of the products, to which they are linked, and assess whether certain patent characteristics support the development of a model that can predict future approval-linked patents. Section 1.4 discusses how this thesis extends prior literature by proposing and testing new hypotheses on patent characteristics and their links to an increased likelihood of receiving regulatory approval. In Section 1.5 I discuss how the models developed from testing these hypotheses can be utilised to predict the likelihood of receiving a drug approval. I then demonstrate how the models' predictive abilities can form the basis of a profitable investment strategy.

The remainder of this chapter is organised as follows. Section 1.2 presents the background and rationale for conducting this study. It also contextualises the study by briefly outlining

the extant research on drug approval prediction modelling and patent characteristics together with their effect on firm value and performance. In so doing, I highlight the principal gaps in the literature and illustrate how the current research contributes to the research area. Section 1.3 discusses the main reasons for sampling within United States (U.S.) pharmaceutical firms and Section 1.4 sets out the overall research questions. Section 1.5 provides an overview of the main empirical findings. Section 1.6 discusses the contributions of the study. Finally, Section 1.7 provides the organisation of the thesis and briefly presents the content within each of the chapters.

1.2 Background

Whilst many promising scientific breakthroughs, such as cancer-inhibiting drugs and gene-sequencing therapies offer new hope for patients, these advances have also made the discovery and development process much more complex, creating increasing outcome uncertainty and substantially increasing trial costs. These advances create new drug candidates, each of which requires many years of research and clinical trial[8] testing, costing hundreds of millions of dollars, with a high likelihood that they will fail to become marketed.

The pharmaceutical industry is known for its creation of innovative drug products, but drug discovery efficiency has been declining steadily over the past 40 years, despite scientific and technical progress within the industry. The number of completely new drug products[9] approved by the FDA has declined markedly in recent years. According to the FDA, the average number of such new product approvals was around thirty-five per year between 1999 and 2001, while this fell to about twenty per year during the period 2005 to 2007. This may be due to several factors, including competitive and market conditions. However, work by DiMasi (1995) and Hay, Thomas, Craighead, Economides and Rosenthal (2014) indicated that today's drug candidates are more likely to fail in clinical trials than those in the 1970s. This is despite many years of clinical trials experience and significant improvements in knowledge of disease mechanisms. Researchers have been unable to improve the low clinical approval success rates for new candidates, with the industry's innovative pipeline becoming sparser in recent times (DiMasi, Feldman, Seckler and Wilson,

[8] A controlled experiment involving a defined set of human subjects (See glossary of terms in Appendix 2 for further detail).

[9] A completely new drug product for the purposes of this study is a new molecular entity (NME). It is a drug that contains an active moiety that has never been approved by the FDA or marketed in the U.S. (See glossary of terms in Appendix 2 for further detail).

2010). For example the underlying odds of success for the approval of an oncology drug that enters phase III trials are less than 50% and have not improved in recent decades despite an expansion in the knowledge of the disease mechanisms (DiMasi, Reichert, Feldman and Malins, 2013). Drug development costs nearly doubled every 10 years between 1950 and 2010 (Booth and Zemmel, 2004 and Munos, 2009), with these costs being mostly attributable to failed research projects. The estimated development costs have increased by a factor of ten since the 1970s, where estimated average costs were $114m (DiMasi, Hansen, Grabowski and Lasagna, 1991) but are now estimated to be $1,395 million (DiMasi, Grabowski and Henson, 2016). In the face of this uncertain and costly environment, there is need to evaluate a drug candidate's likelihood of success with more efficiency and to allocate resources accordingly.

This environment has placed greater emphasis on the importance of revenues generated by highly successful commercialised drugs, which need to recoup the costs of these failures. In the period 1997 to 2006, the number of these blockbuster drugs[10] increased from six to fifty-two, comprising half of all 2006 sales and accounting for three-quarters of prescription drug[11] spending growth over the same period. Currently, the industry is experiencing a productivity crisis with the drug discovery process characterised by rising development costs for new drugs, combined with smaller market opportunities. This was firstly demonstrated in 2007, when the number of blockbuster drugs fell from fifty-two to forty-eight (Collier, 2011).

These changes create difficult conditions for the industry with falls in projected revenues potentially affecting innovation in the industry and creating pressure on budgets. Tighter research budgets may therefore inhibit the prospects for improved new drug products reaching the market. In the mature U.S. market, new products experience stronger competition and are targeting smaller market segments, reducing the market potential of any new drug. Firms are therefore under increased pressure to improve the portfolio decision-making process, increase success rates and reduce development costs and risks.

These low productivity levels could be offset by an increase in the financial value per new drug launched. Ideally any successful project resulting in a new drug should provide enough

[10] A blockbuster drug is defined to be a drug than generates at least $1 billion in annual sales.
[11] A drug that can be obtained only by means of a physician's or dentist's prescription.

revenue to cover the investment in failed projects and provide revenues for the company and its investors. Therefore, commercialising more blockbuster drugs could compensate for low commercialisation rates and increasing development costs.

In the initial stages of the drug discovery process developers select a target disease to investigate. The rising costs within the industry may be one of the factors that influence the choice of this target. In developed economies non-communicable, lifestyle-related diseases such as Type 2 Diabetes are more prevalent that communicable diseases such as Tuberculosis (T.B.). The potential market and revenue stream for any new drug that targets a condition such as Diabetes is therefore greater than for T.B. This makes it more likely that firms may choose to focus their research projects in these disease areas. In the sample of FDA approvals I study (see Section 4.2.2), 83% of the conditions these drugs treat are non-communicable. Investors may also consider a drug candidate's target market as this may influence future firm revenues.

As well as target disease type, drug developers typically use general estimates of trial success rates and regulatory approval rates to assist in project decision-making and portfolio management. Prior literature has also tended to use clinical trial related data when attempting to predict progress through the trial process. These studies are published in health-related journals and used a wide array of heterogeneous factors to explain regulatory success. These include clinical trial design, previous trial outcomes, trial characteristics and drug compound characteristics (Kesselheim, Myers and Avorn, 2011 and Putzeist, Heemstra, Garcia et al., 2012). These studies often focus on trials targeted at disease areas such as cancer or focus on orphan drugs[12]. Their focus on prediction has improved the design of clinical trials, reducing failure rates and reducing development times. Recent work in Clinical Pharmacology & Therapeutics by DiMasi, Hermann, Twyman et al. (2015) proposed an approved new drug index algorithm to predict approval for lead indications of oncology drugs after phase II clinical trial[13] testing. In a review of the literature, Liberti, Breckenridge, Hoekman et al. (2017) concluded that no one element or group of elements within the trial process could explain regulatory success. The other weakness of a trial-data based approach is that this approach can only be employed once the trial process is advanced and where

[12] For a drug to qualify for orphan designation it must treat a rare disease or condition.
[13] A phase II clinical trial is carried out on between 100-300 patients to test for the efficacy and potential side effects of a drug candidate (See glossary of terms in Appendix 2 for a more detailed explanation on this trial phase).

resource misallocation may already have occurred. This approach therefore does not inform private investors of any possible financial implications.

This study seeks to assist the project decision-making process within the industry and to develop a model that predicts future regulatory approval. I employ publicly available patent data, produced by the drug discovery process, to predict successful project outcomes. These data have advantages over trial-related data, being publicly available and also available early in the drug discovery process. I select patent data as patents are a measure of novel knowledge and technological capabilities, which have the potential to be developed into marketable (prescription) products.

Patents are often key components in the discovery and development process and can be matched to resulting products with little difficulty in this industry. Following the approach of previous research, I use patents as proxies for the technical knowledge incorporated into the development activity of the firm. Patents exhibit value in two forms; private economic and public value. Private economic value mostly benefits the holder of the patent (the pharmaceutical firm in this study) and not society. This economic value is of relevance to this study. Patent values vary greatly and a patent can often have a private economic value that is greater than its technological significance. Conversely, some inventions can have great technological significance but have little or no economic value. Schankerman and Pakes (1986) demonstrated that patent value distribution is highly skewed, where many patents have little or no value, and very few have any substantive value. This study attempts to identify those patents that are linked to approval success and that therefore possess higher levels of economic value to the holder.

Researchers had always recognised the potential of patent documents as data sources. Silverberg and Verspagen (2007) observed a relationship between the bibliometric and other properties of patents and their economic value. Other researchers also correlated patent-based data with independent technological and economic indicators, for example firm market value (see e.g. Aboody and Lev, 1998 and Hall et al., 2005). Differing patent measures were used to assess the economic value of a patent and the relationship this had with subsequent firm market value (Hall, Jaffe and Trajtenberg, 2000).

Patents can create a commercial advantage, and this can produce increased future revenues for their holders. Prior literature has attempted to determine the value of individual patents or portfolios. This is not without difficulties, but is an important exercise, as many firms

within the U.S. pharmaceutical industry possess substantial intangible assets in the form of patents.

Patent value can be distinct from patent quality, but they may be linked in many cases. According to Allison and Tiller (2003, p.997) "patent quality and value are interwoven in inextricable ways" such that "value can probably be characterised as quality plus other factors". Value and other patent characteristics may be utilised as proxies for quality. Quality may need to be distinguished from value as value reflects characteristics such as the market for the invention. Value may be linked to therapeutic area, where, for example, diseases that are more common have a larger potential market. Quality, like value, varies from patent to patent.

What features make a good quality patent? In the first instance, a good patent must at least satisfy the legal and statutory requirements of patentability[14] where the invention must be new or novel, useful and non-obvious. All issued or granted patents are assessed against these criteria and are therefore valid patents. Good quality patents may also possess additional characteristics including, for example, the concept of value. Value can be relevant to quality and value has often been taken to be a proxy for quality. Drug patents can combine positive social utility benefits with commercial success. Previous literature recognised that valuable patents differ, in certain aspects, from those of lesser or no value (Allison et al., 2004).

Does the knowledge base of a firm influence its approach to research and the output of this research (patents)? This study demonstrates that a firm with a well-developed knowledge resource-base will produce patents of a higher quality. These higher levels of quality also indicate higher potential economic value. Quality patents eventually lead to new products that have the potential to improve a firm's financial performance when marketed. A valuable patent can increase a firm's market valuation (through profit, sales, return on assets etc.).

Many studies focussed on patent citation behaviour, viewing citations as an indicator that the firm was entering a research area providing technological opportunities, and the chance to engage in a new profitable area of research. Nagaoka (2005) found that certain patent

[14] U.S. patent laws require that, for an invention to be patentable, it must be patentable subject matter, i.e., a kind of subject-matter eligible for patent protection, novel, non-obvious and useful (See glossary of terms in Appendix 2 for further detail).

citation types were positively correlated with a firm's Tobin's Q for U.S. patents held by Japanese firms. Carpenter, Cooper and Narin (1980) found patents associated with inventions assessed to be more "important" had a higher likelihood of being cited by subsequent patents. Researchers worked on the premise that more economically valuable patents would be cited more frequently. Hall et al. (2005) found that for U.S. firms, in multiple industries, the addition of one citation was linked to a 3% rise in a firm's market value and that citations below the median level had no impact on value but that citation levels above this median level did have an impact. A firm's technological superiority may lead to increased performance and earnings. Narin, Noma and Perry (1987), Lanjouw and Schankerman (1999), Deng et al. (1999) and Chen and Chang (2010) found this relationship applied to U.S. pharmaceutical firms. A firm's patent citation levels may be a good indicator of its research prowess and of efficient idea-search processes.

This study builds on prior studies that identified a relationship between the characteristics of the patenting output of firms, their market valuation and future performance. Most of this work involves cross-industry studies focussing on differences in firms' performance. These studies were conducted in different economies and industries under various patenting regimes and market conditions (see Hall and Harhoff, 2012 for a review). The results of the patenting-value relationship tested in these studies, although valid, are based on a range of industries and are often based on survey data and inventor[15]-estimates of patent value. Very few studies within this field attempted to predict outcomes based on these measurable patent characteristics. However, one such study was conducted by Sampat and Ziedonis (2004) who attempted to predict which university patents would be licensed to industry together with the revenues earned by licensing. Their results suggested that patent citations were significant and positively related to the likelihood that a patent was licensed, but not related to revenues. In this study, I utilise some of the patent characteristics found by prior studies to be indicators of patent economic value and/or quality to develop models predicting the likelihood of a patent being linked to regulatory approval. I focus on predicting the likelihood of regulatory success and not on the revenues produced by the drugs linked to my predictions. The revenues are ultimately influenced by many factors, such as market competition, disease type and the potential market size.

[15] One who contributes to the conception of an useful invention (See glossary of terms in Appendix 2 for further detail).

Predicting the likelihood of regulatory success is a key goal of this study. I also investigate whether the predictions I obtain can be utilised as a basis for a profitable investment strategy. According to the findings of the multi-industry study of Chaney, Devinney and Winer (1991), new product introductions benefit stock returns, but completely new products have a higher impact. Indeed, the new product literature has consistently related innovation success to the product's ability to provide benefits and features not offered by alternative products (Holak and Lehmann, 1990 and Henard and Szymanski, 2001). I therefore take my predictions and use them to construct stock portfolios that can be profitable investment tools for investors.

1.3 Selection of the United States Pharmaceutical Industry

This study focuses on the U.S. pharmaceutical industry, where patenting behaviour is widespread and patents are viewed as important intellectual assets with firms relying on patent portfolios to gain competitive advantage (Levin, Klevorick, Nelson and Winter, 1987). Patent rights are strong and enforcement levels are high compared to other industries. Correlations between patent characteristics and product commercialisation are more clearly observable as often, relatively few patents protect a product. This enables the linkage between drug candidates and patents to be more clearly identified. Firms have the protected monopoly and exclusivity afforded by patents to reap the commercial rewards for many years[16]. Without the appropriability regime afforded by patents it is doubtful that firms would invest to the levels required to fund development programmes, as they would be unlikely to recoup their costs. Patents therefore play a pivotal role in the innovation process, establishing and supporting a monopoly over a critical area of scientific knowledge. The private economic value of a firm's patent portfolios is critical to the firm's success.

There are several reasons for selecting U.S. pharmaceutical firms as the sample for this study. The U.S. economy has a very large consumer prescription drug market, totalling $325 billion in 2015[17]. It has the world's largest pharmaceutical industry with U.S. firms making

[16] For patent applications filed after June 7, 1995 the patent term commences on the date that the patent issues and ends on the date that is twenty years from the date on which the application was filed in the U.S. or, if, the application contained a specific reference to an earlier filed application, twenty years from the filing date of the earliest of such applications.

[17] https://www.brookings.edu/blog/up-front/2017/04/26/the-hutchins-center-explains-prescription-drug-spending/, last accessed on 8th February 2019.

up eight of the top twenty global firms[18]. The industry leads the world in the development of new drugs, accounting for $1.3 trillion in economic output, representing 4% of total U.S. output and supporting 4.8 million (direct and indirect) jobs in 2015[19]. In 2012, the Pharmaceutical Research and Manufacturers of America (PhRMA) members invested $48.5 billion in research and development (R&D) and the industry has a 20% share of U.S. domestic business R&D (National Science Federation, 2012[20]). This sector's overall economic impact on the U.S. economy is substantial.

This industry is highly regulated due to the nature of the industry's products and their interactions with human subjects. Its products undergo an extensive, compulsory trial-testing regime prior to final submission to the relevant regulatory approval body (FDA) for marketing approval. These trials increase in rigour and a drug candidate must pass each phase of this sequential test regime. Failure at any stage means a candidate cannot obtain approval. This process results in an industry characterised by long lead times between idea development and a final marketed product. The product cycle takes over ten years to take a drug candidate from the laboratory bench to the pharmacy shelf. DiMasi et al. (1991) and Dranove and Meltzer (1994) estimated that it took twelve to fourteen years for a patented drug candidate to obtain marketing approval. All through this process, there is uncertainty as to whether a drug candidate will reach the market and produce any economic return. The attrition rate is high during this process. For every twenty drug candidates that enter the testing phase only one becomes a marketed product.

There are varying cost estimates for development. In 2013, the PhRMA surveyed its members and estimated costs at $1.3 billion to $1.6 billion per marketed drug. All potential products incur costs that vary according to their progress through the testing regime. Very few candidates make it to market and of those that do only a very few recoup their development costs. Grabowski and Vernon (1994) found that only 20% of approved drugs generate enough revenue to cover R&D investment. However, the rewards for a successful

[18] https://www.rankingthebrands.com/The-Brand-Rankings.aspx?rankingID=370, last accessed on 8th February 2019.
[19] http://phrma-docs.phrma.org/files/dmfile/PhRMA_GoBoldly_Economic_Impact.pdf, last accessed on 8th February 2019.
[20] https://www.centerwatch.com/news-online/2012/04/17/phrma-member-firms-invested-495-billion-in-rd-in-2011/, last accessed on 10th February 2019.

drug can be extremely significant as a successful drug for a widely occurring but dangerous disease e.g. hypertension, can earn billions of dollars annually.

The sample period for this study spans the years 1974 to 2014, as the rate of major drug innovations increased from the 1970s (Grabowski and Vernon, 2000). This increase in productivity resulted from pharmaceutical firms starting to implement findings from basic scientific research into their R&D projects. Prior to this, firms applied a trial and error approach to the discovery process.

1.4 Research Questions

Evidence from prior studies, summarised in Chapter 2, suggests that certain patent characteristics are associated with inventions that have more technological significance and potential economic value. Many research studies have shown a positive relation between patent characteristics and the stock market valuation of a firm (see for example, Hall et al., 2005). These studies argue that characteristics in the information recorded on the front page of a patent and citations received from subsequent patents indicate that the patent is of higher quality. This study proposes that certain patent characteristics together with the firm's publishing behaviour influence approval success rates.

I apply four research questions to U.S.-listed pharmaceutical firms. I build multiple hypotheses using the main research questions of this study as their foundations. These hypotheses are fully discussed in Chapter 3 and are tested empirically in Chapter 5.

Research Question 1

Are there differences in patent characteristics between patents linked to commercialised products and those that do not possess such a link?

Research Question 2

Are there differences in publishing behaviour between firms whose patents are more frequently linked to commercialised products and those that are less successful?

Research Question 3

Does the quality of a firm's patents positively influence the likelihood of the commercialisation of its products, in this case prescription drugs?

Research Question 4

Can certain characteristics of a firm's publishing behaviour positively influence the likelihood of the commercialisation of its products, in this case prescription drugs?

Research Question 5

Can stock portfolios based on model predictions relating firm patent and publishing characteristics to drug approval likelihood earn long-run abnormal returns for an investor?

The first question attempts to discover if patent characteristics differ between commercially successful and unsuccessful patents, by comparing the characteristics of approval-linked patents to those with no approval-linkage. The objective is to understand what factors differentiate potential approval-linked patents. The second question is focussed at the firm rather than patent-level. It assumes that differing firm-level publishing behaviour can influence the innovation process and in so doing, affect the output characteristics of this process, patents. In questions three and four, I test whether improved knowledge of the patent characteristics and firm-level publishing characteristics can support the development of models to predict future approval-linked patents. As discussed in Section 1.5, this study directly contributes to the research area by introducing methodological improvements for developing and testing drug approval prediction models. This produces predictions at an earlier stage in the drug development process, which can be utilised by both firm managers and investors.

The fifth research question focuses on approval prediction from an investor's perspective. Here, the objective is to evaluate whether investors can earn long run abnormal returns by investing in firms predicted to hold approval-linked patents.

1.5 Overview of the Main Empirical Findings

My study hypothesises that increases in certain patent-level characteristics, the number of inventors, non-patent prior art[21] citations, independent patent claims, dependent patent claims and citations received, would each increase the likelihood of approval. My analysis shows that all of these characteristics increased the likelihood of approval. Using data available at the patent publication date as the basis, my model indicated that independent claims had the strongest effect on the likelihood of approval, increasing it by 44.5%. Results for the model including citations received indicated that this parameter had the strongest effect, increasing approval likelihood by 54.8%. Independent claims still had a strong effect, increasing approval likelihood by 37.5%. I also hypothesise that shorter citation lag length[22] measures for citations made to other patents, non-patent publications and citations received, would increase the likelihood of approval. The results for the citation lag length measures (both citations made and received) are found to be either insignificant or in a direction contrary to that hypothesised. I also hypothesised that higher levels of firm publishing and co-authoring would increase the likelihood of approval. However, evidence suggests that higher levels of firm publishing and university co-authoring reduce the likelihood of receiving an approval by around 5%, which is also contrary to the hypothesised relationships.

The main motivation of this study is to examine whether certain firm patent and publishing characteristics can be utilised to assist in the predicting of the likelihood that a patent will be involved in the development of an approved drug.

Firstly, the results from the in-sample analysis support the conclusion that the models developed with those variables found to influence the likelihood of approval can indeed predict the likelihood that a patent will be involved in the development of an approved drug. The model based on data available at the patent publication date can predict 40.51% of approval-linked patents. The model including citations received exhibits very similar results, predicting 40.98% of approval-linked patents. These findings were robust to both changes in the firm and patent composition of the sample. Similarly, the model's results from the out-of-sample analysis also support the view that the hypothesised relationship between the

[21] Prior art is all information that has been made available to the public in any form before a given date that might be relevant to a patent's claims of originality (See glossary of terms in Appendix 2 for further detail).
[22] The citation lag length is the time elapsed between the filing of a patent application and the newest patent or non-patent reference cited in the patent document. For citations received this lag length is the time elapsed between the publication of a patent application and the filing of the first patent application that cites it (See glossary of terms in Appendix 2 for further detail).

likelihood of approval and firm patent and publishing characteristics can predict the future approval status of a patent with an ability that cannot be solely attributed to chance. The model based on data available at the patent publication date can predict 22.86% of approval-linked patents. The model including citations received exhibits similar results, predicting 21.43% of approval-linked patents. Again, the results were robust to changes in the size and timing of the selected estimation and test samples. Overall, these results suggest that for the U.S. pharmaceutical industry, a firm's patent and publishing characteristics can be employed to assist in the prediction of a patent's eventual drug approval-status.

Based on this encouraging model for predictive performance, I proceeded to test whether investors can utilise my publication date model to earn abnormal stock returns. The results demonstrate that an investment in the firms predicted to hold approval-linked patents results in significant long run abnormal returns being earned by an investor. The analysis shows that this model can generate substantial abnormal returns in certain periods. For example, the model generates significant positive buy-and-hold abnormal returns (BHARs) for patents (published during the period 1962 to 2014) predicted to be linked to an approved drug. For these patents, the model can generate BHARs of 30.90% for the publication to approval date holding period, after excluding certain confounding events (earnings announcements and mergers and acquisitions announcements) from portfolio return calculations. For the 1997 to 2014 out-of-sample portfolio, approval date BHARs are 52.20% after excluding confounding events from return calculations. These results suggest that the presence of highly confounding events within the predicted approval portfolios are not the sole drivers of positive and significant portfolio returns.

Overall, I find that approval prediction portfolios generated using the publication date model outperform the market.

1.6 Contribution

My review of the literature reveals a research gap in the theoretical and empirical studies of the link between patent-related measures and the commercialisation of drug candidates. To the best of my knowledge, this study is unique in utilising patent data and data on firm publishing behaviour to predict drug approval outcomes. Most prior studies on drug approval prediction have a pharmacological focus and are published in health-related journals. They utilise trial-related data, have very small sample sizes, are concentrated on specific therapeutic areas and involve a small or limited number of predictive factors. Recent

research by Wagner and Wakeman (2014) investigated drug product market outcomes using European patent-based measures. Their study focussed on drug candidates that had already entered pre-clinical[23] and clinical trials, so any predictions made from their data are based on data for patents that have all entered the pre-clinical or clinical trial phases of the drug discovery process.

In this study, I propose the use of a broader range of both patent-level and firm-level publishing characteristics to predict regulatory approval success rates. My sample is also broader in nature as it includes patents that have not been linked to clinical success. This unique dataset allows me to investigate how patent-level and firm-level publishing characteristics are related to the outcomes of the drug development process. These predictive models can then be employed to make data-driven decisions at a much earlier stage of the discovery process. I hypothesise that certain patent characteristics and firm-level publishing behaviour contain useful information concerning drug development outcomes that allows me to forecast the outcome of drug development projects with increased accuracy. My aim is to develop predictive models to assess the probability of success for drug candidates. This reduces the risk and increases the efficiency of the drug development process, providing the project decision-making process with more data-based selection tools. The models also provide benefits for individuals and investors outwith the pharmaceutical industry, as the predictive information they can provide can be used to construct profitable stock portfolios.

R&D managers often prefer novel projects, so project selection is not always based on objective estimates of a project's costs and returns (Criscuolo, Dahlander, Grohsjean and Salter, 2017). These predictive models can assist decision makers to overcome these biases in their decision-making. The issue for drug developers is that no such models exist for assessing the probability of success for regulatory marketing approval early in the drug discovery process. This prototype tool would have the following advantages over intuitive decision-making: it assists in eliminating bias in decision-making; weights predictive factors based on underlying probabilities, hastens the rejection of and lessens the work needed for drug candidates with a low predicted probability and allows for comparison to norms when only one project is under consideration.

[23] Safety and efficacy tests undertaken in computational models, cells and animals (See glossary of terms in Appendix 2 for further detail).

The intention is not to use these predictive models in a robotic fashion to make go/no-go decisions for drugs in development. Rather that they be considered along with traditional trial success measurements and considerations specific to the drug candidate. Using assigned success probabilities might not preclude failure and, nor is it meant to limit risk taking. Instead, assessing probability of success using a model-based method may assist with resource allocation. This ensures that resources are provided to promising candidates and to limit ill-conceived risk. The economic impact from avoiding costs for failed trials can be substantial. For example, a recent analysis of biopharmaceutical R&D costs found that nearly halving the clinical approval success rate over the last decade or so, in isolation, accounted for more than a 50% increase in R&D costs (Scannell, Blanckley, Boldon and Warrington, 2012). This approach demonstrates that, despite there being a low overall probability of regulatory success, patents possessing certain identified characteristics exceed the average success rate for the industry. Energy and resources should be concentrated on these drug candidates.

To my knowledge, this is the first study to use predictive modelling of drug approval likelihood to inform an investment strategy. I utilise the model predictions to select firm stocks to include in an investment portfolio and evaluate the model's ability to generate long run abnormal returns for investors. The model's performance is evaluated using different estimation windows and portfolio identification methods. The model generates abnormal returns in all selected time periods and for alternative portfolio compositions. Contrary to the efficient market hypothesis, the long run return generated by the investment portfolio derived from my chosen model is statistically different from zero.

1.7 Thesis Structure

The thesis has eight chapters and is organised as follows.

Chapter 2 provides a literature review that firstly examines theory related to the role of knowledge and firm-level absorptive capacity and the part they play in the innovative process. This is followed by a section reviewing prior empirical research on patent characteristics and their use as patent-quality indicators and predictors of firm-level performance and valuation.

The development of hypotheses is the focus of Chapter 3, in which the assumptions and rationales for the formation of the hypotheses are presented.

Chapter 4 discusses the methodology adopted in this study and explains how the sample firms are selected. Specifically, it presents the selection rationale, definitions and computations of the dependent variables, explanatory variables and control variables tested in the empirical models in this study, with discussion of the reasoning underlying their selection. The chapter also outlines the data collection procedures employed.

Chapter 5 presents the descriptive statistics and the results of univariate analyses of the data. It reports the empirical test results of the multivariate regression models and discusses the interpretation of the results.

The model's predictive ability is evaluated in Chapter 6.

The ability of the model as the basis of an investment strategy is evaluated in Chapter 7. Further analysis and robustness tests are conducted in Chapter 7 with a view to explaining some of the results obtained.

Chapter 8, the concluding chapter, provides a summary of the study and a discussion of the empirical findings. This chapter highlights any limitations of this study and offers suggestions for further studies in this research area.

Chapter 2 Literature Review

2.1 Introduction

This chapter presents an in-depth review of previous studies, providing background information and developing a link to the present study. The sections of this chapter outline the theoretical underpinnings that inform this research and have motivated the selection of variables. I examine and review the current state of the literature in these theoretical areas. I discuss the conclusions of this review and outline the contributions that this research can make. This chapter is structured as follows; Section 2.2 outlines the context and motivation of the thesis. Section 2.3 discusses the role played by patenting, and Section 2.4 discusses the literature related to the concept of absorptive capacity and its role in firm innovation. Section 2.5 contains a review of the empirical literature on firm patenting behaviour. Section 2.6 discusses the effects of drug approval decisions on shareholder wealth. Section 2.7 concludes the chapter.

2.2 The Context

Over the last few decades, the developed economies of the world (subsequently referred to as the economy) have shifted from tangible assets to intangible assets being viewed as the key productive assets. This is characterised by rapid technological change and the role of innovation becoming increasingly important. Carneiro (2000) viewed this as a structural change in the productive paradigm. Technology is increasingly embodied in the intangible assets of organisations, e.g. intellectual property (IP), and to a lesser extent in fixed tangible assets. The industrial economy is evolving into a knowledge economy driven by information and knowledge (see Smith, 2002). In the era of the knowledge economy, patents, proprietary technology, trademarks, goodwill, and other innovation-related intangible assets play a central role.

Physical capital is no longer the most important asset for firm success (see, for example, ARM Holdings). Many scholars now view the economy as one of intangibles, where these assets are the determinants and a key source of firms' economic success. One of the key intangible assets is knowledge, with Wenger and Synder (2000) viewing knowledge as the main output of the economy. Romer and Kurtzman (2004) recognised that competition in a knowledge economy is based on innovation and that innovation has emerged as the principal determinant of firm competitiveness. The organisational knowledge base differentiates firms.

The OECD (1999) confirmed that the role of knowledge has assumed a greater importance within the economy. Lev (2003) recognised that intangible assets often have a greater importance to firms than tangible assets, both in terms of value and contribution to growth. Intangible assets may represent the competitive advantage of a firm but often firms do not fully understand their nature or value. It is therefore important to value a firm's stock of intangible assets to gain a clearer understanding of the firm's true value. As the focus of the economy has increasingly turned to intangible assets, the role of firms has also had to evolve and the knowledge-intensive firm has taken on an increasingly important role. According to Alvesson (1995), the knowledge-intensive firm's main activity is the employment of knowledge to produce new products. A firm's survival now depends on its ability to synthesise knowledge and develop it into new innovative products. According to Lim (2009), inventive, knowledge-intensive firms combine internal knowledge sources (created by in-house R&D expenditures) with acquired external knowledge. These firms tend to compete on their ability to solve problems and provide solutions. They promote innovation and creativity and have a more intense knowledge-creation process than traditional firms that relied more on physical capital. The rate of new product introduction can be viewed as a function of a firm's ability to create, maintain and manage knowledge. The knowledge-creation process of these firms tends to be more exploratory in nature with a stronger impact on innovation, often producing radical innovations. Workers within these firms need to be willing to create and share knowledge (in the form of publications or patents) to create and sustain a competitive advantage and improved firm-level performance in the longer-term.

One of the key sources of knowledge-intensive firms in the economy is the pharmaceutical industry. This industry is widely accepted as being composed of knowledge-based firms with human assets being central to their achievements (Hung, Huang, Lin and Tsai, 2005).

Drug discovery and development is a complex knowledge-intensive process. Novelty and continual innovation are essential for corporate success in this industry and a large section of the industry's workforce is engaged in intellectual activities. Knowledge creation is viewed as more important for the firms developing new prescription drug products (subsequently referred to as drugs) with knowledge utilisation more relevant to firms producing generic drugs. Firms have a critical need to develop new knowledge in order to produce successful products in the marketplace. Knowledge creation results in longer-term economic benefits, as new ideas become new drugs and earn revenues. Economic success in

this industry is dependent on how firms develop and apply knowledge to produce new products.

2.3 The Industrial Setting: The U.S. Pharmaceutical Industry

The pharmaceutical industry is known for its knowledge-intensive focus and the creation of innovative products. This thesis focuses on the U.S. pharmaceutical industry as it leads the world in the development of new drugs (Batelle Technology Partnership Practice, July 2013[24]). In 2012, the Pharmaceutical Research and Manufacturers of America (PhRMA) members invested $48.5 billion in research and development (R&D) and the industry has around a 20% share of U.S. domestic business R&D (National Science Federation, 2012[25]). The overall economic impact of this sector on the U.S. economy is substantial. The industry accounted for $789 billion in economic output, representing 2.9% of total U.S. output and supported 3.4 million (direct and indirect) jobs in 2011 (Batelle Technology Partnership Practice, July 2013).

The industry is highly regulated due to the nature of its products and their interactions with human subjects. Products undergo an extensive, compulsory testing regime prior to final submission to the relevant regulatory approval body for marketing approval. For the purposes of this research, this is the U.S. Food and Drug Administration (FDA). The FDA regulatory approval process can be classified into "micro-stages" (McNamara and Baden-Fuller, 2007). Stage 1 is the award of a patent after a new molecular entity (NME)[26] is discovered. Pre-clinical trials are the second stage and involve laboratory and animal studies. If pre-clinical trials are successful, the developing firm progresses to the third stage, which incorporates three phases of human clinical trials. Phase I trials assess the safety on humans, phase II evaluates effectiveness at various doses for a larger group of subjects and the drug is administered to a large number of patients who meet the selection criteria for phase III. These trials increase in rigour and must be completed successfully prior to market approval. A drug candidate must pass each phase of this testing regime prior to progressing to the next. Failure at any stage means a drug cannot obtain marketing approval. Successful phase III trials open up the possibility of progressing to phase IV, which is the submission of a new

[24] http://phrma-docs.phrma.org/sites/default/files/pdf/The-Economic-Impact-of-the-US-Biopharmaceutical-Industry.pdf, last accessed on 16th September 2013.
[25] https://www.centerwatch.com/news-online/2012/04/17/phrma-member-firms-invested-495-billion-in-rd-in-2011/, last accessed on 10th February 2019.
[26] An NME is a drug that contains an active moiety that has never been approved by the FDA or marketed in the U.S. in any form (See glossary of terms in Appendix 2 for further detail).

drug application (NDA)[27] to the FDA. Approval of an application grants marketing rights for the drug to be sold legally. Figure 2-1 shows the timeline of the approval process and the probability of a drug candidate progressing from a lab-based idea to a marketed product (See Appendix 1 for a more detailed explanation of the drug discovery and development process). The focus of my research is the time period between the filing of a patent application with the United States Patent and Trademark Office (USPTO) and the drug approval decision by the FDA.

It is argued in the literature that each stage of this process represents definable progress for the R&D project, with patenting and pre-clinical trials categorised as basic exploration research. Human clinical trials and NDA submission are classified as the exploitation stage of drug discovery (see, for example, DiMasi, 1995; DiMasi, Hansen and Grabowski, 2003; Rothaermel and Deeds, 2004 and McNamara and Baden-Fuller, 2007). This process results in an industry characterised by long lead times between idea development and a product launched to market. The very long product cycle can take over ten years to take a drug candidate from the laboratory bench to the pharmacy shelf. DiMasi et al. (1991) and Dranove and Meltzer (1994) estimated that marketing approval took twelve to fourteen years for a patented drug candidate. There is uncertainty throughout this process as to whether a drug candidate will reach the market. There is a high attrition rate, so that for every twenty candidates that enter the pre-clinical testing phase only one becomes a marketed drug. In comparison, many other high-technology industries have relatively short product development cycles, e.g. consumer electronics.

The highly regulated environment also creates a unique competitive environment for pharmaceutical firms. In the mature U.S. market, new products also experience stronger competition and target smaller market segments, reducing the market potential of any new drug. To compete and remain profitable, firms need to improve the efficiency and productivity of their R&D.

Ideally any successful project that results in a new drug should provide enough revenue to cover the investment in that project together with failed projects, and also provide revenues

[27] For drugs completing the clinical trial testing process, a submission for marketing authorisations is made to the U.S. Food and Drug Administration (See glossary of terms in Appendix 2 for further detail).

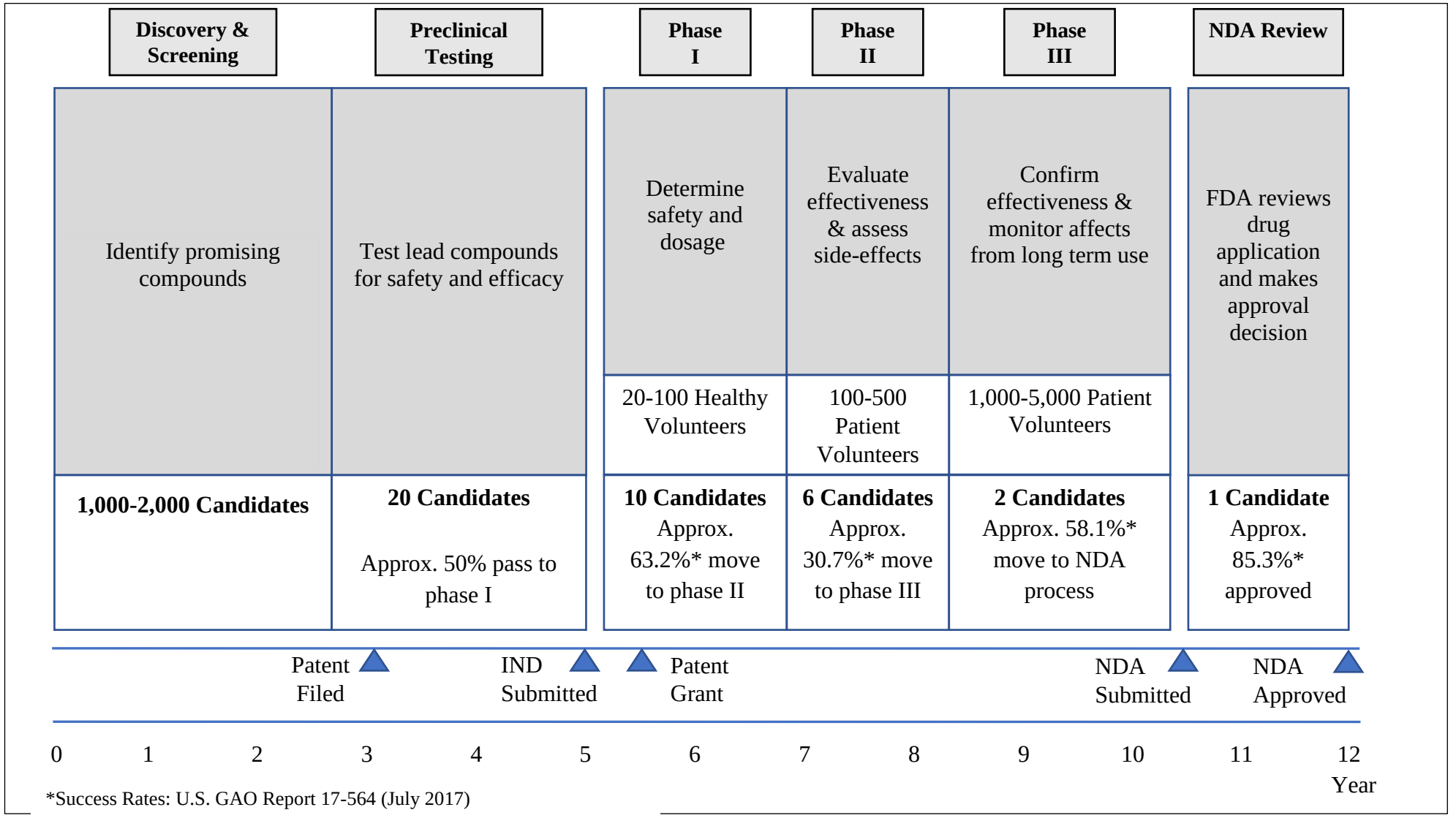
for the company and its investors. Therefore, commercialising more profitable drugs could compensate for low commercialisation rates and increasing development costs.

In the initial stages of the drug discovery process developers select a target disease to investigate (See Figure 2-1). The rising industry cost-base may be one of the factors that influences the choice of this target. In developed economies non-communicable, lifestyle-related diseases such as Type 2 Diabetes are more prevalent that communicable diseases such as Tuberculosis (T.B.). Therefore, the potential market and revenue stream for any new drug that targets a condition such as Diabetes is greater. In addition, they need to formulate a commercialisation strategy that helps them transform their R&D results into a marketable product. Measuring innovative performance in the pharmaceutical industry is different from other industries for multiple reasons. First, the time between the commencement of a research project to the introduction of a drug candidate on to the market is very long, as detailed above. Second, unlike industries where direct consumer demand can be used to measure the success of an innovation, the demand and consumption of drugs is decided by doctors and not by the direct consumers, the patients.

Another unique industry characteristic, created by the high level of regulatory oversight, is the extensive testing process required to gain approval means that pharmaceutical firms consequently have a very high cost base. The costs for creating and developing drugs are high but product manufacture is relatively simple, and costs are relatively low compared to the development phase. There are varying estimates of these development costs, but in 2013, the PhRMA surveyed its members and estimated costs at $1.3 billion to $1.6 billion per marketed drug. However, the rewards can be extremely significant as a successful drug for a widely occurring but dangerous disease, e.g. hypertension, can earn billions of dollars every year through worldwide sales.

The long lead times result in firms with high levels of intangible assets. Firm value is therefore very difficult to measure, especially due to the high levels of uncertainty that exist over the outcome of the regulatory testing process. There is also uncertainty surrounding the future profit flows resulting from the marketing of an approved drug. Firms would therefore benefit from improving their understanding of the future value of their intangible assets. To do this it is necessary to identify the factors that influence successful innovation and to develop indicators to identify these at an earlier stage of a drug's development. These factors can then be utilised as indicators to identify firm-level intangible assets that are of higher

Figure 2-1 Drug Discovery and Development Process



value to the firm and to utilise them to predict the outcomes. This research could therefore assist firms with resource or asset valuation and could be a valuable input into the firm strategic decision-making process.

This industry not only has an impact economically, but it also has social and ethical impacts by its nature. The non-economic impact of this industry is unique in the non-service sector of the economy. The prices for the output of this industry are often high and their procurement is a notable percentage of the U.S. and other countries' gross domestic product (GDP). Developing techniques to assist in the prediction of regulatory success, may assist in improving the efficiency of resource allocation. This in turn would benefit both the industry and the wider society.

2.3.1 The Role of Patenting within the Pharmaceutical Industry

Patenting is one route to appropriate the returns on innovation. For the purposes of this research, appropriability relates to the ability of an inventor or group of inventors to capture the economic value and other benefits that are created by introducing their invention. In the pharmaceutical industry, patenting is extensively utilised to protect the firm's IP. These industrial patents are also enforced with instances of infringement being litigated or settled out of court. Levin et al. (1987) surveyed 650 research managers and found that for drug products, many of these managers viewed patents as the most effective means of appropriability. Patents are effective IP-protection mechanisms in the pharmaceutical industry (see also Mansfield, 1981; Gabrowski and Vernon, 1990; Cohen, Nelson and Walsh, 2000 and Kesan, 2009) and propensity to patent is much higher than for other industries (Arundal and Kabla, 1998). A common definition of the "propensity to patent" is the number of patents per R&D dollar (see, for example, Griliches, 1990, p. 1678 and Hall and Ziedonis, 2001, p. 102). This ratio, however, is sometimes referred to as "research productivity" (see, for example, Lanjouw and Schankerman, 2004, p. 441). This high propensity means that the economic benefits are highly appropriable for gains in the form of future revenues or cash flows from products based on these patents. These gains can in turn fund further drug discovery and development. The industry's IP environment is unique as patents protect products extensively. Firms have the protected monopoly and exclusivity afforded by patents to reap the reward from a commercially successful drug for many years. Without the appropriability regime afforded by patents it is doubtful that firms would invest to the levels required to fund drug development programmes, as they would be unlikely to recoup their development costs. Patents play a pivotal role in the innovation process within

the pharmaceutical industry and any explanation of this process would be incomplete without the inclusion of the role patents play. For the purposes of this research, a patent is defined as a property right over a knowledge asset. It is a set of exclusive rights to a form of intangible asset granted by, in this case, the United States Government to an inventor or their assignee[28] via the USPTO. These rights are granted for a limited time and are given in exchange for the public disclosure of the invention. A patentable idea must be novel, non-obvious and useful.

A patent is a form of IP that, in legal terms is the creation of the mind for which legal rights are recognised. It is a legal document that defines the boundaries of the technology it covers. An inventor has the sole right to make, use and sell their invention for a limited period. This period has been 20 years from the date of the filing of the application since 8 June 1995.

Many patents issued have no real technological influence and have very little private economic value to the firm that holds them. It is not the remit of the issuing patent office to assess the commercial characteristics of a patent, its role is only to assess if the invention meets the necessary patent requirements new, novel etc. (see Chapter 1, Section 1.2). A granted patent is not an indicator of economic or technological value. The value of patents is found in their ability to exclude others from appropriating the economic benefits flowing from the underlying invention. Here, an invention is a brilliant idea that occurs in the research discovery phase of a project. Innovation gets that brilliant idea to the market through the process of developing inventions for use. However, an invention may have little or no market value and so its private economic value will be marginal. This private economic value depends on the patent's economic value and the relative strength of the existing patent protection. The quality of a patent is not the same as its ex-post value, as a patent can be technologically valuable without earning high economic returns (Lanjouw and Schankerman, 1999). Obtaining a downstream indicator of a patent's private economic value is useful to the firm, especially for a firm's strategy.

Firms operating in a knowledge-intensive sector, with high levels of R&D and intangible assets are often more concerned about appropriability conditions, which may be essential for the success of their R&D projects. In the pharmaceutical industry, patents are granted many

[28] The entity that is the recipient of a transfer of a patent application from its owner of record (See glossary of terms in Appendix 2 for further detail).

years prior to a marketable product. The rationale for patenting in the pharmaceutical industry is to stop imitation and to allow a firm to earn monopolistic profits that will allow it to recoup its development costs.

The benefit from patent protection has a cost and that cost is public disclosure. The patent system is based on a contract between the patent holder (inventor) and society. Inventors receive a possibly valuable exclusion right for the lifetime of the patent and society receives information about the invention that may be built upon by other individuals. Patents induce follow-on investment that enables commercialisation of the invention. This view assumes that the patent is awarded at an early stage of the product cycle. The patent ensures that the economic rewards from further effort at the development stage can be appropriated when this stage proves successful. The publication of a patent disseminates knowledge that would otherwise remain secret. This benefit partially mitigates the monopoly power that can be derived from patenting an invention.

Patents are a method of rewarding a firm's innovation process and they may encourage technological change and economic growth. Without them, there can be difficulties appropriating the returns from the creation of knowledge. Assuming that firms are profit seeking they will seek to recoup the benefits from their innovative efforts by pursuing patent protection for their inventions.

2.3.2 Research Motivation

The drug discovery and development process is extremely complex in nature and the main drivers must be identified to create a useful explanation of how this process functions. Because of this complexity, an explanation influenced by a single theory would be inadequate, so multiple theories are used to construct a robust explanation. These theories all have the common attribute of knowledge as a central focus, as knowledge creation and application are the main innovation drivers in the pharmaceutical industry. To understand this process, one must first understand the importance of knowledge, its characteristics and the interactions it has with other firm-level factors.

Firms patent their inventions to protect them from competitors. I examine the role patents play in the drug discovery process and how they can inform our understanding of this process. I do not investigate the reasons why patenting regimes exist within the pharmaceutical industry or why the propensity to patent may vary amongst firms. The

technological quality of the inventive idea that creates a new drug is of key importance to this research. This quality can be assessed by utilising some of the characteristics within the patent document. The focus of this research is therefore on the variation in the characteristics of firm-level patenting and how this may be an indicator of intangible asset value. I also investigate how these characteristics may inform firm-level decisions on potential asset valuation and the selection criteria for research projects.

This research demonstrates that many factors influence a successful drug application. Firstly, the innovative process does not just happen suddenly. Innovation, patents and their characteristics are influenced by firm-level factors. The firm must be able to create new innovative knowledge that can become patentable material. It must also develop in-house R&D and be able to recognise when this is not sufficient to enable a significant inventive step. A firm's process for creating in-house knowledge and absorbing external knowledge can determine the success of its innovative process.

2.4 The Role of Knowledge in the Innovation Process

This thesis seeks to determine which patent-level and firm-level characteristics enhance a firm's ability to create new drug products that obtain regulatory approval. By analysing these characteristics, I will identify whether the factors that drive this success are internal to the firm or whether external factors play a key role.

This section reviews the literature addressing the determinants of firm-level competitive advantage and their role in the innovation process.

The resource-based view of the firm (RBV) proposes the idea that a firm's competitive advantage flows from a set of unique firm-level resources (Nelson and Winter, 1982). For an organisational competence to be a source of competitive advantage it must be heterogeneously distributed within an industry, be unavailable via market transactions at less than its true marginal value and must be difficult or costly to replicate (Wernerfelt, 1984; Barney, 1986 and Peteraf, 1993). Barney (1991) proposed that intangible resources are more likely to generate a competitive advantage as they tend to be rare, valuable and are difficult to imitate or substitute. Several authors have suggested that a firm's unique capabilities in

R&D are important sources of competitive competence (Dierickx and Cool, 1989 and Nelson, 1991)[29].

The RBV of the firm focussed on explaining why firms' performance in the same industry varies over time. However, very few papers attempted to test the core premises of the RBV using empirical methods (Hoopes, Madsen and Walker, 2003). Authors also questioned the theory's treatment of knowledge resources as RBV sees knowledge as a generic resource[30] that can provide a competitive advantage if combined with other firm-level resources. Whilst the theory focussed on the importance of firm-level resources, it ignored how firms use their resources and interact with their external environment.

Knowledge resources are particularly important to ensure that competitive advantages are sustainable and this is especially true for the pharmaceutical industry. As these resources are difficult to imitate, they provide a foundation for sustainable differentiation between firms (Wiklund and Shepherd, 2003). The RBV literature recognised that knowledge underlies the capabilities that sustain a firm's competitive advantage. The existence of performance differences between firms was attributed to knowledge asymmetries (capabilities and competences) and knowledge gained increasing attention as an important source of competitive advantage (see, for example, work by Kogut & Zander, 1992; Amit & Schoemaker, 1993; Peteraf, 1993; Von Krogh and Roos, 1996 and Spender, 1996). The firm's resource base was acknowledged to be comprised increasingly of knowledge-based resources (Roos, Roos, Dragonetti and Edvinsson, 1997; Stewart, 1998; Sveiby, 2001 and Marr and Spender, 2004).

This viewpoint was further developed within the knowledge-based view (KBV)[31] of the firm, which considers knowledge to be the most important strategic resource (Grant, 1996). Proponents of the KBV argue that knowledge-based resources are hard to imitate, are socially complex, immobile and heterogeneous and thus are major determinants of sustained

[29] Several studies have confirmed that there are significant and persistent differences across firms in their ability to conduct research and to develop new products (see, for example, work by Clark and Fujimoto, 1991; Leonard-Barton, 1992; Henderson, 1993 and Eisenhardt and Tabrizi, 1995).
[30] Knowledge and other intangible resource measures utilised were often unrefined, e.g. R&D intensity or patent counts.
[31] The KBV of the firm is viewed as an extension of the RBV. (Roos, 1998; Hoskisson, Hitt, Wan and Yiu, 1999; Sveiby, 2001; Bontis, 2002; De Carolis, 2002; Huizing and Bouman, 2002 and Balogun and Jenkins, 2003).

competitive advantage (Spender, 1996). They recognised that, as innovation is a knowledge-based process then an innovative firm's performance is based on its knowledge assets.

Kogut and Zander (1992) viewed the role of the firm as creating, transferring and transforming knowledge into a competitive advantage. Innovation is a function of the firm's ability to manage, maintain, and create knowledge (Cohen and Levinthal, 1990 and Smith, Collins, and Clark, 2005). It connects previously unconnected ideas and knowledge or recombines previously connected ideas and knowledge in novel ways (Kogut and Zander, 1992 and Nahapiet and Ghoshal, 1998). According to Nonaka (1994), knowledge is the only true, lasting source of competitive advantage.

Knowledge is unlike other resources as is not scarce, it does not depreciate, and it can be shared without diminishing. Expanded through its application, it can generate increasing returns through its systematic use (Kim and Mauborgne, 1999) and can be used simultaneously in several locations (Wilcox King and Zeithaml, 2003). Knowledge has two forms, tacit knowledge (knowing how) and explicit knowledge (knowing about facts and theories): tacit knowledge is the knowledge of an individual, which is unspoken and not codified, e.g. a researcher's own knowledge and experience. Explicit knowledge is written down, is accessible to others and codified, e.g. patents. Knowledge needs to be documented and codified to protect the competitive advantage it creates.

An important issue in the KBV is the transfer of knowledge and the difficulty involved in its transfer (Nonaka, 1994 and Szulanski, 1996). The KBV literature also focuses on the different categories of knowledge that exist (Grant, 1996 and Nonaka, 1994). For knowledge to be absorbed by the firm, it often requires internal firm processes to convert tacit knowledge into explicit knowledge. One path by which tacit knowledge becomes explicit knowledge is through the patenting process and the creation of new drug products.

Publishing scientific papers provides evidence of the quality of a firm's knowledge base. Scientific publications are peer-reviewed and so act as an independent indicator of knowledge quality. A firm's knowledge base facilitates the production of new products and processes or the improvement of existing ones making them more efficient or effective (Nonaka and Takeuchi, 1995). According to Katila and Chen (2008), a firm needs to develop exploratory, exploitative knowledge and time its development relative to its rivals. Based on previous theoretical and empirical studies I hypothesise that firms increase and improve their knowledge base by allowing their researchers to publish scientific papers, co-author papers

with universities or other firms and by organising their researchers into larger research teams. In so doing they improve their research performance and increase the likelihood that their research output will result in an approved product.

2.4.1 Internal and External Knowledge

Amit and Schoemaker (1993) argue that the value of a resource to a firm does not reside solely in the resource itself but in its utilisation. Many firms may have access to the same knowledge resources, but it is a firm's capability in using these resources that allows differentiation in firm-level innovative performance. According to Kogut and Zander (1992) the combinative capabilities of a firm allow it to create new applications from existing knowledge and are the source of continuous innovation.

I believe the variation in firm regulatory approval success within the pharmaceutical industry is determined by a firm's use and combination of knowledge from differing sources. I am interested in empirically analysing whether differing firm-level knowledge-based capabilities are related to commercialisation success.

The knowledge available to a firm can either be generated internally by its own research teams or generated externally by researchers outwith the firm. The internal knowledge of a firm is its own in-house technological knowledge. R&D is a measure of a firm's knowledge creating activity and allows firms to more effectively appraise, assimilate and apply new external knowledge. It can therefore be used as an indicator (proxy) for knowledge inputs and as an innovation ability indicator, as it highlights the innovation ability of the firm. The tacit, specific and complex knowledge developed internally can generate a long-lasting competitive advantage, as this knowledge is difficult to imitate in the short term (McEvily and Chakravarthy, 2002). Gambardella (1992) concluded that U.S. drug manufacturing firms with more successful in-house scientific research programmes exploit outside scientific information more efficiently, improving research productivity. Barriers to imitation, e.g. patents, also create differences in firm knowledge bases that lead to a competitive advantage for the firm. If a firm has difficulty acquiring external knowledge (possibly due to its tacit nature) then it has no option but to create knowledge internally.

Based on the KBV of the firm and a dynamic capabilities perspective, Teece, Pisano, and Shuen (1997) investigated the dependence of firms' product innovation on the set of practices through which firms identify, gain access to, assimilate and utilise external and

internal knowledge. These firm-level dynamic capabilities belong to a firm's fundamental learning processes. In this thesis, I focus mainly on the knowledge-based and dynamic capability views as they stress elements such as the role of absorptive capacity (knowledge exploration, transformation and exploitation), and the supporting role of other firm-level capabilities in developing absorptive capacity (Volberda, Foss and Lyles, 2010).

The external knowledge sources of a firm are the knowledge it acquires and develops from interactions with other organisations. Knowledge can also spill over from external sources due to its free good characteristics. The use of external knowledge sources broadens a firm's knowledge base and keeps the firm up to date with current developments (Grant, 1996). Fabrizio (2009) argued that acquiring external knowledge can produce a richer knowledge base for a firm, allowing the firm greater flexibility and greater adaptability especially in a dynamic environment. Knowledge sharing can also have a positive effect on new product development (Hong, Doll, Nahm and Li, 2004). The level of external knowledge used by a firm can be measured by using the average level of patent citations to scientific literature or to patents held outwith the firm. In their study of major U.S. pharmaceutical firms, Narin et al. (1987) found that when a firm utilises external scientific discoveries there is an increase in citations to NPR in patent documents. Cockburn and Henderson (1998) showed that the ability to maintain closer ties to the scientific community is a key factor driving a firm's ability to recognise and utilise upstream research findings.

Although important, external sources of technological knowledge are not a substitute for internal R&D. Internal R&D is needed because, "It frequently requires a substantial research capability to understand, interpret and to appraise knowledge that has been placed upon the shelf." (Rosenberg, 1990, p. 171). In this context, Rosenberg argued that internal basic research is essential for making strategic decisions about the future product line of the firm. In-house R&D is necessary for monitoring and evaluating research conducted outwith the firm, and it ensures that the firm gets the most from external sources of basic research, e.g. universities. Even though scientific knowledge circulates in the outside environment, firms must undertake their own basic research in order to understand and utilise external science. Pavitt (1991, p.114) suggested that scientific research provides "skills, methods, and a web of professional contacts", which equips firms to exploit outside scientific findings. Cohen and Mowery (1984) argued that internal R&D facilitates this integration because it produces an intermediate good; this is firm specific knowledge that enables the firm to absorb external technology.

Finally, there is reason to expect that a firm's internal research and external collaborations are complementary activities (Arora and Gambardella, 1994 and Powell, Koput and Smith-Doerr, 1996). The firm's internally developed research capability allows it to better evaluate, understand, and assimilate new knowledge to which it is exposed via external collaborations. Hence, collaborations with university scientists should be more beneficial, in terms of accessing knowledge, to firms with more internally developed research capabilities. Likewise, internal research is of more value to firms that also engage in external collaborations because these firms increasingly exploit the absorptive capacity provided by their internal research.

Which is the best source of knowledge to use? Studies in this area have had mixed results. Various studies have shown that a combination of internal knowledge creation and external knowledge acquisition improves innovative performance (see, for example, Chang and Lee, 2008 and Deng, 2008). Chou (2005) found that external knowledge could influence internal knowledge creation. Rothaermel and Hess (2007) and Cassiman and Veugelers (2006) found that internal R&D and external knowledge sourcing are complementary innovation activities where their interrelatedness improves a firm's innovation performance. However, Hess and Rothaermel (2011) and Vega-Jurado, Gutierrez-Gracia and Fernandez-de-Lucio (2008) found the two sources of knowledge to be substitutes. The use of internal and external knowledge sources does not need to be mutually exclusive. Cohen and Levinthal (1990) found that these sources could be interdependent and complimentary. They proposed that firms must excel at internal knowledge creation and develop an absorptive capacity (an ability to absorb external knowledge) before they can learn from external sources.

Based on prior literature, citing NPR in firm patents, firm-level publishing and co-authoring behaviour indicates that the firm is undertaking both in-house research and utilising external knowledge sources in its research projects. This behaviour is expected to have a positive influence on the likelihood that the firm will produce patents that will be linked to an approved drug.

2.4.2 Speed of Learning

Is there an optimal speed for the firm learning process? In this thesis, I argue that pharmaceutical firms with greater success gaining regulatory approval are more likely to have a faster learning process. These firms utilise younger external knowledge compared to those that are less successful in gaining regulatory approval. Nerkar (2003) found that in the

U.S. pharmaceutical industry the quality of new technological knowledge is affected by the age of the prior art it references. According to March (1991), the speed at which a firm learns is determined by the available resources and the technology cycle time (TCT). The TCT is the median age of the patents cited on the front page of a patent document. This measure assumes that the more recent the age of the cited patents, the more quickly one generation of inventions replaces another. It can be measured by the median age of citations to prior art, where the learning process is faster when citations are more recent. Empirical papers by Narin, Carpenter and Woolf (1984) and Bierly and Chakrabarti (1996) found, for U.S. pharmaceutical firms, that utilising newer knowledge increased innovation and exploration and led to an increase in firm performance and profitability. They also found that the TCT is significantly faster for firms that predominately generate new knowledge internally and slower for firms that rely on external sources of new knowledge.

Citing newer NPR and patents in patent documents indicates that the firm is utilising newer knowledge in its research projects and so learning at a faster rate. Shorter non-patent and patent citation lag lengths are therefore expected to have a positive relationship with the likelihood of drug approval.

2.4.3 Depth and Breadth of the Firm Knowledge Base

Teece (1998) argued that firms need to accumulate, articulate, codify and use knowledge resources to create value and enhance performance over time. Merely processing knowledge resources may not be sufficient. He states that, "the competitive advantage of firms in today's economy stems not from market position but from difficult to replicate knowledge assets and the manner in which they are deployed." Carlucci, Marr and Schiuma (2004) found that the process of developing knowledge requires dynamic capabilities. These allow the firm to reconfigure its resources to improve performance. These dynamic capabilities are unique to a firm and as such will be a source of competitive advantage and superior long-term performance according to Teece et al. (1997).

A firm's depth of knowledge makes the search for new ideas more efficient allowing the firm to accelerate new product development. According to Lerner (1994) and Hall et al. (2005), if new knowledge is not closely related to a firm's existing knowledge, then the firm must work harder to absorb and assimilate it. If a firm's focus is too narrow then this will lead to inertia through a lack of adaptability. The deeper a firm's knowledge base, the wider the range of new knowledge it can assimilate. If a firm has more in-depth specialised

knowledge then this allows a local knowledge search to produce more successes. There is a positive relationship between the depth of technological knowledge and firm performance, but this benefit will reduce over time (Dosi, 1982).

A firm with breadth of knowledge can more easily integrate a wide range of ideas and technological knowledge from different areas. Knowledge breadth is required when a product is complex in nature. If a portion of a firm's knowledge is made obsolete by revolutionary technological change then it will more easily recover and maintain its competitive advantage with a broader knowledge base. Knowledge breadth creates flexibility, the ability to work in multiple areas and facilitates the combination of knowledge from different therapeutic areas (see, for example, Reed and DeFillipi 1990; Kogut and Zander, 1992 and Henderson and Cockburn, 1994). Henderson and Cockburn (1994) and Pisano (1994) found that for pharmaceutical firms, greater knowledge breadth led to improvements in firm performance and drug productivity.

If a firm references NPR in its patent documents and co-authors publications with other organisations, then this indicates that the firm is utilising external knowledge sources in its research projects. This behaviour is expected to broaden and/ or deepen the knowledge base of the firm and in so doing positively influence the likelihood of drug regulatory approval.

2.5 The Role of Firm-Level Absorptive Capacity in the Innovation Process

One of the goals of this thesis is to determine which firm-level characteristics influence differences in innovative performance within the pharmaceutical industry. Why do certain firms produce patents that result in successfully marketed drugs, whilst other firms do not? I argue that it is differences in the research capabilities of firms and their willingness to look outwith firm boundaries for research inspiration are key drivers of this differential.

According to the KBV of the firm, the foundation of a firm's performance lies in its ability to generate, combine, recombine, and exploit knowledge (Grant, 1996). Accordingly, a firm's capacity to absorb knowledge is important for value creation. Cohen and Levinthal (1990, p.128) introduced the term "absorptive capacity" which they define as the "ability to recognise the value of new information, assimilate it, and apply it to commercial ends".

The central argument of Cohen and Levinthal's approach is that the technological responsiveness of an established firm is a function of its absorptive capacity. It enables the firm to respond to technological opportunities and to appropriate the technological knowledge made available through R&D spillovers by competitors, and/ or organisations external to the industry (Cohen & Levinthal, 1989, 1990). They state that "a firm's absorptive capacity is not, however, simply the sum of the absorptive capacities of its employees" (Cohen and Levinthal, 1990, p.131). The firm-level attributes of external knowledge acquisition, assimilation and exploitation for innovative purposes must also be considered. Research teams need to have the capability to create and utilise knowledge in such a way that their firms can learn and innovate (Cohen and Levinthal, 1990).

Knowledge spillovers come at a cost to the recipient as firms are constrained with the amount of information they can process and use. Firms must invest resources in order to absorb knowledge spillovers. Cohen and Levinthal noted that this investment may take many forms, and that many mechanisms exist for building absorptive capacity. However, the form of investment that they primarily emphasise is prior and related R&D. According to Cohen and Levinthal (1990), the relationship between R&D and absorptive capacity arises because knowledge is either hard to codify or tacit (Polanyi, 1966) and is often embedded in the routines of the organisation (Nelson and Winter, 1982). From this perspective, R&D has two facets; it increases a firm's productivity and its absorptive capacity. They argue that firms undertake R&D not only as a direct input to innovation, but also as a means of absorbing external knowledge.

A firm's ability to utilise knowledge depends on the relatedness of the new knowledge to existing knowledge stocks (Hall et al., 2005 and Nagaoka, 2007). This theory is broadly consistent with empirical data, where Jaffe (1986) found that the interaction between a firm's R&D expenditure and spillovers is strongly correlated with the firm's performance. Similar evidence is offered by Gambardella (1992), Henderson and Cockburn (1996) and Lane and Lubatkin (1998). However, these studies are unclear as to whether R&D made some firms more successful because they were better at capturing spillovers or because their R&D was more productive.

The organisation may enlarge its knowledge base through the new application of pre-existing knowledge in the firm (Szulanski, Jensen and Lee, 2003) or through the generation of new knowledge by new combinations of pre-existing knowledge (Gratton and Ghoshal, 2003).

Even in circumstances where external, explicit knowledge involves high acquisition costs to the firm and is also simultaneously available to competitors, there are still knowledge-building opportunities available. If this knowledge is combined with the unique internal knowledge of the firm it may still result in new and exclusive knowledge (Zack, 2002).

Cohen and Levinthal suggested that a firm's prior scientific or technological knowledge must meet two criteria for it to be relevant enough to facilitate the understanding and valuation of new external knowledge (Cohen and Levinthal, 1990, p.136). First, the firm must possess some amount of prior basic knowledge that is associated to the new knowledge. Understanding the relevant basic knowledge permits the firm to understand the assumptions that shape the external knowledge and, thereby, to be in a better position to evaluate the importance of the new knowledge.

The next challenge is how to internalise this new knowledge. Following Nelson and Winter (1982), Cohen and Levinthal (1990) noted that the assimilation process is influenced by a firm's tacit, firm-specific knowledge regarding its established systems for processing knowledge. Complementary know-how enhances the transferability of technology between organisations because the more tacit, unobservable, complex and interdependent is the knowledge in question, the greater the need for the firm to have some basic understanding of that technology.

Given the increasingly significant role played by external knowledge flows in recent years, absorptive capacity has gradually become a key driver of a firm’s competitive advantage (Cockburn and Henderson, 1998). Resources such as knowledge, learning capacity, culture, teamwork and human capital contribute the most to the sustained competitive advantage of the firm (Hitt, Bierman, Shimizu and Kochhar, 2001 and Barney, 2001). Escribano, Fosfuri and Tribó (2009) suggested that the role of absorptive capacity is more pronounced in environments of tight IPR protection (such as the pharmaceutical industry).

In contrast with Cohen and Levinthal, Zahra and George (2002) stated that absorptive capacity is an organisational capability that enables a firm to be more competitive. They define absorptive capacity as “a set of organisational routines and processes by which firms acquire, assimilate, transform, and exploit knowledge to produce a dynamic organizational capability” (p.186).

They extended Cohen and Levinthal's theory to include knowledge acquisition and transformation in their model of absorptive capacity and define it as "a firm's capability to identify and acquire externally generated knowledge that is critical to its operations" (p.189). They also included assimilation in their model and define it as "the firm's routines and processes that allow it to analyze, process, interpret, and understand the information obtained from external sources". Once new knowledge is assimilated and transformed, it must be exploited; commercialisation of opportunities requires the exploitation of value in order to gain and maintain competitive advantage (Porter, 1980 and Zahra and George, 2002).

Pharmaceutical firms absorb knowledge either through utilising academic papers to explore new innovative ideas or by encouraging co-authoring with university researchers or researchers employed by other firms. This external knowledge is then assimilated into the firm by incorporating it into its drug research projects and by referencing this knowledge (in the form of scientific papers) in its patents. The knowledge is finally exploited when the firm creates patents based on this knowledge that are then developed into approved drug products.

The literature suggests two separate roles played by absorptive capacity with respect to external knowledge (Cohen and Levinthal, 1989; Arora and Gambardella, 1994 and Zahra and George, 2002). First, absorptive capacity helps the firm to identify more available knowledge flows. Second, for a given quantity of identified external knowledge, the degree by which the firm derives benefits also depends on its absorptive capacity. Heterogeneity in the level of absorptive capacity translates into differences in the benefits from otherwise similar stocks of external knowledge, both because the firm can identify more of them and because it can exploit them more effectively than other firms.

A substantial body of theoretical work suggests that idiosyncratic research capabilities are likely to be a particularly important source of strategically significant "competence" in science-and technology-driven industries (Dierickx and Cool, 1989). Cohen and Levinthal (1989, 1990) used R&D intensity to measure absorptive capacity. However, Zahra and George (2002), criticised this measure and argued that R&D intensity does not reflect expenditure on all dimensions of identification, assimilation, transformation, and exploitation of knowledge. They developed the concept of potential and realised absorptive capacity. Potential absorptive capacity includes knowledge acquisition (e.g. measured by years of R&D experience) and assimilation capabilities (e.g. measured by the number of patent citations to other organisations' research). Realised absorptive capacity includes

transformation (e.g. measured by the number of new research projects or the number of new product ideas) and exploitation capabilities (e.g. measured by the number of patents or the number of new drug product approvals). According to Zahra and George (2002), it is realised absorptive capacity that is positively related to future value creation of the firm.

Absorptive capacity enhances a firm's ability to judge the probability of successfully turning a given piece of basic research into a profitable product. Firms with greater absorptive capacity are more likely to pursue projects with a higher probability of success due to their superior knowledge. If the probabilities for a research stream change, due to new discoveries beyond the boundaries of the firm, then firms with higher absorptive capacity will sense this change more quickly and adjust their research efforts in accordance with the new information.

2.5.1 Empirical Studies on Absorptive Capacity

Nicholls-Nixon (1995) examined the role of absorptive capacity in pharmaceutical firms' responses to the competency-destroying technological changes created by the emergence of biotechnology. She measured pharmaceutical firms' absorptive capacity in three ways, using the number of biotechnology patents held by the firm, the number of new products it had on the market or under development and the firm's reputation for expertise in the applications of biotechnology. She suggested that responsiveness to technological change is a function of how firms develop their absorptive capacity. Firms with high levels of absorptive capacity invested more in their own R&D, utilised alliances, had more in-house expertise with relevant technologies and managed communications with alliance partners more effectively.

A considerable and growing literature describes and tests the importance of firm research activities in identifying, assimilating, and exploiting external knowledge. Literature considering the role of scientific knowledge in the innovation process characterises search for innovation as an uncertain process in which knowledge can provide a guide (Fleming, 2001). This work demonstrates the importance of scientific knowledge in the search process for innovations. The limited empirical research in this vein confirms that a better knowledge base and connectedness to external scientific knowledge; generates more impactful inventions, especially when the search process is complex (Fleming and Sorenson, 2004). Knowledge provides researchers with an understanding of the fundamental principles underlying a system, which may allow them to better anticipate the result of various possible experiments without proceeding with the experiment (Fleming and Sorenson, 2004).

Two aspects of firm research are associated with building absorptive capacity; conducting internal basic science research and forging external linkages with university scientists. In several empirical studies on the determinants of success in the pharmaceutical industry, researchers demonstrated a positive and significant relationship between internal basic research and collaboration with university scientists and the number of pharmaceutical firm patents generated (Cockburn and Henderson, 1998) and the importance of biotechnology firm patents (Baum, Calabrese and Silverman, 2000 and Zucker, Darby and Armstrong, 2002). Whilst Gambardella (1992) found that firms with superior in-house research programmes were better positioned to exploit public science, this ability was also correlated with research productivity. These studies provide some evidence of the expected innovative performance benefits of absorptive capacity.

A re-examination of the empirical literature from a social network perspective highlights the importance of connectedness. Social networks and informal interactions provide an important avenue for the exchange of tacit, research-related knowledge. Liebeskind, Oliver, Zucker and Brewer (1996) reported that scientists at two biotechnology firms improved the integration of external knowledge by collaborating with outside researchers. These connections are important to the development of the absorptive capacity necessary to internalise available knowledge and improve the research performance of a firm. Von Hippel (1994) suggested that those firms that reach outside the organisation for critical knowledge perform more effectively than those that do not. Cockburn and Henderson (1998) and Fabrizio (2009) found a relationship between a firm's research productivity and its rate of co-authorship with universities and public sector laboratories. Fabrizio also found that a firm's own basic science research and collaborations with university scientists are associated with a faster pace of knowledge exploitation for pharmaceutical and biotechnology firms. Zucker and Darby (1995) showed that biotechnology firms that collaborated with star scientists at universities (or employed them) were faster adopters of new technological knowledge.

The development of the firm's ability to access external knowledge sources is not limited to activities that take place strictly within the firm. As emphasised by Zucker, Darby and Armstrong (1994), Cockburn and Henderson (1998) and Zucker, Darby and Brewer (1998), "connectedness" to outside knowledge sources (particularly scientists) provides benefits in terms of accessing and exploiting external knowledge. Other work has focused on the ability of firms to use connections and collaborations with university and other public sector

scientists to gain advantage by accessing and developing public sector science (Zucker et al., 2002). Informal interactions are particularly important avenues for knowledge transfer between universities and industry (Cohen, Goto, Nagata et al., 2002), and not only help to identify relevant scientific research, which may or may not be published, but also provide the firm with access to the tacit knowledge complementary to published research results.

Henderson and Cockburn's (1994) study of pharmaceutical firms found that the more they emphasise publications and external reputation when evaluating their research staff, the higher their rate of new drug development. Based on previous empirical findings I expect that firms that allow their researchers to publish academic papers and co-author papers with universities or other firms are improving their own basic science research. In so doing they will improve the firm's absorptive capacity and its ability to absorb external basic scientific research in the future. Collaboration with university researchers will enable firms to become faster adopters of new technological knowledge, which will be reflected in the citing of newer NPR in their patent documents. This will then improve the firm's search for novel ideas, improving research performance and increasing the likelihood that their research output will result in an approved drug product.

2.5.2 Absorptive Capacity and New Product Development

A firm's ability to develop a new drug depends on its absorptive capacity, which is defined by foundational knowledge and technical capabilities developed through continuous R&D (Hall et al., 2005).

Absorptive capacity has been examined in the innovation and technology management literature in different contexts and found to impact New Product Development (NPD) performance. The role of absorptive capacity has also been addressed in product innovation literatures (see, for example, Maes and Sels, 2014), and is recognised as playing an important role in learning.

University-generated science is an important source of external knowledge for innovation in the pharmaceutical industry. According to McMillan et al. (2000), public science is a key input for innovation in the industry. University research results are a valuable knowledge source for many firms, with their importance growing through time. Narin, Hamilton and Olivastro (1997) found that 73% of references in U.S. industrial patents are created by publicly funded institutions. Many studies discuss how industries use university-based basic

scientific research in the development of new products and processes (Narin and Olivastro, 1992; Mansfield, 1998; McMillan, Narin and Deeds, 2000; Cockburn and Henderson, 2001; Grossman, Reid and Morgan, 2001 and Cohen et al., 2002). Medicinal drug patents cite significantly more scientific publications than patents in other fields (Narin et al., 1997), and these patents cite a greater number of basic scientific research journals (Narin and Olivastro, 1992). Pharmaceutical firms rely on basic science developments in biology and biochemistry, with many new drugs having their origins in university-based discoveries. Industry-based researchers believe that new products could potentially have been delayed without access to this research (Mansfield, 1981, 1998 and Collins and Wyatt, 1998).

Fleming and Sorensen (2004) find that using scientific knowledge in patented inventions results in innovations of greater importance. This is especially true for inventions that are complex in nature and that rely on a science-centred innovation search process.

Following on from these empirical findings I expect that firms that co-author with university researchers and cite academic papers in their patent documents will have an increased likelihood of producing patents that are linked to an approved drug.

2.6 Patent-Based Measures of Innovation

Studying patents and their characteristics may facilitate the measurement of intangible assets. Knowledge- and research-intensive firms have a large percentage of their assets in the form of intangibles, for example patents or drugs in development. Measuring these intangibles is problematic, as the actual private economic value of these assets is not realised until the product is released to the market.

2.6.1 Patent Counts

Schmookler (1962, 1966) was one of the first researchers to make use of patents statistics. He recognised that patents may have a role in the innovative process and could perhaps be an informative indicator. Patents are a technology-related variable that can be viewed as an important intermediate output of the innovation-driven R&D process. Investment in R&D creates intangible intellectual capital within a firm, in the form of patents, and this can affect the market value of the firm. The impact on market value was utilised as a partial indicator of the expected success of the inventive efforts of the firm. Patent strategy can affect firm value and performance generated through the appropriability outcomes. In the late 1970s, Zvi Griliches also recognised the potential value of this data source. Early research proposed

that the number of patents applied acts as a proxy for the rate of new knowledge generation, allowing the measurement of knowledge creation across industries. These studies had a multi-industry focus and the majority studied the U.S., using patent counts as an indicator of the value of the intangible innovative assets present within a firm. Researchers such as Griliches (1981), Ben-Zion (1984), Jaffe (1986), Connolly, Hirsch and Hirschey (1986) and Connolly and Hirschey (1988, 1990) found a positive relationship between the value of a firm and its intangible assets (e.g. patents). Jaffe (1986) also found that a higher ratio of patents per unit of R&D outlay occurs in sectors where firms spend more on R&D.

Griliches, Pakes and Hall (1986) found that patents explain changes in market value levels but are less important and add little in the presence of R&D expenditures. Griliches (1981, 1984 and 1990) also utilised patent counts as a proxy for innovation. These works modelled R&D, patents and performance measures (e.g. productivity growth, profitability and market value). Griliches (1984) studied the observed strong relationship between R&D expenditures and patenting levels. Pakes (1985) found that patents are an indicator of current and past inputs into the R&D process. Market value is an indicator of the outputs of the R&D process. Unexpected changes in R&D expenditures and patenting levels can have a large impact on a firm's value (Cockburn and Griliches, 1988; Griliches, Hall and Pakes, 1991 and Hall, 1993). The study by Griliches and his co-authors found that information on changes in patenting rates does not provide additional information beyond that contained in R&D, as patenting rates account for only a very small fraction of changes in the stock market value of the firm.

Some UK and European studies linked patent data to external performance measures, e.g. firm market value. One of these studies by Hall, Thoma and Torrisi (2007) found that Tobin's Q is positively related to R&D and patent stock. Blundell, Griffith and Van Reenen (1995, 1999) and Toivanen, Stoneman and Bosworth (2002) conducted similar studies for the UK. Toivanen and his co-authors found that the UK equity markets value R&D, but the market value is more sensitive to the first announcement of R&D expenditure in the firm's accounts. Other researchers, such as Griliches, Hall and Pakes (1988) combined patent data with the firm-level financial data and found that patents were less strongly correlated with firm performance than with R&D expenditures and R&D intensity. Work on patent data was still combined with R&D–related data, with R&D being the dominant variable in many papers. Bosworth and Mahdian (1999) examined UK pharmaceutical R&D expenditures and patents and found they had a positive impact on market value with both measures performing

an approximately similar role in the explanation of market value. McMillan, Mauri and Hamilton (2003) found that, for the U.S. pharmaceutical industry, the number of patents controlled is positively linked to the development of new drugs and firm performance.

There were problems with this approach as researchers noted that the distribution of patent economic value and technological importance was highly skewed in nature. That is, a small number of patents were worth a great deal and many patents had little or no economic value. Schankerman and Pakes (1986) found, in their study of UK, French and German patents, that the economic significance of individual patents is variable, with only a few highly valuable patents. The work of Harhoff, Narin, Scherer and Vopel (1999), Harhoff, Scherer and Vopel (2003) and Gambardella, Harhoff and Verspagen (2008) also supported this view.

There are problems therefore in using patent stocks or counts, as unimportant patents will tend to overshadow valuable ones. Griliches et al. (1991), in their U.S. study, found patents were a noisy measure of the underlying economic value of the innovation linked to them. He understood the potential value, to researchers, of utilising the information contained within the text of the initial pages of patent documents, which could have some importance in outlining the innovation process itself (see also Trajtenberg, 1990). Trajtenberg found that simple patent counts cannot be very informative with reference to innovative output. “It has long been thought that the detailed information contained in a patent document may have a bearing on the importance of the innovations disclosed in them” (Trajtenberg, 2002). He recognised that simple patent counts can have lower information content and realised that other patent characteristics and statistics could be utilised as potential economic indicators. There are difficulties with this approach, which attempts to measure private returns to innovative activity.

2.6.2 Patent Characteristics

Researchers had always recognised the potential of patent documents as data sources. In the late 1970s, the USPTO started to digitise its patenting information. A few researchers (see, for example, Jaffe and Trajtenberg, 1996 and Hall, Jaffe and Trajtenberg, 2001) took on the task of collating this multi-industry data so that the patent data could be converted into a useful resource and they created an extensive new dataset. Patent documents had only been studied to a limited extent previously, not due to the documents themselves; they were a rich source of data, but due to the practical difficulties of collating the sheer volume of data. The scope of the original data made it virtually impossible to process manually, although some

researchers did attempt this (Trajtenberg, 1987). Because of the work (sponsored by the National Bureau of Economic Research (NBER)) of Jaffe and Trajtenberg and others the newly created NBER patent data file project opened many new avenues for empirical patent research. They organised the data in a systematic and useful manner and created new measures from the patent data.

Patent data research can be classified into two separate strands. Firstly, there are studies that substantiated the role of patents and citation-based measures as indicators of technological impact by only examining patterns and relationships occurring within the patent data (see, for example, Lanjouw and Schankerman, 1999 and Harhoff, Scherer and Vopel, 2003). Secondly, other researchers correlated patent-based data with independent technological and economic indicators, such as firm market value (see, for example, Aboody and Lev, 1998 and Hall et al., 2005). Differing measures are also used to measure the economic value of a patent, firm market value (Hall et al., 2000) or survey-based measures of value (Harhoff, Scherer and Vopel, 2003).

Surveys of patents-holders were conducted to determine value by Sanders, Rossman and Harris (1958) for the U.S., Harhoff et al. (1999) and Harhoff, Scherer and Vopel (2003) for Germany, Gambardella et al. (2008) and Giuri, Mariani and Brusoni et al. (2007) used survey data for European firms and Nagaoka and Walsh (2009) looked at Japan. These surveys measured the value of owning a patent right relative to someone else owning the patent. This approach does not cover the same value as the approach that utilises patent renewal data to measure value. The renewals-based approach measures the patent asset value (Scherer, 2010).

2.6.2.1 Patent Citation Data

Studies have identified that quality is reflected by the citation level of a patent. By inference, patent quality also includes private economic value. Citations are linked to the technological importance of an innovation. (see, for example, Trajtenberg, 1990 and Jaffe, Trajtenberg and Fogarty, 2000)

Patent documentation contains different types of citations. Citations are not generic, some are added to avoid patent infringement, and some are added by the patent examiner. The USPTO states that all relevant citations (described as prior art) must be included. There is a duty of disclosure where an applicant is obliged to inform the patent office of any relevant

documents that are known to them that may have a bearing on the claims that result from the patent application. It covers all information reasonably deemed necessary and this needs to be included.

Patent citations should not be compared to bibliographic citations. Citations in patents have a legal function or purpose. Bibliographic journal citations indicate sources of influence or demonstrate the novelty of the research. They indicate and acknowledge the existing knowledge upon which the paper builds. This is not necessarily the case with patent citations as these often added to assist the evaluation team to assess the patent's inventiveness and to qualify the patent's claims.

Citation data was one of the first pieces of data to be utilised from the patent front-page. Researchers recognised (see, for example, Griliches et al., 1986 and Trajtenberg, 1990) that this form of patent data could be utilised to quantify the private economic value and technological importance of an invention and to measure knowledge flows and spillovers. From these data, the pace of technological change could be determined.

There are two main citation classifications: backward citations and forward citations.

(a) Backward Citations

Patent citations are references to all prior art (this is all publicly available documentation that contributes materially to the patented entity report) and are cites of already existing knowledge. These citations can be references to other patents, non-patents literature (journal articles, newspapers, technical manuals etc.) on which the citing patent builds. Prior art can be provided by the patent examiner and/ or the inventor and is relevant for assessing the invention and its claims. Patent citations signal knowledge spillovers and knowledge flows (see, for example, Jaffe, Trajtenberg and Henderson, 1993 and Jaffe and Trajtenberg, 1996). They also indicate the firm's absorptive capacity for recognising, assimilating and applying external knowledge to produce a commercial product (Cohen and Levinthal, 1990).

A patent with higher numbers of citations to previous patents (backward citations) is likely to be more derivative in nature and therefore less valuable. The invention it covers builds on existing technology and is not as innovative (Hall et al., 2001). If a patent is very novel it might have very few backward citations to already existing patents, as it is a radical invention

that is breaking new ground. An invention that is incremental in nature will have higher numbers of backward patent citations as it builds on already existing knowledge.

Lanjouw and Schankerman (1999) found that backward citations were positively related to patent quality in the U.S. pharmaceutical industry. Harhoff et al. (1999) and Harhoff, Scherer and Vopel (2003) supported these findings for German and U.S. held patents covering a range of industries.

Non-patent references (NPR) are citations to papers within the patent document and researchers consider them to be an indicator of knowledge use in an industry. These can be references to scientific papers, technical journals etc. They are indicators for the use of external knowledge. Most NPR are to journal articles; this demonstrates that technological developments are linked to scientific knowledge and can signal a patent's originality and novelty and also science-technology interactions. Fewer NPR indicate a less radical form of invention. Nagaoka (2005) found these references were positively correlated with a firm's Tobin's Q for U.S. patents held by Japanese firms. Fabrizio (2009) found that citations to scientific literature occurred at a faster speed than references to patents.

If a patent has fewer backward citations to patents but more to NPR, this may be an indicator of a more radical inventive step grounded in more basic science.

(b) Forward Citations

Forward citations are citations to a patent from patents that are filed at some future date. Studies have linked these citation levels to the technological importance of an innovation. They are viewed as an indicator of knowledge relatedness, knowledge flows and broader dissemination of the knowledge within a patent. More successful patents that are technologically important or economically valuable tend to have higher citation levels. The NBER dataset found that patents receive 50% of their total citations within the first 10 years. Citations received by patents can be viewed as a single indicator of their technological quality. They also signal potential future private economic value of the invention. They are an indicator of the probability of the invention becoming a marketable drug and a source of future profits.

Patents in a well-developed area of technology will be watched more closely by competitors and may receive higher levels of forward citations and receive them within a relatively short time lag.

Forward citations reflect the confidence that third party experts have in the idea. They indicate that competitor firms, or the industry in general, assess the patent as having value (technological and potential economic value). They act as a non-financial leading indicator of a firm's innovation capabilities and its future profitability. They contain information on the firm's technological superiority, which in turn can enable a firm to increase its future earnings. These citations can be an early indicator of downstream economic value (Lanjouw and Schankerman, 1999). The higher frequency of forward citations cannot prove anything on its own and needs to be related to other factors. Higher citation frequency can also be an indicator of a rapidly expanding research area with many participants, so increasing the level of citations. If an invention is very radical, forward citations may have a longer time lag as it will take competitors, who are unfamiliar with the area, longer to recognise the potential of the idea. Forward citations may then increase when the patented invention is launched to market.

Carpenter et al. (1980) found that patents associated with more "important" inventions had a higher likelihood of being cited by subsequent patents. Hall (1999), in her U.S. multi-industry study, proposed that it might be more informative if patent counts were weighted by subsequent citations and found that the market value of a firm was strongly related to its knowledge assets when using this measure. Trajtenberg (1990) introduced the use of forward citation data as a more meaningful measure of inventive activity and found that citation-weighted patent counts can be a useful measure of a patent's economic value. Hall et al. (2000) found that at a given date the market value of a firm is related to the stock of forward citations after that date. The social and private value of patents also appears to increase with citation intensity and that citations to a firm's own patents are associated with twice as much market value as citations from outwith the firm. Jaffe et al. (1993) identified forward citations as indicators of knowledge spillovers. Initial work in this area only dealt with very small samples because citations made by subsequent patents had to be traced to the original patent of interest. This required a large research effort to achieve a reliable method of collating forward citation statistics.

Jaffe et al. (1993) p. 578 state that "knowledge flows sometime leave a paper trail in the form of citations to patents". Trajtenberg (1990) identified that patent-to-patent citations reflect spillovers and pathways of innovative trajectories. Downstream patents cite upstream patents upon which they have built. Citations are related to both private economic value and social value. Citations are manifestations of knowledge linkages and allowed the development of non-trivial robust indicators of innovation. These bibliographic indicators were used to map innovation patterns and assess connections. Researchers worked on the premise that more economically valuable patents would be cited more often. Hall et al. (2005) found that, for U.S. firms in multiple industries, the addition of one citation was linked to a 3% rise in a firm's market value and that citations below the median level had no impact on value but that citation levels above this median level did have an impact. This work indicated a non-linear relationship between a firm's patent forward citation levels and its market value. Sampat and Ziedonis (2004) viewed patent citation data as a proxy for knowledge flows and utilised them as a predictor of successful licensing outcomes.

Many studies have found that better quality patents have higher numbers of citations from subsequent patents (forward citations). Citations indicate the value of a firm's innovation and underline the potential of the firm's patent portfolio. If a patent has above average citation levels, this may result in more knowledge spillovers and therefore that the patent is technologically more important and more likely to be more economically valuable. Citations imply that the original patent proved valuable and was viewed as worth further work. Kogan, Papanikolaou, Seru and Stoffman (2012) found that, for the U.S., the market reacted positively to patent application, and grant information, and the cite level increased subsequently for that patent. They can reflect the level of market interest in that patent or technology (See also Czarnitzki, Hall and Oriani, 2006 and Grandi, Hall and Oriani, 2009). There is evidence that patent citation measures (as a proxy for patent quality) are positively associated with the future benefits from innovation. Patent citations can be viewed as a non-financial indicator of a firm's innovative capabilities and an indicator of improved firm performance. Other studies have found that an increase in patent citation levels is associated with an increase in abnormal stock returns and that these effects can last up to three years (Gu, 2005). The strength of the relationship between citations and earnings levels increases over time and then levels off. Citations can provide an indication of the nature of a firm's technological superiority. This superiority may lead to increased firm performance in the future and increased earnings. Narin et al. (1987), Lanjouw and Schankerman (1999) and Chen and Chang (2010) found this relationship for U.S. pharmaceutical firms. A firm's

citation levels may be a good indicator of its research prowess and an efficient idea-search process.

Citations were viewed as a, probably noisy, indicator signal of the technological relationship between the citing and cited inventions. The process of technological change is viewed as a cumulative process, through which inventors build on the work of previous inventors and at the same time contribute to the pool of future available knowledge. Citations are an imperfect map of the links between pieces of innovative output and knowledge. They are a road map that links inventive ideas and traces the flows of technological knowledge.

2.6.2.2 Timing of Citations

Whilst the number of citations a patent receives has an important role, the timing of these citations can also be an informative factor. The more quickly a patent is cited may be an indicator of its importance. Most citations occur over a fifteen-year window, but citation distribution through time varies between patents. Forward citations can be viewed as more relevant and reflect patent quality if they occur closer to the patent publication date[32]. Therefore, the selection of the citation window length can have an impact. Some studies have found that a shorter citation window is a better indicator of patent importance. Lanjouw and Schankerman (1999) and Deng et al. (1999) found that citations received within five years of the patent grant date are more likely to reflect patent quality, although Lanjouw and Schankerman found that this window was extended to ten years for the pharmaceutical industry.

2.6.2.3 Patent Claims

Patent claims, in technical terms, are the extent of the protection conferred by a patent, or the protection sought in a patent application. The claims are of the utmost importance during both prosecution and litigation. The claims list the technical areas to which the patent lays legal claim. These claims legally define a patent's coverage and the novelty of the patent. Many claims have been interpreted as an indicator of the general feasibility, validity,

[32] For patent applications filed prior to 29th November 2000 the patent grant date and patent publication date are the same. USPTO patent applications for a patent filed on or after 29th November 2000 are published promptly after the expiration of a period of eighteen months from the earliest filing date. This means the patent can be publicly available prior to its grant date.

originality and potential profitability. Pharmaceuticals patents have a high number of claims compared to computer and other industries evidenced in the NBER dataset.

In studies of U.S. multi-industry patents, Lerner (1994) and Egan (2012) found that claims had a positive impact on the value of the firm. Lanjouw and Schankerman (1999, 2001), Allison et al. (2004) and Reitzig (2002) found claims had a relationship to the likelihood of litigation.

2.6.2.4 Patent Scope

Researchers have proposed several measures of patent scope; these included the number of patent citations received by a patent, the number of patent claims or the number of specific International Patenting Classification (IPC)[33] subclasses to which a patent is assigned (Lerner, 1994). The scope of a patent defines the extent of the legal coverage that a patent provides. Patenting scope is viewed as a value driver by many studies, bearing on its economic value, and hence its inventive effect. A broader right pre-empts more substitutes than a narrow right (Menell and Scotchmer, 2007). This has implications for rivalry within an inventive field; high levels of rivalry can be problematic as found by Barzel (1968), Kamien and Schwartz (1976) and Dasgupta and Stiglitz (1980).

The broader the scope of a patent the more difficult it is for rivals to find a competitive alternative. A patent with a broad scope can limit the opportunities available to other inventors. Lanjouw and Schankerman (1999, 2001) found that the narrower the scope the more likely it was that the patent would be challenged; this result was supported for the European biotechnology/pharmaceutical industry, by Harhoff and Reitzig (2004). Lerner (1994) results showed that, for the U.S. biotechnology industry, the broader the patent scope the more likely was litigation (see also Knight, 1996 and Harhoff, Scherer and Vopel, 2003).

Litigation against patents implies that a patent is of higher value and higher quality. Litigation is a very rare occurrence, normal rates are 1 to 2% in the U.S., but Lanjouw and Schankerman (2001) found the probability of litigation of valuable patents is approximately

[33] IPC classes are used to classify patents according to the different technology areas they cover. Patents that relate to many different technologies are broader in scope than those that are specific to a more limited range of technology classes.

25% for the pharmaceutical industry (see Moore, 2000; Harhoff, Scherer and Vopel, 2003; Allison et al., 2004 and Allison, Lemley and Walker, 2011).

Austin (1993) investigated the scope of patents in the U.S. biotechnology industry and their market value. He found that there was no statistically significant evidence that patents with a broader scope are more valuable than those with a narrower scope.

2.6.3 Science Linkage

Other researchers were interested in the role of knowledge creation within the innovation process. Mansfield (1995) found that there were spillovers from academic research into private R&D. Cohen et al. (2002) and Zucker et al. (2002) found evidence of university-industry cooperation in R&D. They were interested in the linkage between basic (scientific) research carried out in universities and research institutes and commercial (corporate) R&D activities. Basic research (conducted by universities and research institutes) does not persist across industries in the short run but it is useful in the long run as a knowledge stock. This stock can be used when looking for technical solutions, to solve unseen problems and discovering market opportunities. However, in the pharmaceutical industry basic research has an immediate relevance to the drug discovery process.

Firms have tended to move away from the vertical (in-house) model of R&D toward a "network strategy" of innovation based on the exploitation of external knowledge resources. This is illustrated by the increased industrial funding of university research (for OECD countries whose funding was 3% of GDP increasing to 6% in 2003 according to Vincent-Lancrin (2006). Czarnitzki, Glänzel and Hussinger (2007) and Czarnitzki (2009) found that academic patents tend to be more highly cited than non-academic patents. Cohen et al. (2002) found that scientific publications were also a highly rated channel of communication between universities and firms (Carnegie Melon Survey).

Agrawal (2001) found certain firm characteristics for firms that choose to interact with universities. Firms need absorptive capacity in order to be able to recognise and absorb the external knowledge created by university research. Firms must have knowledge in each area to be able to absorb the external knowledge from that area and they must improve their absorptive capacity in order to attain competitive advantage. Firms need to become participants in science rather than mere users of scientific knowledge.

Innovation surveys support this linkage; Yale Survey (early 1990s), Carnegie Melon Survey and the European Community Innovation Surveys (4 surveys 1991 to 2004). Mansfield also conducted surveys in 1991, 1995 and 1998. The linkage was also supported by patenting data e.g. joint university-industry patents, NPR etc.

At an individual level, there are inventor-author linkages and at the firm-level, there are strategic alliances, co-patenting and co-authoring. However, the low numbers of co-patents and alliances may limit the applicability of this measure (Hicks and Narin, 2001).

Narin and Noma (1985) examined the nature of NPR (see also Van Vianen, Moed and van Raan, 1990 and Harhoff, Scherer and Vopel, 2003). Many researchers examined citations to NPR (Van Looy, Zimmermann, Veugelers et al., 2003) and found a relationship between the science intensity of patents measured by NPR and technological productivity (output per unit of input, in this case R&D expenditures). They investigated the nature of science-technology relationships implied by citation links. NPR signal the direct influence of science on technology. Narin et al. (1997) and Meyer (2000) conducted patent case studies and found that NPR should not be interpreted as indicating a direct, unidirectional link or influence from science on technology.

Harhoff et al. (1999) found that linkages to basic scientific research improve technology interactions and innovative performance. Fleming and Sorensen (2004) and Nagaoka (2007) found a positive and significant relationship between science and patents. Tijssen (2000 and 2001) conducted a survey of inventors in the Netherlands and found that citations represent genuine direct observable links between research and technical innovations. NPR should be considered a general indicator of the interaction between science and technology; they are not an indicator that science leads directly to invention. Scientific references do not represent a direct link between the citing patents and the cited article. Scientific references indicate the relevance of scientific findings and technological developments. There is no direct causal link and an increase in NPR only implies the relevance or relatedness between science and technology. Scientific literature plays an indirect role as background information and indicators based on NPR may prove useful. Narin and Olivastro (1992) showed that there was an increase in science citations in U.S. patents in the 1980s and that the median age of scientific papers fell.

Callaert, Van Looy, Verbeek et al. (2006) found that in the USPTO, pharmaceutical patents had the highest average number of NPR, when compared to other industries. This is the only

sector where NPR intensity is greater than the technology intensity (Verbeek, Debackere, Luwel et al., 2002 supported this finding). Tijssen (2001) found that there is a link between patent assignees or inventors to author information in articles, implying that the assignee or inventor is involved in knowledge flows. The patent-science-intensity proxy is defined by backward citations to science articles. Patents that list science references show that science literature is a greater influence on the inventive process. Patents with no references to external non-patent publications indicate that contacts within firms have more influence on the inventive process. Scientific-NPR signal relatedness to science instead of a direct knowledge spillover; they show science-industry interactions and act as a source of inspiration.

Callaert, Grouwels and Van Looy (2012) found that, for Belgian science-intensive technologies, scientific knowledge was the source of inspiration for nearly 30% of inventions patented in the USPTO. Tijssen (2000) concluded that a university or research institute connection was an important influence on invention. Patents may not contain science references, but they can still be inspired by scientific research. These references are just not relevant to the patent. Sternitzke (2009) found that external scientific knowledge has commercial relevance, it is not just background information. Roach and Cohen (2013) found that survey data combined with quantitative data indicated that NPR are a better measure of knowledge originating from public or basic research. Patent references are less adequate when explaining the value of public research and underestimate its value.

Most patents will have more than one inspiration source. Influences from professional expertise and in-house research are also important. Meyer and Persson (1998) assessed the role of tacit, scientific knowledge and its importance to the inventive process.

2.6.4 Patents as a Signalling Mechanism

Pharmaceutical firms have inside knowledge of potential drug candidates. They make a choice to patent and gain future market exclusivity and in so doing they disclose their ideas and place them in the public domain. The level of disclosure varies; some patents disclose a lot and some a minimal amount of information. A patent itself can be an indicator of the firm's belief in the potential of a drug, as a firm does not patent all the candidates it discovers.

Austin (1993) and Darby, Liu and Zucker (2004) examined the impact of individual patent applications, grants or announcements on market value. Horstmann, MacDonald and

Slivinski (1985) found that the act of patenting signals the information accrued by the inventor during the research phase of product development.

Researchers have also studied U.S. patent data in its role as a signal to investors (Mann, 2005; Mann and Sager, 2007; Cockburn and MacGarvie, 2009; Graham, Merges, Samuelson and Sichelman, 2009; Sichelman and Graham, 2010 and Hsu and Ziedonis, 2011). Haeussler, Harhoff and Mueller (2009) and Helmers and Rogers (2011) used a similar approach with European and UK data. Patents act as a signal to investors that a firm possesses a valuable asset even in the absence of a current profit stream. Technology intensive start-ups often use patents this way. Graham et al. (2009) and Sichelman and Graham (2010) found that this view was supported by a survey of American entrepreneurs. A firm signals the quality of its research to investors and competitors. This is done via patenting, publishing, joint publishing, joint patenting and publishing with universities. This signals its strong links to basic scientific research and can be viewed as a quality indicator.

In the initial stages of the drug discovery process firm patenting decisions may be influenced by the disease area targeted by their drug candidate. The potential market and revenue stream for any new drug that targets a condition such as Diabetes is greater than for T.B, as described previously in Section 2.3. This increases the likelihood that firms may choose to focus their research projects in these disease areas as there may be increased profit-earning opportunities. Investors may also consider a drug candidate's target market as this might influence future firm revenues and the ability of their investments to earn increased returns.

In my sample of FDA approvals (see Section 4.2.2), 83% of the conditions these drugs treat are non-communicable. If a firm creates a drug candidate that treats a non-communicable condition it signals that the drug may have the ability to earn higher future revenues and may increase the likelihood of FDA approval as there is an existing track record for approving drugs for these conditions.

2.7 The Impact of Regulatory Drug Approval on Shareholder Wealth

One of the goals of this thesis is to confirm that certain patent-level and firm-level characteristics can be identified with patents that are linked to drugs that receive regulatory approval. Using these identified characteristics, I seek to predict patents that will be linked to approved drugs, utilising these predictions to construct an investment portfolio.

FDA approval of an NDA removes the uncertainty experienced by investors during the drug approval process. The submission of an NDA signals the firm's belief that the drug research project is viewed as a worthwhile invention. The empirical literature consistently finds that there is a significant positive market reaction to drug approval announcements across several markets (Bosch and Lee, 1994; McNamara and Baden-Fuller, 2007 and Hamill et al., 2018 for the U.S., Dedman, Lin, Prakash and Chang, 2008 for the UK and Hamill et al., 2013 for both the U.S. and UK).

Early studies evaluated FDA approvals, rejections, and other announcements for spillover effects and information leakage (Bosch and Lee, 1994; Ingrain, Shastri and Shastri, 1994; Torabzadeh, Woodruff and Sen, 1998 and Sharma and Lacey, 2004). Collectively, this literature reports significant increases in shareholder wealth for drug approvals and decreases for rejections. Sharma and Lacey found a greater reaction to rejections than approvals.

Bosch and Lee (1994) found that FDA decisions had very large wealth effects, with large price movements associated with approval or rejection. Their results indicated that there seemed to be some uncertainty about FDA decisions present almost up to the announcement day. Hamill et al. (2013) investigated the value implications of FDA application and approval announcements for firms listed on both the New York (NYSE) and London (LSE) stock exchanges. Applications increased shareholder wealth for NYSE-listed firms only and approvals increase shareholder wealth in both markets. For the UK, a significant market reaction around approvals is consistent with Dedman et al. (2008), who concluded that the UK market reaction increases as a drug candidate progresses through the drug development process, being greater for late stage rather than early stage announcements.

Drug-in-progress announcements are a significant performance indicator and have been shown to significantly increase shareholder wealth along the various stages of the exploration-exploitation continuum (McNamara and Baden-Fuller, 2007 and Dedman et al., 2008). In intangibles-intensive industries, non-financial performance indicators provide important signals to investors of firm performance and can substitute for financial disclosure (Amir and Lev, 1996; Eccles, Herz, Keegan and Phillips, 2001; Ely, Simko and Thomas, 2003 and Espinosa, Gietzman and Raonic, 2009). A key value driver for biotechnology and pharmaceutical firms is the non-financial announcements from outcomes at various stages of the drug approval process. An announcement of a firm's success, at any stage of the drug development process, is at the discretion of management (unlike patent grant

announcements) and, as such, falls into the voluntary disclosure category. This is classified in the literature as a non-financial performance indicator.

For a sample of biotechnology and pharmaceutical firms listed on the NASDAQ, London, Paris, Frankfurt and Milan stock exchanges, McNamara and Baden-Fuller (2007) assessed the market's reaction to announcements of various stages of drug development. For each of the stages, they report a significant positive market reaction, with evidence of 'peaks' in the market's reaction at the latter end of exploration (patenting and pre-clinical trials) and exploitation stages (human clinical trials and NDA grant). They also report that investors value exploration and exploitation announcements by large firms and focused exploration activities for small firms. However, projects that are part of a strategic alliance partnership are not value-relevant. Himmelmann (2009) found, that for German biotechnology firms, patent grant and FDA approval announcements had a greater impact than announcements linked to individual trial commencement or successes.

Sarker and de Jong (2006) studied the four defined stages (FDA initial review, approvable drug, approval or rejection) of the FDA approval process and found positive announcements increased shareholder wealth, but the size of the impact depended on firm and drug characteristics.

Dedman et al. (2008) analysed drug-in-progress announcements for the UK biotechnology and pharmaceutical sector. They explored whether firms' announcements are concentrated at the exploration stage for smaller firms and the exploitation stage for larger, dominant firms. Their findings confirm the relative importance of drug approval announcements, with approvals significantly increasing shareholder wealth.

Following from these studies, I predict that investing in a portfolio containing stocks of the firms holding patents predicted to be approval-linked will earn abnormal stock returns for investors.

2.8 Conclusion

This chapter began by outlining the context and motivation for this study, discussing the unique features of the pharmaceutical industry, the drug discovery process and the key role that patenting plays in this environment. I reviewed some of the theoretical literature on the determinants of innovative success. Of these theories, the theory of absorptive capacity is

viewed as the most promising with its emphasis on the role of internal knowledge creation and external knowledge assimilation and exploitation. The literature suggests that firms that create their own in-house research using larger research teams and encouraging its researchers to publish academic papers, as well as patents, are better placed to produce more innovative ideas. Firms that harness external knowledge sources, as indicated by the citing of NPR in their patent documents or by encouraging co-authoring with universities or researchers employed by other firms are also more likely to experience innovative success. This success will be demonstrated both in terms of a superior ability to discover novel ideas and also to exploit these ideas commercially by transforming them into patents and eventually approved drug products.

The final two sections of the chapter discuss firstly the empirical literature on the influence of patent characteristics on patent quality and firm value and performance. I emphasise the role of certain patent characteristics in value creation and how they might increase the likelihood that a patent is linked to an approved drug. I next review the empirical literature on the effect of drug-related announcements on shareholder wealth, focussing on the positive influence of such announcements. In the next chapter, I utilise these literatures to formulate testable hypotheses.

Chapter 3 Development of Hypotheses

3.1 Introduction

The central proposal of this research is that firms must possess certain patent- and firm-level characteristics to succeed in the drug discovery process. For the purposes of this study, a drug is a prescription drug product that requires a doctor's purchase authorisation. The drug discovery process is the process from original idea through pre-clinical and clinical testing to final FDA approval. I investigate the relationship between these indicators and the outcomes from the product development process: i.e. the likelihood that a drug discovery obtains FDA regulatory approval.

In this chapter, testable hypotheses are developed in light of the previous literature. The arguments put forward in support of these hypotheses are anchored in the literature discussed in Chapter 2. To assist in the construction of these arguments and to create testable hypotheses in this study, summaries of the main findings from prior key literature are presented in the following subsections. They provide support to the construction of the hypotheses. The following subsections discuss the rationale for the hypotheses in relation to patent-level and firm-level variables.

In Section 3.2 hypotheses H_{1a} to H_{1e} are developed, to analyse how patent-level characteristics, available at the patent publication date, can act as indicators of patent quality and how this affects the likelihood of regulatory approval for a prescription drug product. In Section 3.3, hypotheses H_{2a} to H_{2d} are developed, to test the proposal that elements of a firm's absorptive capacity affect patent quality and so the likelihood of regulatory approval. Section 3.4 develops hypotheses H_{3a} and H_{3b} which investigate the effect of the quantity and timing of citations received from subsequent patents on the likelihood of approval. Finally, Section 3.5 develops hypotheses H_{4a} to H_{4d}, to analyse how certain firm characteristics may influence approval likelihood. A summary of this chapter is presented in Section 3.6.

3.2 Patent-Level Characteristics: Available at Patent Grant/Publication Date

In research-intensive industries, such as Pharmaceuticals, a large proportion of a firm's value is held in the form of intangible assets, e.g. patents. Patent quality and economic value is not directly observable in the early stages of a patent's life. Prior studies have examined a range of patent characteristics and utilised selected patent-based variables as proxies to enable the

measurement of patent quality and economic value. Certain patent-based variables exist within the patent document and are within the control of the firm, e.g. patent claims, becoming available once the patent application is published. These variables were some of the earliest utilised by researchers. Other patent-based variables, such as patent forward citation data, are available later in the patent's life and are often generated outwith the firm.

In studies of U.S. pharmaceutical firms, Narin et al., (1987) found that the citations made to scientific literature were positively correlated to the technological strength of a firm (in terms of patents produced and drug sales). Patel and Ward (2011) found a positive relationship between citations to a firm's own patents and firm performance. Lanjouw and Schankerman (1999 and 2004) found, in their study of U.S. firms, that patent claims were a good indicator of a patent's quality. A range of patent-based studies are discussed more fully in Chapter 2.

The following sections discuss the patent-based characteristics that I argue are the key indicators of a patent's quality, potential economic value and predictors of success in the drug approval process.

3.2.1 Number of Inventors: Research Team Size Indicator

Firms do not know the outcome of any particular research project at an early stage of development (see Pakes, 1986 and Moore, 2005). Continual development of new molecules creates a substantial number of patents, which assists the firm in maintaining its long-term competitiveness. A firm must review its portfolio of patents, evaluate their private value and decide which to progress to future development. Firms would favour those patents that are created by high-quality human capital and have an increased likelihood of gaining regulatory approval, so creating future economic value for the firm.

Inventors work with a number of colleagues when generating knowledge. This allows inventors to combine their knowledge with the complementary knowledge of co-inventors. Such collaboration provides an important mechanism for screening out inferior ideas early and increasing the probability of finding a solution to a research problem. Complementary pieces of knowledge can be pooled together from inventors with different knowledge backgrounds, giving access to a wider range of knowledge (Palomeras and Melero, 2010). The wider pool of knowledge available to larger research teams allows them to recombine knowledge in complex ways (Hoetker and Agarwal, 2007). I therefore propose that research teams that include multiple inventors will have access to differing knowledge sets and will

be more likely to develop new and novel ideas that in time will be developed into successful new drug products.

It follows from this argument that firms whose research teams are larger will produce more technologically important and valuable patents. This importance and value will be reflected in higher levels regulatory approval.

In view of the previous discussion, Hypothesis 1a of this study states that:

H_{1a}: Firms whose patents have a larger inventor team size (number of inventors listed on the front page of the patent) will have an increased likelihood of receiving regulatory drug approval from the FDA

The variables employed in hypothesis H_{1a} are:

- inventor team size, defined as the number of inventors listed on the front page of the patent

Further discussion of these variables is presented in Chapter 4.

3.2.2 Backward Citations

Patents contain a list of references to existing prior art on their front page, which outline the state of the art in that technological field. The USPTO requires, by law, that a patentee reference all relevant prior art[34]. These references indicate the foundations of the ideas embodied within the focal patent and are available at the patent publication date. The prior art referenced in patents is codified knowledge that can be divided into two categories - patent references and NPR. As the name suggests, patent references are references to earlier patents/patent applications, either domestic (U.S.) or foreign (non-U.S.). NPR are references to any published document other than earlier patents/patent applications. Where citations reference the knowledge created by a firm's researchers (in the form of patents or publications) this can indicate the internal knowledge base of the patent. They can also

[34] See the United States Code regarding patent law; Title 35 of the United States Code, Chapter 30, section 301.

indicate an external knowledge inflow into the firm from the patents of other firms or other published material.

Backward citations can be used to assess the invention's patentability, its technical novelty, and define the legitimacy of the claims stated within the patent application. Narin et al. (1987) found that, for the U.S. pharmaceutical industry, these citations could enhance the knowledge creation capabilities of a firm and increase the firm's knowledge stock. Where there are many backward citations to patents, the new invention represents an incremental improvement over existing technology and that the drug candidate being developed is entering a popular technological field, with increased competition and lower levels of potential future profits.

The number of patents cited in a drug patent application can be a proxy for the amount of patented knowledge used in the innovative activities that created the drug. Firms that utilise their own patents/in-house research and cite them in a patent show that internal knowledge has been utilised to develop the drug. Firms that utilise other firms' patents/external research publications show that external knowledge has been utilised to develop the drug. Citation sources from outwith the firm can encourage innovation and improve the firm's potential to develop new products. In both cases, firms whose research builds on previously patented knowledge, rather than non-patented research, are demonstrating that existing knowledge has been utilised to develop the drug. The use of previously patented knowledge can indicate that a firm is less creative and the inventions they produce may be less novel in nature. Harhoff and Reitzig (2004) found, in their study of the European biotechnology industry, that the inclusion of many backward citations to previous patents tended to indicate an invention that was more derivative in nature and that had lower economic value. In these cases, knowledge is flowing within already existing knowledge boundaries and this may be detrimental to innovative activity and restrict research creativity.

Who adds the citations will affect the strength of the relationship between the focal patent and the prior art it references. Citations are added by attorneys and examiners as well as by inventors and therefore do not indicate any technological influence on the invention. Jaffe et al. (2000) found that, from the results from their survey of U.S. inventors, around 50% of citations do not correspond to a technological relationship between the cited and citing invention. However, another survey response, from the same survey, indicated that around 70% of prior art referenced had some influence either on the invention or on the drafting of

the application. These results indicate that the relationship between focal patents and their cited prior art is not random. Citations can therefore provide an informative role in tracing the technological development of an invention and indicate potential economic value, although there is a degree of noise in the relationship.

Drug inventions need to be novel for a drug to obtain a NME classification from the FDA. I argue that the patents that create NMEs will have relatively low levels of citations to prior patents, as the firms that create them will tend to follow a research process that is more open to external scientific influences, especially those from academia. Successful firms need to be more open to new ideas and cannot be blinkered in their view of research possibilities. The use of citations to prior patents indicates a narrow approach to innovation. The focus of this study is therefore on backward citations to NPR.

3.2.2.1 Non-patent References

Prior to the 1980's the drug discovery process involved random screening, in this environment knowledge generated outside the firm was of limited usefulness. Advances in biomedical science greatly increased firms' scientific knowledge and firms started to apply deductive, scientific methods to the drug discovery process[35]. These changes meant that the ability to access and utilise scientific developments, generated outside the firm, became increasingly important to productive research. Branstetter and Ogura (2005) found that this change in the drug discovery process was reflected in a change in the nature of U.S. inventive activity from the mid-1980s onwards. They found that there was an increased emphasis on the use of the knowledge generated by university-based scientists, especially in the area of bio-pharmaceutical patents.

In the pharmaceutical industry, basic science[36] research has historically played an essential role in supporting the industry and NPR convey information on the proximity of the patented invention to this research[37]. The prevalence of NPR reflects the closeness of an invention to scientific knowledge, indicating an exploratory research path and can act as an indicator of a science-technology linkage. Gittelman and Kogut (2003), in their study of the U.S.

[35] The Economist, More Than a One-Drug Wonder (20th December,1986)

[36] Basic science is systematic study directed towards the greater knowledge or understanding of the fundamental aspects of phenomena without any focus on the commercialisation of this knowledge.

[37] The importance of academic research to U.S. industrial innovation is supported by survey data. See Mansfield (1991and 1995) and the Carnegie Mellon survey (1994).

biotechnology industry, found that NPR act as a proxy for the "science intensity" of a patent. They found that the higher the proportion of NPR in the prior art of a patent the more valuable it is. Inventions that build upon scientific publications are more likely to be influential in the future than inventions that do not build upon scientific publications.

Firms that are more closely linked to the scientific community have a stronger capability in exploiting this externally generated knowledge. NPR indicate that researchers are keeping up with developments in basic science and continuously improving their skill set and knowledge[38]. These citations can provide information on the research approach of the firm; that is how innovative and radical a firm is with its research philosophy. These attributes can enable firms to produce more pioneering inventions. Innovative firms will experience higher levels of research productivity, enjoy superior innovation performance (in particular with respect to innovations that are new to the market), and produce patents that are technologically important and of higher potential private economic value to the firm. Gambardella (1992), in his case studies of several large U.S. drug manufacturers, found that utilising basic science focuses a firm on the most promising research paths and can increase the productivity of in-house research.

Where a patent has few backward citations to prior patents and many citations to NPR, this can be an indicator of the novelty of the invention and sometimes the invention may be a completely new use of a scientific discovery. The invention may be an early entrant into a new field and be of potential technological importance. This can indicate the existence of a latent market with low levels of competition and the potential of future commercial opportunities and a source of potential private economic value for the firm. If the invention is opening a new field the firm may gain a first entrant advantage and have the potential to earn monopoly profits due to patent exclusivity rights.

I expect to find a relationship between a firm's ability to utilise external knowledge sources and success in the pharmaceutical innovation process. I explore this relationship by studying the characteristics of a firm's prior art citations detailed on the first page of the firm's patent documents. Firm patents that contain citations to NPR and to scientific journal articles in

[38] Henderson and Cockburn (1994) in their study of several major pharmaceutical firms found that NPR are an indicator of the expertise of a firm's research team(s). Scientific knowledge can also assist in identifying areas that can be exploited commercially and firms that have science-driven research also enjoy increased research productivity.

particular are for the purposes of this study assumed to experience closer connections to the basic science environment that exists outwith the boundaries of the firm. These NPR are assumed to be a measure of the science intensity of a firm's patents.[39] At the firm level, these references provide an indicator as to the ability of a firm to identify, absorb and commercially exploit this external knowledge. I argue that the ability of a firm to produce technologically novel patents will result in a higher likelihood of obtaining regulatory drug approval.

Hypothesis 1b of this study states that:

H_{1b}: Firms with higher levels of science intensity (patent prior art references to scientific publications) will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{1b} are:

- the non-patent references (NPR) contained on the front page of a patent

Further discussion of these variables is presented in Chapter 4.

3.2.3 Patent Claims

The number and content of a patent's claims define the legal rights conferred by that patent (OECD, 2009). Only items covered by the patent claims can be legally protected and enforced. They are under the control of the patentee but are subject to external validation by patent attorneys and the USPTO patent examiners. They are an indicator of the novelty and originality of an invention, an invention's potential commercial opportunities and an early indicator of potential economic value to the firm. A patent can only claim aspects of the invention that are not already covered by prior art (including other patents). Any aspect of the invention covered by patent-related prior art cannot be claimed by the focal patent without legal infringement.

Increased numbers of claims are an indicator that there are few existing products in that particular technological field. Patent claims can be either broad or narrow in the claims they

[39] Van Looy, Magerman and Debackere (2007) found this in their country-level study of the biotechnology industry. Whilst Gittleman and Kogut (2003) found support in their study of the U.S. biotechnology industry.

make for an invention. A firm can make more numerous or broader claims in an undeveloped field, where there are more commercial opportunities, without the fear of legal action as its claims are not restricted by the existence of prior art. This increased number of claims acts as an indicator of a larger inventive step. If a patent makes a large inventive step in an active market then there is an increased potential for future profits to flow from that invention. Nagaoka (2007) found that a patent with a higher number of claims had an increased chance of being commercialised. In a developed field where many firms are engaged in similar activities and an invention is more incremental in nature, building on prior patent-related knowledge, claims need to be less numerous and narrower to avoid infringement of other patents.

Claims indicate the technological breadth of the knowledge and capabilities embodied in the patent and indicate the pioneering nature of the patent. Claims link technological importance to potential economic value as they can assist in the commercial exploitation of a technological innovation. Lanjouw and Schankerman (2001 and 2004) found that, for industries where patenting is important, such as pharmaceuticals, there tend to be higher levels of claims as firms want to protect the potentially important commercial areas covered by these claims. Patents that contain more claims cost more to draft, file and prosecute. Albert, Avery, Narin and McAllister (1991) and Tong and Frame (1994) found that there was a positive and significant relationship between the number of claims and expected patent value.

Hypothesis 1c of this study states that:

H_{1c}: Firms whose patents contain higher numbers of claims will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{1c} are:

- the patent claims (those claims listed on the front page of the focal patent)

3.2.3.1 Claim Type

Patent claims are either independent or dependent. Independent claims are the most important type of claim; they describe in a concise manner the solution to a known technical problem and outline the degree of the inventive step. The number of independent claims indicates the size of the inventive step and the breadth of the patent. These claims stand

alone, without dependence on another claim and show the scope of patent protection more accurately. Dependent claims recite a connection to another claim. They add to the breadth of a patent and provide a fall-back option in legal disputes. Barney (2002) observed that the number of priority (independent) claims indicated the novelty and quality of a patent.

I expect patents with a larger number of independent claims to make a greater inventive step and are more likely to produce drug candidates that are NMEs. It follows from this argument that patents containing larger numbers of claims (especially if there are higher numbers of independent claims) are more technologically important and of larger potential economic value. This importance and value will lead to an increased likelihood that these patents will receive regulatory drug approval from the FDA. Therefore, Hypotheses 1c(i) and 1c(ii) of this study state that:

$H_{1c(i)}$: Firms whose patents contain higher numbers of independent claims will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis $H_{1c(i)}$ are:

- the independent claims of the patent (those independent claims listed at the end of the focal patent)

$H_{1c(ii)}$: Firms whose patents contain higher numbers of dependent claims will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis $H_{1c(ii)}$ are:

- the dependent claims of the patent (those dependent claims listed at the end of the focal patent)

Further discussion of these variables is presented in Chapter 4.

3.2.4 Backward Citation Lag Length

The backward citation lag[40] of a patent is the length of time between the application or grant year of the focal patent and the publication year of the previous patents or NPR it cites. A shorter lag length indicates the use of newer knowledge and so an invention that is more novel and radical in its nature. Nagaoka (2007) found, that for U.S. firms, a short citation lag relative to the prior patent literature indicates a significantly higher patent quality in terms of both patent citations received and the number of claims per patent. The longer the length of this lag, the older the science utilised which might imply that the invention is more incremental in nature. I expect that patents that utilise newer knowledge will produce more successful drug candidates. Following from this argument, firms whose patents cite prior art that is closer in time to the citing (focal) patent are using newer or more novel knowledge in their patents. This in turn increases the likelihood that they will produce patents that are more technologically important and of higher potential economic value to the firm. These patents will have an increased likelihood of successfully completing the drug discovery process.

In view of the previous discussion, Hypothesis 1d and 1e of this study state that:

H_{1d}: Firms whose patents have a shorter citation lag to non-patent prior art references (NPR) will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{1d} are:

- the citation lag length to non-patent prior art (length of time, in days, from the application year of the focal patent to the publication year of the latest non-patent prior art cited in the focal patent)

H_{1e}: Firms whose patents have a shorter citation lag to patent prior art references will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{1e} are:

[40] I assume that backward citations represent a flow of knowledge from the cited document to citing inventor.

- the citation lag length to patent prior art (length of time, in days, from the application date of the focal patent to the publication date of the latest patent prior art cited in the focal patent)

Further discussion of these variables is presented in Chapter 4.

3.3 Firm-Level Characteristics Available at Patent Grant/Publication Date: The Role of Absorptive Capacity

In the present study I view absorptive capacity as the "ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends", Cohen and Levinthal (1990, p. 1). The central argument of Cohen and Levinthal's approach is that, the technological responsiveness of an established firm is a function of its absorptive capacity. It enables the firm to respond to technological opportunities and to appropriate the technological knowledge made available through research spillovers by competitors, and/or organisations external to the industry (Cohen & Levinthal, 1989, 1990). I argue that a firm's ability to develop a new drug depends on its absorptive capacity.

The absorptive capacity of a firm provides it with the capability to create new inventions but also to successfully commercially exploit them. R&D projects, patents and number of scientific publications represent the firm's ability to organise its activities so that tacit knowledge can be successfully transformed into intellectual property. External networks (knowledge related to external parties such as strategic alliances and co-operations with other firms, co-authoring activity and relations with public research institutions) also play an important role in the creation and exploitation of new knowledge.

In the context of drug research projects, the firm has the task of adding new projects and abandoning existing ones. The firm must determine whether some projects are valuable enough to receive further investment and development, whilst others are no longer valuable enough to continue and must be abandoned. A firm's absorptive capacity can influence its choice of research project portfolio.

One important external source of knowledge applicable to industrial innovation, especially in the pharmaceutical industry, is university-generated science. Firms rely on basic science developments in biology and biochemistry, and many new drugs have their origins in discoveries at universities or government research facilities. Several researchers have

described industry use of university-based basic scientific research in the development of new products and processes (see for example, Narin and Olivastro, 1992). Firms in the "drugs and medical products" industry report that their innovations draw heavily from academic research and that new products and processes would have been delayed without access to this research (Mansfield, 1991, 1998). According to Narin et al., 1997, patents in the drugs and medicine category cite significantly more scientific publications than patents in other fields.

A firm's absorptive capacity is related to the ability of individuals in the organisation to assimilate, process and transform external knowledge flows. Some firms might possess high-quality human capital resources, e.g. highly qualified scientific researchers. However, without the necessary firm-level characteristics, such as research teams with good connections to academia, their inventions may never enter the final stages of the development process. Value is only created for the firm if these components interact in a particular way. I argue that the firm must possess a well-developed absorptive capacity to achieve success in the drug discovery process. This study aims to outline the key roles these components play in assisting a firm to achieve the goal of a successful NDA.

Firms that encourage external collaborations (joint research projects/alliances, co-authoring in academic journals etc.) are less likely to need to carry out additional research to absorb external knowledge. Their research teams will already have a good working knowledge of external research in their particular field.

3.3.1 Firm-level Publishing and Co-authoring Behaviour

In pharmaceutical firms, inventors do not work in isolation but collaborate in research teams, within and outwith the boundaries of the firm, where they pool their complementary knowledge to create quality research outputs. These skilled research teams are needed to drive the R&D process. For this industry, human capital can be viewed as a collective firm-level resource that improves the absorptive capacity of the firm.

The value that research teams create for the firm is best represented by the intellectual assets they create for the firm. These include patents, academic publications and new products. Human capital plays an important role in developing patents as inventors are the key contributors of the knowledge underlying patents (Liu, 2014). The teams' human capital or their knowledge, skills and capabilities determine whether the patents they have created hold

sufficient value for the firm to develop, maintain and transform into marketable products, so influencing firm performance. Gambardella (1992) demonstrated that trends in publishing behaviour were closely linked to research performance and that the number of publications was an approximate measure of a firm's in-house scientific capabilities[41].

Publications can act as an indicator of the level of scientific knowledge, capabilities and skills within the firm and are a measure of the firm's scientific excellence. A firm's researchers may also co-author publications with academic researchers. Gittelman and Kogut (2003) found that 70% of the papers published by employees of U.S. biotechnology firms were co-authored with academic scientists. This co-authoring behaviour improves the quality of the human capital of the firm-based researchers, as academic researchers possess tacit knowledge that can be learned from, absorbed and utilised in the firm's proprietary R&D. This sharing of ideas assists in the creation of radical innovations such as NMEs. It is important to a firm's innovative success to create conduits that allow access to this knowledge. Jong and Slavova (2014), in their study of UK biotechnology firms, found that co-authoring improved a firm's innovative performance as measured by products in development. Whilst, Almeida, Hohberger and Parada (2011) found that firm researchers who co-authored produced higher quality patents.

It follows from this argument that firms who undertake research alliances with external parties, allowing their research teams to publish scientific journal articles, co-authoring within and outwith the firm and funding such collaborative work are more likely to produce technologically important and valuable patents. Many prior studies have measured the quality of a firm's human capital by factors such as the qualifications and experience of researchers. However, I argue that it is the academic connectedness of a firm's researchers, their willingness to co-author with other firms and the quality of the research output (patents) that are the key metrics for assessing the quality of the human capital embodied in a firm's researchers. Firm researchers that possess high quality human capital will produce high quality patents that in turn will produce drug candidates that are more likely to attain regulatory approval from the FDA.

[41] Gittelman (2007) also found that publishing in quality academic journals was a good indicator of research quality and acted as a proxy for a firm's scientific expertise.

In view of the previous discussion, Hypotheses 2a, 2b, 2c and 2d of this study state that:

H_{2a}: Firms whose research teams publish scientific journal articles will produce patents of greater technological importance that will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{2a} are:

- scientific journal articles published by a firm's researchers

H_{2b}: Firms whose research teams co-author scientific journal articles with academic researchers will produce patents of greater technological importance that will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{2b} are:

- the percentage of scientific journal articles co-authored with academic researchers

H_{2c}: Firms whose research teams co-publish scientific journal articles with researchers from other firms will produce patents of greater technological importance that will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{2c} are:

- indicator that the firm's researchers co-author scientific journal articles with other firms

H_{2d}: Firms that fund research projects resulting in co-authored scientific journal articles, will produce patents of greater technological importance that will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{2d} are:

- indicator that the firm funds research projects that result in co-authored scientific journal articles

Further discussion of these variables is presented in Chapter 4.

3.4 Patent-Level Characteristics: Available after Patent Grant/Publication Date

In Section 3.2 I discussed some of the patent characteristics that are available at the patent grant/publication date. There are also other patent-based characteristics that are not available at this date but occur later in the patent's life e.g. patent renewal information. Several researchers (Schankerman and Pakes, 1986; Lanjouw, 1998 and Harhoff et al., 1999) found that patents that were renewed were more valuable in terms of technological importance and economic value. However, patent renewal data are not available in a timely manner, only being available 4, 8 and 12 years after a patent is granted. The relationship between patent renewal and the likelihood of drug approval is therefore not investigated in this thesis. Another of these characteristics are forward citations (citations received from subsequent patents).

3.4.1 Forward Citations

Forward citations occur throughout a patent's life and can be regarded as an indicator of a patent's innovativeness and quality. Pandit, Wasley and Zac (2011) viewed them as a measure of potential economic value.

Patents cited as prior art by many subsequent patents contain ideas that are recognised as being technologically or commercially important[42] and upon which numerous inventors build. Citations by other firms indicate that they think the technology covered by the focal patent has potential. These citations indicate the level of commercial interest and activity in that technological field and reveal the existence of downstream research activity in that field.

The correlation between the value of an invention that underlies a patent and the number of citations a patent receives by subsequent applications has been clearly established by prior literature. Numerous studies have illustrated the existence of a strong positive relationship between forward citations and the technological importance of patents (for example see work by Pakes, 1985; Deng et al., 1999; Hall et al., 2005; Gambardella et al., 2008 and Serrano, 2010). In his study of U.S. patents Nagaoka (2007) found that, for biotechnology and pharmaceutical patents, the number of forward citations received is a good indicator of the

[42] In their study of USPTO patents Carpenter et al. (1981) found that patents linked to technologically important products were likely to be cited more than twice as much by subsequent patents than patents that did not possess this link.

quality of that patent. In their study of patent auctions, Fischer and Leidinger (2014) found empirical support for the role of forward citations in patent valuation. There exists a strong theoretical and empirical foundation for utilising forward citations as a measure of patent quality and value.

A patent can have high levels of forward citations if it is an early entrant into what turns out to be a popular/crowded technology area where a radical invention is quickly adopted. It would be expected that the higher the technological quality of an invention, the greater the number of subsequent inventions that should build upon this focal invention. These citations operationalise the market potential of the focal patent, independent of the technological sophistication and quality of the underlying invention. A focal patent may receive many forward citations indicating its technological importance but this does not always follow that the invention is economically relevant. An invention can be technologically important, but not economically so, as economic success depends on other factors not covered by this research.

A contrary viewpoint is that if an invention is very radical in nature it may not receive a high number of forward citations. It may be that other firms working in the same technology field do not, initially, recognise the importance of the invention as it is too unfamiliar in nature. Scientific and technical advances that are related to the firm's core knowledge are more easily integrated into innovative activities than advances that are not related. More recent scientific advances are associated with uncertainty and risk. Firms may need to conduct additional research to apply patents that embody newer technical knowledge.

Prior studies have shown inconclusive results when patent indicators are related to the commercialisation of drug development. This could be due to factors unrelated to patent indicators such as, for example the disease the drug treats, potential number of patients and market competition. Wagner and Wakeman (2014), in their study of European pharmaceutical patents and their commercialisation, found that patent indicators such as the level of forward citations (over a five year window) and backward citations could be informative as to the speed of the development process. These indicators were also correlated with the outcomes of various stages of the development process.

On balance, I expect that patents that are perceived to be technologically or commercially important by other firms will receive higher levels of citations from subsequent patents. These patents will produce drug candidates of superior quality that will be more likely to

successfully complete the approval process This in turn increases the likelihood that these patents experience higher levels of regulatory drug approvals from the FDA.

In view of the previous discussion, Hypothesis 3a of this study states that:

H_{3a}: Firms whose patents receive citations from subsequent patents (forward citations), within a five-year time window, will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{3a} are:

- the citations the focal patent receives from subsequent patents (referred to as forward citations) within a five-year time window

3.4.2 Forward Citation Lag Length

The forward citation lag or lag length of citations received from subsequent patents may indicate the technological or commercial importance of the cited patent. It indicates the speed at which the content of the focal invention is recognised by other firms. The speed with which forward citations are received indicates the impact that the focal patent has had on other industry participants. If citations are received at a faster rate it is a signal that others have recognised the technological or economic importance of that patent. Nagaoka (2007), in his study of U.S. patents, found that for biotechnology and pharmaceutical patents the length of the forward citation lag is a good indicator of the quality of that patent.

It follows from this argument that firms whose patents receive citations from subsequent patents at a faster rate are viewed as more technologically or commercially important by other firms. This increases the likelihood that these patents will have a higher potential economic value to the patent holder and will experience higher levels of regulatory drug approvals from the FDA.

In view of the previous discussion, Hypothesis 3b of this study states that:

H_{3b}: Firms whose patents receive citations from subsequent patents (forward citations) at a faster rate will have an increased likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{3b} are:

- the forward citation lag for the focal patent (length of time in days, from the publication date of the cited patent to the application date of the citing patent)

Further discussion of these variables is presented in Chapter 4.

3.5 Other Firm-Level Characteristics

Previous literature has employed a variety of firm characteristics to explain the relationship between firm patenting, firm valuation and performance. However, their relationship with the likelihood of FDA regulatory drug approval is not the primary focus of this study. Prior patent-based studies (for example see Deng et al., 1999; Fabrizio, 2009; Pandit et al., 2011 and Liu, 2014) have recommended the use of certain control variables.

3.5.1 Firm Size

Firm size tends to vary widely in the pharmaceutical industry, ranging from the smallest start-up to the largest multi-national firm and this reflects in the large variability in patenting and commercial activity.

Previous research suggests that firm size may be important in explaining the heterogeneity in firm patenting. Larger firms may spend more on R&D and have larger research teams so tend to produce more research outputs (e.g. patents). The high costs of patenting have a larger impact on smaller firms so they may have to be more selective in the ideas they patent. Importantly, size matters because larger firms can apply the results of their R&D over a greater output than smaller firms can. Studies that have considered firm size in the context of patenting and firm performance have come to inconsistent conclusions about the role of firm size. DiMasi (2014) found that larger firms were less successful in obtaining U.S. regulatory approval but smaller firms had lower returns in terms of sales. In one prior study, Jensen (1987) employed actual new products as a dependent variable and reports that firm size has no effect when introduced as an independent variable. However, studies that have proxied for new products have shown firm size effects to be significant (Bound, Cummins, Griliches, Hall and Jaffe, 1982; Acs and Audretsch, 1989; Shan, Walker and Kogut, 1994; Nohria and Gulati, 1996 and Rothaermel and Hess, 2007). These inconclusive results from prior literature make it essential that this study must control for firm size.

Hypothesis 4a of this study states that:

H_{4a}: The size of a firm influences the likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{4a} are:

- firm size (measured by total assets)

Further discussion of these variables is presented in Chapter 4.

3.5.2 Firm Age

Prior literature has found a variety of findings of the effect of firm age on patenting. Some studies found that patent counts tend to vary with firm age (McNamara and Baden-Fuller, 2007 and Khoury and Pleggenkuhle-Miles, 2011). Barnett (1990) found that firm age had a negative effect on innovation levels within a firm as older firms can fail to keep up with recent advances in science and are more complacent in their outlook. Older firms are viewed as being less able to incorporate radical change in an efficient manner. However, Sorensen and Stuart (2000) found a positive effect, as increased experience seemed to enhance a firm's innovation ability.

Hypothesis 4b of this study states that:

H_{4b}: The age of a firm influences the likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{4b} are:

- firm age (measured in years)

Further discussion of these variables is presented in Chapter 4.

3.5.3 Firm Profitability

Profitability has been found to be an important firm-specific variable that affects patenting (Deng et al., 1999). It is one of the key characteristics that are associated with the extent of patenting and the performance of the firm. When profitability is high firms tend to spend

more on R&D and so patenting increases. If profitability is low, firms will have fewer resources to devote to R&D and producing patented inventions. Firms that are more profitable may have more patents and therefore a greater eventual number of approved drugs.

Hypothesis 4c of this study states that:

H_{4c}: If the firm experiences higher levels of profitability it will produce greater numbers of patents. This increases the likelihood of producing patents of greater technological importance that are more likely to receive regulatory drug approval from the FDA.

The variables employed in hypothesis H_{4c} are:

- firm return on assets

Further discussion of these variables is presented in Chapter 4.

3.5.4 Firm Prior Experience of the FDA Drug Approval Process

Prior experience of the FDA regulatory drug approval process can increase the likelihood that a firm will have a successful application outcome. Heemstra, de Vrueh, van Weely et al. (2008) found that experienced firms were more successful in obtaining European regulatory approval, with Heemstra, Leufkens, Rodgers et al. (2011) confirmed this finding for the U.S. regulatory process. However, Putzeist, Heemstra, Garcia et al. (2012), in their European-approved drug study, found that the most approvals belonged to firms with no prior experience. Firms with experience of the FDA approval process will already be familiar with the approval system and may have an existing working relationship established with FDA staff.

Hypothesis 4d of this study states that:

H_{4d}: If the firm possess prior experience of the FDA regulatory drug approval process this increases the likelihood of receiving regulatory drug approval from the FDA.

The variables employed in hypothesis H_{4d} are:

- indicator of firm experience of the FDA regulatory drug approval process

Further discussion of these variables is presented in Chapter 4.

All the hypotheses formulated in this chapter are summarised in Table 3.1.

3.6 Summary

This study hypothesises that regulatory approval success for a drug product is positively associated with certain characteristics of the patents cited in the drug approval application. Subsequent approval success is more likely if a firm has larger research teams, integrates more basic research and utilises more recent patented knowledge into its R&D programme. It is hypothesised that the likelihood of success increases if the firm also possesses certain firm-level characteristics that improve its absorptive capacity such as publishing scientific papers and co-authoring with academia.

The hypotheses in this chapter are tested using univariate and multivariate logistic regression techniques. In the next chapter, Chapter 4, the variables utilised in this study, the research design, and the methodology are presented. The explanation on the selection of variables and their construction is also discussed in Chapter 4.

Table 3.1 Summary of the Expected Impact of Patent and Firm Characteristic Variables on New Drug Application Success

Hypothesis	Variable	Nature of Variable	Measurements level and scale	New Drug Application Success
H_{1a}	Size of inventor team	Numerical	Continuous, Interval	Positive
H_{1b}	Non-patent references in backward citations to prior art	Numerical	Continuous, Interval	Positive
H_{1c}	Number of claims on front page of patent	Numerical	Continuous, Interval	Positive
$H_{1c(i)}$	Number of independent claims at the end of the patent	Numerical	Continuous, Interval	Positive
$H_{1c(ii)}$	Number of dependent claims at the end of the patent	Numerical	Continuous, Interval	Positive
H_{1d}	Patent reference backward citation lag length	Numerical	Continuous, Interval	Negative
H_{1e}	Non-patent reference backward citation lag length	Numerical	Continuous, Interval	Negative
H_{2a}	Employees of firm publish scientific journal articles	Numerical	Continuous, Nominal	Positive if articles published
H_{2b}	Percentage of published scientific journal articles co-authored with academia	Numerical	Continuous, Nominal	Positive if articles co-published
H_{2c}	Employees of firm publish scientific journal articles with employees of other firms	Categorical	Continuous, Nominal	Positive if articles published
H_{2d}	Firm funds projects that produce co-authored scientific journal articles	Categorical	Continuous, Nominal	Positive if funding exists
H_{3a}	Citations from subsequent patents (forward citations)	Numerical	Continuous, Interval	Positive
H_{3b}	Forward citation lag length	Numerical	Continuous, Interval	Negative
H_{4a}	Firm size	Numerical	Continuous, Interval	Positive
H_{4b}	Firm age	Numerical	Continuous, Interval	Positive
H_{4c}	Firm profitability	Numerical	Continuous, Interval	Positive
H_{4d}	Firm prior experience of the FDA drug approval process	Categorical	Continuous, Nominal	Positive if prior experience exists

Notes: All hypotheses are stated on a "ceteris paribus" basis, where all other conditions are equal. Categorical variables take the value 1 if the variable exists and 0 if it does not. The expected sign is the hypothesised relationship between FDA regulatory drug approval likelihood and the stated variable.

Chapter 4 Data Collection and Methodology

4.1 Introduction

I discussed the research questions, hypotheses and variables in Chapter 3. This chapter discusses the data utilised and the methodology applied in the empirical analysis. Variable definitions are stated, data sources outlined, and the data collection process discussed. Sample characteristics, explanations and details are provided for any variable transformations. The chapter also discusses the selection of a logit model as the main modelling tool for drug approval prediction. This chapter forms the basis of results presented in the following three chapters. Section 4.2 outlines the sample selection criteria together with sample construction and discusses the study period for which data are collected. Section 4.3 provides definitions and computations of the proposed variables. Section 4.4 presents the research design and methodology and Section 4.5 presents the empirical models, derived from the hypotheses that were stated in Chapter 3 and are constructed using the variables defined in Section 4.4. Section 4.6 outlines the methodology employed to evaluate model performance and Section 4.7 provides an overview of the techniques used to evaluate potential model-based portfolio investment performance. Section 4.8 provides a summary of the chapter.

4.2 Sample Construction

This section describes sample selection procedures regarding the sample period, sample firms, data collection for both dependent and independent variables, sample construction procedures and the data transformation techniques applied. The section is sub-divided into four subsections. Sections 4.2.1 to 4.2.4 discuss and justify the sample selection procedure for the sample firms and their patents.

4.2.1 Sample Period

The data for FDA drug approvals were extracted from the FDA's website via the Drugs@FDA database.[43] The extracted data cover the period from January 1974 to December 2014. The selection of the initial year for drug approval data was chosen as the start of the sample period because in 1974 the FDA began to code drugs using a classification

[43] Available at http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm?fuseaction=Reports.ReportsMenu, last accessed on 10th November 2017.

system that indicated how novel a drug was, this distinction was not available prior to this point. The selection of the initial year was also driven by the changes in the drug development process within the industry. From the 1970s, pharmaceutical firms began to introduce the findings from basic scientific research into their own R&D projects. Prior to this point, a trial and error approach was applied to the drug discovery process. The productivity and research effort of firms within the industry improved and the rate of major drug innovations increased (Grabowski and Vernon, 2000).

Most new drugs are patented prior to the submission of an NDA and are usually applied for in the pre-clinical or early clinical stages of drug development (e.g., Grabowski, 2008 and Olson, 2013). The digitisation of patent data commenced in the early 1970s, thus making the extraction of data from patent documents much easier and less time-consuming. To discover if an approved drug was linked to a U.S. patent, I searched for the approval in the FDA database Approved Drug Products with Therapeutic Equivalence Evaluations, commonly known as the Orange Book[44]. This lists all patents associated with a particular drug. These patent data can also be downloaded as a .csv file from the National Bureau of Economic Research (NBER) website[45].

4.2.2 Drug Approvals

For the purposes of this study, a drug is defined as a new prescription drug (a drug that is available only with written instructions from a doctor or dentist to a pharmacist) and has small molecule (non-biologic) formulation. Medicinal products that are excluded from the sample are those requiring a Biological License Application from the FDA (large molecule drugs) as the FDA only started approving notable numbers of these in the late 1980s with the increase in AIDS-related research. I also excluded vaccines, blood and blood products, gene therapies, cellular therapies, tissue and tissue products, diagnostic aids, new esters or salts. These exclusions are often based on naturally occurring substances that cannot be patented, for example radioactive elements in diagnostic aids. Also excluded from the sample are new active ingredients (coded chemical type 2 by the FDA), new drug dosage

[44] Available at http://www.accessdata.fda.gov/scripts/cder/ob/docs/queryno.cfm. The Orange Book identifies drug products approved based on safety and effectiveness by the FDA under the U.S. Federal Food, Drug, and Cosmetic Act and lists the drug-related patents and drug approval dates. Professor Bhaven Sampat also kindly provided a dataset containing all patents that are listed in the FDA Orange Book for all 1868 drugs approved from 1956 to 2012. This dataset contained approvals for both NMEs and other drug types. See footnote 46, page 101 of this thesis.

[45] Available at http://www.nber.org/data/fda-orange-book-data.html, last accessed on 10th November 2017.

forms (coded type 3) and new drug combinations (coded type 4). New drug formulations or manufacturer (coded type 5); new drug indications (coded type 6), drugs already marketed without an NDA (coded type 7), drugs switched to over-the-counter (OTC) marketing (coded type 8), a new indication or claim for a drug product that is being reviewed under a different NDA (coded type 9) and non-consolidated new drug indications (coded type 10) are also excluded. Following a similar approach by Sheck, Cox, Davis et al. (1984), Mattison, Trimble and Lasagna (1988), DiMasi (2001), DiMasi et al. (2003), only details of those approvals coded NDA chemical type 1 (indicating a NME) are collected. Drugs classified as NMEs for purposes of FDA review can often provide important new therapies for patients (source: FDA). NMEs can experience increased levels of uncertainty vis a vis their future success in the marketplace. Whilst they are more innovative and may carry more risk, they also have the potential to earn higher profits especially if they have no therapeutic equivalents. Totally new drugs have a potentially greater private economic value and can influence the financial performance of the firm's launching them.

The source of the approvals data was the Original New Drug Approvals by Month section of the Drugs@FDA database. These monthly reports list all applications approved for the first time during any month. Each month is manually selected and then a drug approval report request is submitted. A list of all approvals (coded type 1 to 10) for the chosen month is then displayed, this monthly list is downloaded into Excel; 492 months of data are downloaded month by month into Excel. Data gathered includes NDA number, approval date, drug name, active ingredient, firm name and the review type. Once all data are downloaded, they are refined to remove duplicate entries and to only include approvals that are classified as NME drugs (coded type 1), producing a sample of 930 NDAs. When recording approvals, the FDA links the approved drug to the firm that is the most recent holder of the right to market that drug which is not always the original applicant firm that submitted the NDA. To determine which firm submitted the original NDA the Drug Details section of the NDA record is accessed. The Approval History, Letters, Reviews, and Related Documents tab is selected and a PDF document of either the NDA review (containing the approval letter) or a PDF of the approval letter is displayed. The original application letter is then checked and the application date, applicant firm (if different to that on approval record) and disease the drug treats is recorded. The examination of these letters highlighted instances where the original applicant differs from the firm in receipt of FDA approval. PDF approval letters are available from March 1997. Not for every NDA in the sample has an available approval letter.

Other data sources are utilised to confirm both the original submission date and applicant firm for approvals where the approval letter is unavailable. These sources include the Harvard Project on U.S. Pharmaceutical Regulation (The FDA Project) and non-current issues of Approved Drug Products with Therapeutic Equivalence Evaluations (Orange Book). The FDA Project's NME Approval Time Data dataset[46] covered the period 1950 to 2006 and is utilised to identify missing NDA application dates and applicant firms. Data collected are also checked against data from several prior studies by Kenneth Kaitin and co-authors analysing FDA approvals for the period 1985 to 2001[47]. As a result of this triangulation process, 36 additional NDAs entered the sample, increasing its size to 966. They related to approved drugs that were no longer marketed and therefore no longer included in the Drugs@FDA monthly-approved drug lists. 124 approvals recorded by the FDA against non-U.S. firms were identified as the original applicant being a U.S. firm. All non-U.S. firms, non-quoted U.S. firms and joint ventures were removed from the sample. This created a sample of 587 NDAs coded as NMEs and applied for by US-listed firms that was then utilised to collect patent-related data. A final sample of 290 NDAs was created based on this patent data. The sample reduced in size due to missing patent data, mismatch of patent and FDA applicant details and patent classification[48] issues. A summary of the procedures used to remove NDAs and the effects this had on sample size are presented in Table 4.1.

4.2.3 Firm Selection

The sample includes those firms that have at least one NDA approved by the FDA during the period 1974 to 2014. The approvals of interest cover drugs that are completely new to the U.S. market and are designated as NME by the FDA. As the focus of this study is firms that are quoted on a U.S. stock market, I used the Compustat and CRSP databases to determine whether these approval firms were listed and if so on which markets they were quoted. Firm names were searched for and the Compustat datatype EXCHG (stock exchange code) was used to determine if the firm was listed on a U.S. market and that Compustat financial data are available for. Sample firms were also located within CRSP and share price

[46] These data are available in .csv file format at http://people.hmdc.harvard.edu/~dcarpent/fdadatabasic.csv.
[47] See Kaitin, Richard and Lasagna (1987), Kaitin, DiCerbo and Lasagna (1991), Kaitin, Manocchia, Seibring and Lasagna (1994), Kaitin and Healy (2000) and Kaitin and Cairns (2003).
[48] Patents are classified in the U.S. by a system using a 3 digit class to describe similar groupings of patent art (See glossary of terms in Appendix 2 for further detail).

Table 4.1 U.S.-Listed Drug Approvals: Sample Selection Procedure

Initial Sample: All New Molecular Entity Drug Approvals, Sample Size 966 (930 approvals source: Drugs@FDA and 36 approvals source: FDA Project and Kaitin et al. Articles)	
Sample Procedure	**Approvals Removed**
Remove approvals issued to non-U.S. firms	331
Remove approvals issued to non-quoted/private U.S. firms	41
Remove approvals issued to joint ventures	7
Remove approvals issued to U.S.-quoted Firms: FDA applicant does not match patent assignee	159
Remove approvals for Diagnostic Aids	51
Remove approvals issued to U.S.-quoted Firms: No USPTO patent exists	32
Remove approvals with application dates prior to 1974	22
Remove approvals where patents are filed after drug approval	18
Remove approvals where patent(s) associated with the approval have expired	12
Remove approvals where patents are not within the selected 17 U.S. pharmaceutical patent classifications	3
Final Sample of Approvals issued to U.S.-quoted Firms	**290**

Notes: This table reports the steps for the initial matching of FDA drug approval applicant data to patent assignee data. A total of 379 patents were removed from the sample of 966 approvals, as the applicant was either a non-U.S./non- quoted firm or a joint venture. A further 159 approvals were removed as the patent linked to the approval was not held by the applicant. 51 approvals were removed from the sample as they were for diagnostic aids. In 44 cases, either the patent linked to the approval had expired or there was no patent listed for the approval. 22 approvals were removed as their application dates were prior to 1974. In addition, 3 approvals were linked to patents that had classification codes outwith the 17 U.S. patent classifications selected for this study. This produced a final sample of 290 approvals.

and return data availability ascertained. The firms in the sample were listed on the New York Stock Exchange (NYSE), National Association of Securities Dealers Automated Quotations (NASDAQ) exchange or American Stock Exchange (AMEX). Care was taken to track changes created by firm name changes and merger activity, to ensure collection of the correct firm-specific Compustat and CRSP data. Non-quoted and non-U.S. firms are excluded from the sample as these firms may have different financial structures compared to U.S.-quoted firms and are subject to differing regional and international laws, while

unlisted firms may have differing factors influencing their research strategy and patenting activity. U.S.-quoted firms are also more familiar with the U.S. (FDA) regulatory approval system. There may also be a possible linkage between the market of listing and firm size. Larger, more mature firms tend to be quoted on the NYSE, whilst smaller, newer firms are listed on the NASDAQ, AMEX etc. Ideally, non-quoted firms would be included in the sample, but I do not have access to a data source that provides data for U.S. non-quoted firms, so preventing their inclusion in the sample.

I exclude NDAs where the firm submitting the NDA did not submit the NDA-linked patent application. These firms either licensed or acquired the NDA-linked patent, developing this patent and taking the resulting drug candidate through the clinical trial process to NDA submission. This approach is similar to that of DiMasi et al. (1991), DiMasi, Grabowski and Vernon (1995), Schwartz (2004) and DiMasi (2014), which is discussed in greater detail in Section 4.2.4.1. This excludes 159 approvals, 39 of which have U.S.-listed applicants. As a result 29 U.S.-listed firms are excluded from the sample.

The exclusion of these firms reduces the sample size and may restrict the results of this study to those firms who create self-originated NMEs (firms who create their own patents and are involved in the drug discovery and development process from start to finish). My results may not be applicable to firms whose research strategy is to license in or acquire patents. This decision was motivated by my research focus. I investigate the influence of patent characteristics on approval outcomes but also whether firm-level publishing characteristics play a key role. If a firm does not create the NDA-linked patent then it would be impossible to measure the influence a firm's publishing behaviour may have on the creation of patents with an increased likelihood of approval.

96 firms were identified, from an original sample of 246 unique applicants. However, when creating the Panel dataset and running the conditional logistic regression (discussed later in Section 4.5) 15 firms were removed from the sample as they only had patents for approved drugs with no match in the non-approval patent sample that met the selection criteria. The final sample size consisted of 81 firms (listed in Appendix 3) and 290 FDA drug approvals.

4.2.4 Patent Sample

The unit of analysis in this study is patents granted by the USPTO to the firms in the sample. This section describes the procedures regarding the creation of this patent sample. The

section is further divided into two subsections. Section 4.2.4.1 presents the criteria for selecting patents associated with FDA approved drugs, while section 4.2.4.2 discusses the selection procedures for the sample firm's patents that are not associated with an approved drug.

4.2.4.1 Sample for Patents Related to Approved Drugs

It is necessary to construct a patent sample for patents that are linked to the sample of drug approvals. This is done by utilising various data resources to ensure that as many approval-linked patents are included in the sample as possible. Patent numbers associated with drug approvals were extracted from the Sampat Orange Book patent dataset[49] for the period from 1974 onwards. For the period from 2012, these numbers were collected from the FDA's online version of the Orange Book by using the Orange Book Search facility and searching individually on each of the approval numbers. This search lists any unexpired U.S. patents that are linked to an approved drug. To ensure the sample being constructed contains all relevant patents, the active ingredient(s) for the approvals in the sample are also searched in The Merck Index[50]. If any U.S. patents were found that were linked to the active ingredient of the drugs in the sample, but were not listed in the Orange Book, they were recorded along with details of the firm holding the patent.

Some of the drugs in the sample have no patents associated to them. This occurs when a drug containing a certain active ingredient is approved for the first time, but the active ingredient is not itself patentable. This can be because the active ingredient is a naturally occurring substance, such as insulin, and cannot therefore be protected by a patent. In addition, some new drugs are un-patentable because their active ingredient has been disclosed in a journal article or a previous patent application. This means the drug ingredient does not meet the patentability criteria of novelty and non-obviousness. If the patent(s) associated with the drug have expired prior to the submission of the NDA into the FDA approval process the patent is excluded from the sample. This approach is in line with prior literature (e.g.

[49] Provided electronically by Bhaven Sampat, associate professor in the Department of Health Policy and Management at Columbia University's Mailman School of Public Health, on 30th September 2013.
[50] The Merck Index is an encyclopaedia of chemicals, drugs and biologicals and contains monographs on single substances or groups of related compounds. It also lists the patents, from all jurisdictions, that relate to a particular chemical compound and the firm that was granted the patent.

Magazzini, Pammolli and Riccaboni, 2008). Non-USPTO patents protecting a drug's active ingredients are also excluded from the sample.

Finally, there were cases where there was a valid USPTO patent for a particular NDA but the patent assignee (patent owner) was not the firm that submitted the NDA. There can be many reasons why this may be the case. The patent may be held by a small firm, research institution or university that does not have the resources (financial and/or scientific expertise) to take the drug candidate through late-stage clinical trials prior to FDA approval. The patent is often then licensed to the applicant firm. According to DiMasi et al. (1991) this usually occurs after the candidate has completed the pre-clinical stage of its development process. In addition, firms can merge, or a firm changes its research focus causing the firm to sell off some of its patent portfolio.

Following a similar approach taken by DiMasi, Grabowski and Vernon (1995), Schwartz (2004) and DiMasi (2014) the sample only includes self-originated drugs. These are drugs discovered, patented and developed by the same firm. This type of drug and its associated patent(s) is the focus of this study. I am interested in the characteristics of the research activity initiated within U.S. firms, their publishing behaviour and whether this can be utilised to predict FDA approval for the drugs created by these firms. The sample does not include drugs based on patents licensed or acquired from firms or institutions that differ from the firm submitting the NDA for FDA approval, as the research strategy of these two groups may be significantly different and the publication behaviour of the licensee could not be assumed to have impact on the research upon which the patent is based.

This matching procedure results in a final sample where an NDA is matched to at least one valid USPTO patent. If the patent application date is after the NDA submission date then the patent is removed from the sample. This is necessary as this study focuses on patents that are involved in the discovery and development of drugs and not those patents that are merely an adaptation or improvement to an already existing innovation (Hemphill and Sampat, 2011).

Some firms in the sample applied for patents prior to their Initial Public Offering (IPO) date, and financial data are unavailable on Compustat. If no alternative patents are available for the NDA, then the patent and linked NDA are removed from the sample. This produced an initial sample of 1697 patents associated with an FDA approved drug. A summary of the

procedures used to remove patents and the effects this had on sample size are presented in Table 4.2.

Table 4.2 Approved Drug Patents (966 FDA-approved Drugs): Sample Selection Procedure

Procedure	Number of Patents in Sample
USPTO patents with a link to FDA-approved drugs: 966 FDA-approved drugs	1,697
Remove approvals for which there is no patent linkage: 32 FDA-approved drugs and 0 patents removed. Leaving 934 FDA-approved drugs.	1,697
Remove patents for approvals issued to non-U.S. firms, non-quoted U.S. firms or Joint Ventures: 379 FDA-approved drugs and 93 patents removed. Leaving 555 FDA-approved drugs.	1,604
Remove patents for approvals issued to U.S.-quoted firms where the FDA applicant firm does not match the patent assignee: 159 FDA-approved drugs and 334 patents removed. Leaving 396 FDA-approved drugs.	1,270
Remove patents for approvals for diagnostic aids: 51 FDA-approved drugs and 107 patents removed. Leaving 345 FDA-approved drugs.	1,163
Remove patents for approvals that have application dates prior to 1974: 22 FDA-approved drugs and 38 patents removed. Leaving 323 FDA-approved drugs.	1,125
Remove patents for approvals where the patents are filed after the approval date: 18 FDA-approved drugs and 248 patents removed. Leaving 305 FDA-approved drugs.	877
Remove patents for approvals where the patents associated with the approval have expired: 12 FDA-approved drugs and 44 patents removed. Leaving 293 FDA-approved drugs.	833
Remove patents for approvals where there are patent classification and Compustat data issues: 3 FDA-approved drugs and 201 patents removed. Leaving 290 FDA-approved drugs.	632
Final Sample: Patents Associated with 290 FDA Approved Drugs	**632**

Notes: This table reports the steps for determining the final approval-linked patent sample size. Of the original 966 FDA-approved drugs in the sample, 32 had no patent linkage. The remaining 934 approved drugs were linked to 1,697 USPTO patents. 93 patents were removed from this sample, as the applicant was either a non-US/non-quoted firm or a joint venture. A further 334 patents were removed as the patent linked to the approval was not held by the applicant. 107 patents were removed from the sample as they were for diagnostic aids. In 38 cases, the approval to which the patent was linked had a pre-1974 application date. 248 patents were

removed as the patent was filed after the drug approval date. 44 patents were removed due to expiration. In addition, 3 approvals were linked to patents that had classification codes outwith the 17 U.S. patent classifications selected for this study. This produced a final sample of 290 approvals associated with 632 patents

4.2.4.2 Sample for Patents Not Related to Approved Drugs

Once the FDA approval linked patent sub-sample is constructed a sub-sample is constructed containing USPTO patents that are not linked to an FDA-approved drug. This is necessary to test the hypotheses outlined within Chapter 3. The starting point for the construction of this comparative sample is the NBER Patent Data Project. The dataset contains data on over three million patents and these data are downloaded into Stata for further analysis.

Whilst constructing the sample of patents linked to drug approvals it is noted that for FDA drug approvals approved in 2014 (the latest in the sample) the oldest patents linked to these approvals were granted in 2003. An assumption is made that patents granted up to 31st December 2002 can be considered unlikely to be linked to an FDA drug approved after 2014, unless they are listed as drug-linked in an edition of the FDA Orange Book, in the Merck Index or in the Harvard FDA Project dataset or they are subject to a patent term extension. This is in line with drug development time scales estimated by DiMasi et al. (1991). All patents with grant dates from January 2003 were removed from the NBER patent sample to exclude patents where there is the possibility that they may still be associated with an approved drug. Also removed are patents held by non-U.S. firms and other forms of organisations e.g. universities and individuals. Only U.S. corporate patent assignees (coded 2[51] in the dataset) are retained and this reduces the data sample considerably. These patents cover a range of industries but in this study only patents on pharmaceutical products are considered relevant to the analysis. The patent sample is therefore further refined to include only pharmaceutical patents. Relevant pharmaceutical-related patents are identified by reference to their U.S. patent classification (See Appendix 4 for a full list of classes and their descriptions). In order to do this the patents matched to FDA approvals are re-examined and a note taken of their patent classification codes, seventeen individual U.S. patent classifications are identified. In order to compare similar patents, my sample of patents is limited to these seventeen 3-digit patent classes as these are the most closely associated with the patents for approved drugs. For example, prior studies such as Nairn et al. (1987) used a

[51] The assignee codes are 1 unassigned, 2 assigned to a U.S. non-government organisation, 3 assigned to a non-U.S. non-government organisation, 4 assigned to a U.S. individual, 5 assigned to a non-US individual, 6 assigned to the U.S. (Federal) Government and 7 assigned to a non-U.S. government.

defined list of U.S. patent classes to define pharmaceutical patents. Lanjouw and Cockburn (2001) used a similar approach but utilised a set of international patent classes and Magazzini, Pammolli and Riccaboni (2012) used both class types. These U.S. patent classes are applied as a filter to identify pharmaceutical patents within the non-approval patent data sample. An assumption is made that all patent classes that do not match this group are not pharmaceutical patents and are excluded from the sample.

Patents are also checked against U.S. active ingredient patents in the Merck Index, patents listed in the FDA Orange Book and also the patents for which a term extension has been granted under the Price Competition and Patent Restoration Act of 1984 (the Hatch-Waxman Act)[52]. If a patent has a term extension under this act, then it must be associated with an approved drug and therefore excluded from the sample. Checking against term extension data and Orange Book data also allows the exclusion of patents linked to approved drugs that are not coded as NMEs and so excluding patents linked to other drug types.

A patent has the name of the organisation to which the patent is assigned on its front page; this assignee can be a subsidiary or a separate division of a larger firm. In addition, over time, firms can change their names, merge or cease trading. The pharmaceutical industry in particular is a very merger-intensive industry with a large number of corporate structural changes. The NBER Patent Data Project assists in overcoming this issue as it also contains data linking patent assignee names to corporate entities and Compustat codes. Using Stata, these data and the patent data sample can be merged to allow patents to be matched to Compustat codes, so enabling access to the financial data required in this study. Although the NBER Patent Data Project cleaned up many of the assignee names and linked them to Compustat codes, further work is required to identify all subsidiaries and code them correctly to reflect all merger, acquisition and divestiture activity occurring up to 31st December 2014. This process required a large amount of manual effort to ensure that the corporate structures reflected in the sample remain accurate through time. There are many variants in the spelling of firm (patent assignee) names and some firms are listed as U.S. corporations but do not have a listing on a U.S. stock exchange. The patent sample is manually checked again to

[52] The act provides that a novel drug can receive an extension term equal to one-half of the time the drug spends in human clinical trials through to NDA submission. The complete period of NDA review can also be added to the extension, if this time period is less than or equal to 5 years. A list of patent term extensions are available at http://www.uspto.gov/patent/laws-and-regulations/patent-term-extension/patent-terms-extended-under-35-usc-156.

ensure only U.S.-listed firms classified under U.S. SIC codes 2833, 2834, 2835 and 2836[53] are included in the sample. Any duplicate patents or patents assigned to joint ventures are removed. Finally, any patents owned by firms with no Compustat data were removed in addition to patents with other patent-related data issues. This leaves an initial sample of 19,104 U.S. patents that have no link to an FDA-approved drug.

This study assumes that the probability of a patent being linked to an approved drug can be modelled as a logit function. The logit function classifies the patent as approval-linked or non-approval-linked based on its conditional approval probability. This classification is conducted by computing the odds of the patent being linked to an approved drug conditional upon its observed characteristics and its assignees publishing characteristics. A conditional logistic regression approach is therefore used to test the hypothesised relationships outlined in Chapter 3. For every firm with an approval-linked patent the patent data are organised in Stata so that the firm also has a patent not associated with an approved drug. That is, the firm holds both successful and unsuccessful patents otherwise Stata drops these firms from the regressions. Note that the sample comprises "clinical" failures and successes and not commercial ones. I code a drug patent as successful, if it passes through all the stages of drug development and obtains an FDA approval, irrespective of its commercial success. This patent list is then used to download patent-level data. In the process of constructing a panel dataset before running regressions, I found that certain of the sample firms did not have any matching unsuccessful patents. I therefore had to collect a further sample of non-approval-linked patents to create an unbalanced sample. A summary of the procedures used to remove patents and the effects this had on sample size are presented in Table 4.3.

[53] SIC code 2833 covers: medicinal chemicals and botanical products, 2834 covers pharmaceutical preparations, 2835 covers in vitro and in vivo diagnostic substances and 2836 covers biological products, except diagnostic substances. All these SIC codes belong to the industry group 283: drugs.

Table 4.3 Sample Selection Procedure for Patents Not Related to Approved Drugs

Procedure Applied	Number of Patents
Initial Sample: NBER Patent Data Project 1976 to 2006: Patents available (No Duplicates)	3,210,361
Remove patents assigned to non-U.S. entities, individuals and the U.S. Government (515,578 patents removed)	2,694,783
Remove patents in patent classifications out-with the selected 17 U.S. patent classes (2,590,098 patents removed)	104,685
Remove patents without a match to a Compustat code (36,525 patents removed)	68,160
Remove patents linked to non-U.S. firms (27,587 patents removed)	40,573
Remove duplicate patents and patents listed in the approval-related sample (11,588 patents removed)	28,985
Remove patents granted after 31st December 2002 (5,152 patents removed)	23,833
Remove patents assigned to firms without SIC code 2833, 2834, 2835 or 2836 (2,945 patents removed)	20,888
Remove patents assigned to joint ventures (872 patents removed)	20,016
Remove patents with missing patent and Compustat data (912 patents removed)	19,104
Additional patents to allow a logit regression to be conducted (4,252 patents added)	23,356
Remove duplicate patents based on firm and grant date (11,031 patents removed)	12,325
Final Sample: Patents Not Associated with FDA Approved Drugs	**12,325**

Notes: This table reports the steps for determining the final non-approval-linked patent sample size. Of the original 3,210,361 patents in the sample, 515,578 had links to non-U.S. entities, individuals and the U.S. Government. 2,590,098 patents were removed from this sample, as they were not within the selected 17 U.S. patent classes. A further 36,525 patents were removed as the firm holding the patent had no Compustat code. 27,587 patents were removed from the sample as they were held by non-U.S. firms. 11,588 patents were removed, as they were duplicate patents or listed in the approval-related sample. 5,152 patents were removed as their grant dates were after 31st December 2002. 2,945 patents were removed as they were assigned to firms without SIC code 2833, 2834, 2835 or 2836. 872 patents were removed as the patents were assigned to joint ventures. 912 patents were removed due missing patent and Compustat data. In addition, 4,252 patents were added to the sample to allow a logit regression to be conducted. 11,031 duplicate patents were removed based on firm and grant date, leaving a final sample size of 12,325 non-approval linked patents.

4.3 Data for Patents and Firm Publishing

All patent data are collected from the Google Patents[54] database and the USPTO. The data sourced from patent documents directly include the count of the number of citations to non-patent related references (titled other publications on the patent document), the date of cited patents, the publication year of NPR, the number of inventors and the number of patent claims. The data collected include the total number of claims, number of dependent claims and the number of independent claims. Other patent variables were constructed from patent-based information. These were backward citation lag lengths for both patent and NPR citations. The number of citations a patent received from other subsequent patents and the time taken for these citations to be received (forward-citation lag length) were also calculated. A detailed description of all these variables and any calculations involved in their construction are discussed later in this chapter.

The distributions of the patent variables were viewed graphically using histograms, descriptive statistics analysed for all variables and they show that the data are significantly skewed for many of these variables. Therefore, I use the natural log transformation to allow the distribution to become closer to normal. Following a number of studies such as Lanjouw and Schankerman (2004), Lerner (2010), Singh and Fleming (2010) and Gambardella, Harhoff and Verspagen (2011) amongst others, in cases where the variable can have zero values one is added to the value of this variable prior to the natural logarithmic transformation so that the smallest variable value is one.

The study is also interested in testing whether the publication behaviour of a firm can act as a measure of the firm's research quality and in so doing be an indicator of the likelihood of success in the drug approval process. The publications data for firms in the sample are collected from the Web of Science database[55]. There are three publications variables for which data are collected: total firm publications, publications co-authored between the firm and academia and publications co-authored between the firm and another firm. I also identify whether the firms in the sample are involved in any research alliances with either academia or other firms. Firm publications are identified by searching for the firm name in the author-

[54] Available at https https://patents.google.com/. All U.S. patent documents available on Google Patents originate from the United States Patent and Trademark Office (USPTO).
[55] Available at http://apps.webofknowledge.com.ezproxy.lib.gla.ac.uk/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=S1wsTQq6t8oAyCPXE9q&preferencesSaved=.

address field, selecting the period 1974 to 2014, limiting the research domain type to science and technology and limiting the document type to article. This procedure generates the total number of journal articles published by the firm. University co-authored publications are identified by using the publications' search parameters but with the additional search term "university" in the author-address field. This procedure generates the total number of journal article publications co-authored between the firm and at least one university researcher. These measures are then averaged over the sample period. Publications co-authored with other firms are identified by using the publications search parameters but with the additional search terms "Inc." or "corporation" in the author-address field. These publications may reflect co-authorship with one firm or multiple firms. This measure is coded 1 where co-authorship occurs and 0 otherwise. Alliances are identified by observing if the firm is listed in an article's funding sources so indicating that it funds the project the journal article discusses. This measure is coded 1 where funding occurs and 0 otherwise.

4.4 Variable Description, Measurement and Model Specification

This section focuses on justifying the methodology used in this study, defining variables and specifying the models used. The section is further divided into three subsections: Sub-section 4.4.1 defines the dependent, independent and control variables for the examination of the impact of patent and firm characteristics on the likelihood of FDA drug approval. Sub-section 4.4.2 provides detail on these variables and Sub-section 4.4.3 provides an explanation of the impact these characteristics have on FDA drug approval.

4.4.1 Influence of Patent and Firm Characteristics on FDA Drug Approval

The hypotheses H_{1a} to H_{1e} presented in Section 3.2 of Chapter 3 jointly postulate a relationship between the characteristics of a firm's patents and the likelihood of FDA drug approval. Firms are more likely to be associated with an FDA approved drug where their patents are created by larger inventor teams, which cite NPR, utilise newer prior art and possess higher numbers of claims, both independent and dependent. Hypotheses H_{2a} to H_{2d} presented in Section 3.3 of Chapter 3 jointly postulate a relationship between the characteristics of a firm's publication behaviour and the likelihood of FDA drug approval. Firms that publish scientific journal articles, co-author articles with university scientists or staff from other firms and fund research project alliances are also more likely to be associated with an FDA approved drug.

4.4.1.1 Dependent Variable

This study explores the effect of a firm's patent characteristics, publication output and collaboration behaviour on the likelihood of FDA drug regulatory approval. The empirical models detailed later in this chapter are constructed with the objective of estimating the effects of these variables on the likelihood of successful drug approval outcomes. Accordingly, the dependent variable to be estimated in this study is:

The likelihood of the approval of a New Drug Application (NDA)

The dependent variable is a binary variable that indicates whether the independent variables in the model are associated to an FDA approved drug (coded 1) or are not associated with an approved drug (coded 0).

This study views the approval of a drug as a measure of success for a firm, in that the firm is exploiting its patents and publishing/collaboration behaviour to create a marketable new drug.

4.4.1.2 Independent Variables: Patent Characteristics

These are pieces of information contained either on the front pages of a patent document (non-patent and other patent-related prior art citations), inventors, assignee (owner of the patent) or on the final pages (claims) and are available at the patent publication date. This study does not analyse data from the abstract, drawings or the description sections of patent documents. There may be an issue of possible endogeneity with some of the patent-related variables as a patent can be drafted in a manner to infer value. Patent characteristics may not have an equal impact, some having more importance than others. Patents are not a homogeneous group, being heterogeneous in their characteristics. As the model-estimation process employs various characteristics it takes account of this heterogeneity, so strengthening the potential explanatory power of any models formulated from this process. This study seeks to investigate and explain the variation that occurs and what impact this may have on the likelihood of drug approval.

Patents are only granted if they meet the three patentability criteria: novelty, non-

obviousness and usefulness[56]. To explain why an invention is new, inventors cite prior art, such as earlier granted patents, patent applications and journal articles.

Independent Patent Variable 1: Citations to Non-patent References (NPR)

NPR are one of the independent variables that are included in the empirical analyses of the models in this study (model specifications are discussed later in this chapter). This variable indicates the number of references to non-patent material within the prior art listed on the front page of a patent and are an indicator of the patent's link to scientific knowledge. Here a reference to the scientific literature is taken to be an indication that the inventor (firm-based scientist) utilised scientific knowledge in the process of invention or was influenced by its existence. Previous literature has shown that these references increase forward citations (a widely accepted measure of patent quality) and increase the economic value of patents. Following a number of studies such as Narin et al. (1997)[57], Gittelman and Kogut (2003) and Arts and Veugelers (2012) amongst others, these references are viewed as an indicator of patent innovativeness and how closely it is related to basic scientific research. This measure is represented by the variable NPR_{it} in the models outlined later in this chapter.

Independent Patent Variable 2: Research Team Size

In the pharmaceutical industry, inventors work with a number of colleagues to generate knowledge. Such collaboration provides an important mechanism for early screening of inferior ideas and pooling complementary knowledge. Larger research teams allow inventors to pool and recombine knowledge from a wider scope of knowledge. Firms whose research teams are larger will be more capable of producing more technologically important and valuable patents. The size of the research team is proxied by the number of inventors listed on the front page of the patent and is represented by the variable $INVENTORS_{it}$. The number of inventors has also been used as a proxy for the cost of the research project; the more resources involved, the more research intensive and expensive the project is.

[56] At least one aspect of the patent must be new, the patent has to involve an inventive step and the invention it covers must be capable of use and provide an identifiable benefit.
[57] Carpenter, Cooper and Narin (1980) found that patents in highly scientific areas contain as many citations to the scientific literature as research papers. See also Fleming and Sorenson (2004) and Carpenter and Narin (1983).

Independent Patent Variable 3: Non-Patent Reference and Patent Citation Lag Length

I am interested in these two types of backward citation lag as they measure the age of prior art the patent is built upon. Patents in classes that experience more rapid technological change will tend to cite more recent prior art. The time lag between the cited patents or NPR and the citing patent represents the speed with which prior knowledge is utilised by subsequent patents, with shorter time lags indicating a faster pace of knowledge exploitation. Deng et al. (1999) viewed the patent citation lag measure as a measure of the technology cycle time and proxies for how quickly firms are innovating. It indicates the age of the citation, the age or maturity of the science the focal invention is based on and the speed at which prior knowledge is utilised in the new invention. The calculation of the non-patent reference citation lag follows an approach by Fabrizio (2009) and represents the time difference between the publication year of the non-patent reference cited on the front page of the patent and the application year of the citing patent. These are the only reliable dates available and the lag is then expressed in days. A shorter lag length indicates the use of newer knowledge and so therefore an invention that is more novel and radical in its nature. It is represented by the variable NCL_{it}. The backward citation lag for patents is calculated from the application date of the citing patent to the grant date (publication date for patents applied for after 29th November 2000) of the cited patent, expressed in days and is represented by the variable PCL_{it}. Both measures include citations to a firm's own patents/publications and external citations.

Independent Patent Variable 4: Patent Claims

Patent claims are detailed towards the end of the patent document, but their total number is listed on the front page. These claims create the legal boundaries of the exclusive rights of a patent owner. Only items covered by the patent claims can be legally protected and enforced. They are under the control of the patentee but are subject to external validation by patent attorneys and the USPTO patent examiners. They are an indicator of the novelty and originality of an invention, an invention's potential commercial opportunities and an early indicator of the potential private economic value of the patent to the firm. Novel ideas can make many claims if they open a new productive area. Nagaoka (2007) found that a patent

with higher numbers of claims had an increased chance of being commercialised[58]. All patent claims are either independent or dependent. Following an approach by Allison et al. (2004), Reitzig (2004) and Chang, Hao, Chen, and Yuan (2014), claim data are split into independent and dependent claims. A patent must have at least one independent claim and are those that stand alone, without dependence on any other claim, while dependent claims include a reference to another claim.

These claims outline a precise description of the solution to a known problem, define the novel features of the invention and describe the degree of inventive step the invention provides. The more independent claims a patent has the broader its legal protection. Prior studies such as those by Barney (2002), Allison et al. (2004) and Chang et al. (2014) found that valuable patents contained more independent claims.

The variable $CLAIMS_{it}$ represents the total claims measure, the dependent claims measure is represented by the variable $DCLAIMS_{it}$ and independent claims by the variable $ICLAIMS_{it}$ in the models outlined later in this chapter.

Independent Patent Variable 5: Patent Forward Citations

The patent forward citations are defined in this study as the citations a patent receives from any subsequent patents. These citations are listed under the "Referenced By" section on the front page of a patent. The patent that receives these citations is the focal patent. The citations that a patent receives signal that the cited patent contributed to the state of the art in a technology field and hence the citation count is often used as a measure of the patent's value. According to numerous prior studies there exists a strong positive relationship between forward citations and the technological importance of patents (for example see work by Pakes (1985), Trajtenberg (1990), Deng et al. (1999), Hall et al. (2005), Gambardella et al. (2008) and Serrano (2010)). There is therefore a strong foundation in the literature for utilising forward citations as a measure of patent quality and value. These citations are viewed as a good proxy for the technological significance and economic value of a patent. Prior studies[59] also found that forward citations are positively correlated with the market value of the firm. The more citations a patent receives may mean that it is viewed as being

[58] Prior studies found that claims also proved to be a good predictor of quality and value. See for example Albert et al. (1991), Tong and Frame (1994) and Harhoff, Narin, Scherer and Vopel (1999).
[59] See for example, Hall et al. (2001) and Harhoff et al. (1999).

of more interest and value to other inventors. "Thus, the number of times a patent document is cited may be a measure of its technological significance" p. 87 OECD (2002)[60].

There are some data issues surrounding the use of forward citation data. There is a time lag until a patent begins to receive citations. A patent therefore accumulates citations over time, so as time passes forward citations increase. Data will therefore be right censored, as citations to more recent patents cannot be observed. To account for this, a time-limited window is applied to the citation data. For example, previous studies (e.g. Reitzig, 2003; Harhoff, Scherer and Vopel, 2003 and Liu, 2014) have employed differing citation windows for their research. As this study wishes to provide timely indicators of a patent's importance, a five-year window[61] is employed. Selecting a time limited window ensures that every patent has the same period to demonstrate its technological significance and possible truncation issues are resolved. The calculation of the forward citation window is based on the time difference between the grant date of the patent cited on the front page of the citing patent and the application date of the citing patent for patents applied for prior to 29th November 2000[62]. After this date, patents are published eighteen months after filing. All citations that occur within 5 years of either the grant or the publication date are counted. The measure includes all citations by subsequent patents (U.S. and non-U.S.) and those from U.S. patent applications and is represented by the variable FC_{it}. I perform crosschecks for sensitivity/robustness of the empirical results by using alternative time windows and these are discussed in Chapter 5.

Independent Patent Variable 6: Forward Citation Lag Length

Fleming and Sorenson (2004) treated forward citations as a measure of the speed of diffusion of the knowledge contained within a patent. The shorter the citation lag length the faster the patent's idea is recognised as useful and with potential for commercial success by other researchers. Cassiman, Veugelers and Zuniga (2008) found that patents with a shorter forward citation lag have a greater technological impact. For patents applied for prior to 29th

[60] Benchmarking science-industry relationships, OECD Publications, Paris, France
[61] I calculate the number of forward citations received within a five-year time window to avoid any bias from older patents that have a longer time period in which they receive citations compared to more recent patents.
[62]According to the USPTO, patent applications for a patent filed on or after November 29, 2000 are published promptly after the expiration of a period of eighteen months from the earliest filing date for which a benefit is sought under title 35, United States Code (eighteen-month publication or pre-grant publication (PGPub)). See 35 U.S.C. 122(b).

November 2000 the calculation of the forward citation lag is based on the time (in days) between the filing date of the citing patent (patent B), and the grant date of the patent cited (patent A) on the front page of patent B. If the cited patent, patent A, was filed from 30th November 2000 onwards, the time between the date of publication of patent A is used instead of the grant date. The time elapsed from the filing date of the citing patent, patent B, to the grant or publication date of patent A, is represented by the variable FCL_{it}.

4.4.1.3 Independent Variables: Firm Publishing Characteristics

This study investigates whether the publishing behaviour of a patent-holding firm influences the likelihood of regulatory drug approval success.

Independent Publishing Variable 1: Publication of Scientific Journal Articles

This study uses firm-level publications in scientific journals as an indicator of the level of scientific knowledge and skills within the firm. It seeks to measure the impact of a firm's publishing behaviour on the quality of a firm's research output. Publication counts can be regarded as a measure of the quality of innovative activity. There are several reasons why firms publish scientific journal articles. Cockburn and Henderson (1998) interviewed research managers in pharmaceutical firms and found that publishing allowed firms to use the open scientific networks to develop skills, so improving the absorptive capacity of the firm and allowing them to utilise the advances made by public science. Stern (2004) found that there is a productivity effect from publishing behaviour, with firms that encourage academic publishing gaining earlier access to scientific discoveries that have commercial opportunities. Narin et al (1987), in their study of U.S. pharmaceutical firms, utilised the number of scientific articles published by a firm's researchers as a measure of human capital within the firm. Firms who publish scientific journal articles are therefore more likely to discover new drugs that will be approved by the FDA. This measure is represented by the variable PUB_{it} in the model and represents the firm's average annual publication activity.

Independent Publishing Variable 2: Co-authoring with Universities

Some firms choose to co-author with academia. This study is interested in what influence this activity has on firm research and eventually on its drug approval outcomes. To do this co-authoring behaviour by firms and universities is analysed.

The relationship between public research and firm-level R&D needs to be examined as this relationship plays a key role in the commercial success of pharmaceutical firms. Co-authoring allows the generation and exchange of new knowledge, can be used as a measure of private-public collaboration, measures of the strength of the interaction between the two environments and produces empirical evidence of the interaction of firms with academia (see for example Zucker and Darby, 1995; Liebeskind et al., 1996 and Zucker et al., 1998). Co-authorship is assumed to demonstrate evidence of joint problem solving and requires a significant investment by the firm in terms of time and resources. It often reflects joint research and provides the opportunity for knowledge exchange. The writing of joint publications requires higher levels of commitment by both individual researchers and the firm. The firm is choosing to invest in the development of connections to the public research network. Co-authorship therefore provides strong evidence of research collaboration. This study measures the influence of co-authoring behaviour on firm-level research quality as measured by the likelihood of a successful drug approval.

Following a number of studies such as Cockburn and Henderson (1998) and Fabrizio (2009) amongst others, I proxy for the degree of connectedness with university scientists with the percentage of the firm's publications in a year that are co-authored with university scientists. This measure is represented by the variable $COAUTHAC_{it.}$ and represents the firm's average annual academic co-authoring publication activity.

Independent Publishing Variable 3: Co-authoring with Firms

Co-authorship with other firms in the industry can be assumed to demonstrate evidence of joint research and requires a significant investment by the firm in terms of time and resources. It reflects the decision by at least two firms that they will benefit more from collaboration than they would lose from sharing knowledge and may increase their likelihood of attaining a commercially successful idea in the future and is represented by the variable $COAUTHFIRM_{it}$. This variable takes the value 1 if the firm co-authors scientific journal articles with another firm and 0 otherwise.

Independent Publishing Variable 4: Firm Funding Behaviour

Funding research projects with academia or other firms in the industry can be assumed be demonstrate evidence of joint research and a financial commitment by the firm. It reflects the decision by at firm that they will benefit more from collaboration than they would gain

from investing these funds in one of their own research projects and may increase their likelihood of attaining a commercially successful idea in the future and is represented by the variable $FUNDER_{it}$. This variable takes the value 1 if the firm funds a research project and 0 otherwise.

4.4.1.4 Control Variables

These variables are chosen because they have been previously established as firm characteristics that have an impact on the dependent variable, but their estimated relationship with the dependent variable is not the primary focus of the study.

Previous literature has shown that researchers employ a variety of variables in explaining the relationship between patenting by a firm and its valuation. Some variables are used in common by researchers as control variables despite the fact that the research questions differ. Certain control variables have been recommended by prior patent-based studies (for example see Deng et al., 1999; Fabrizio, 2009; Pandit et al., 2011 and Liu, 2014). It is acknowledged that there may be other omitted variables not controlled for in the model.

Control variables in the present study include firm size, firm age at patent application, profitability and prior experience of the FDA approval process. All these variables are used in this study as control variables in the analysis of patenting and firm-specific characteristics. Chapter 3, Section 3.5, provides the justification for the use of these particular control variables, while the following subsections present the definitions of the control variables and discuss any weaknesses associated with their use.

Control Variable 1: Firm Size

Firm size tends to vary widely in the pharmaceutical industry, ranging from the smallest start-up to the largest multi-national firm and this reflects in the large variability in patenting and commercial activity.

Previous research suggests that firm size may be important in explaining the heterogeneity in the extent of firm patenting. There may be variation in the propensity to patent amongst firms. Larger firms may spend more on R&D and have larger research teams so tend to produce more research outputs (e.g. patents). The high costs of patenting have a greater impact on smaller firms so they may have to be more selective in the ideas they patent. Smaller firms also tend to be younger and therefore will have more limited experience of the

patenting or FDA regulatory system. Importantly, size matters because larger firms can apply the results of their R&D over a greater output than smaller firms can. Studies that have considered firm size in the context of patenting and firm performance have come to inconsistent conclusions about the role that firm size plays. DiMasi (2014) found that larger firms were less successful in obtaining U.S. regulatory approval, but smaller firms had lower returns in terms of sales. On one hand, Jensen (1987) employed actual new products as a dependent variable and reports that firm size has no effect when introduced as an independent variable. On the other hand, studies that have proxied for new products have shown firm effects to be significant (Bound et al., 1982; Acs and Audretsch, 1989; Shan, Walker and Kogut, 1994; Nohria and Gulati, 1996 and Rothaermel and Hess, 2007). These inconclusive results from prior literature make it essential that the study must control for firm size.

Ahuja and Morris Lampert (2001), Rothaermel and Deeds (2004) and Liu (2014) used the log of employees, with log of sales being utilised by Halperin and Chakrabarti (1987) and Sakakibara (2010), whereas Coombs and Deeds (2000), Jong and Slavova (2014) and Kapoor and Klueter (2015) employed the log of total assets to capture firm size. It is noted that there is no consistency in the use of any single proxy for size and so it is left at the discretion of the researcher to choose amongst these commonly used proxies.

Data were collected from Compustat for the number of employees (Compustat code: EMP), total assets (Compustat code: AT), Sales/Turnover (Net) (Compustat code: SALE), market capitalisation (product of CRSP variables Price and Number of Shares Outstanding), research and development expense (Compustat code: XRD) and cash (Compustat code: CH).

In this study, firm size is computed as the logarithm of total assets. The reason for using total assets instead of sales or employees as a proxy for firm size is that there are data issues with the sales and employee variables[63].

Total assets in the year the patent is filed (expressed as the natural logarithm of total assets) is used as the measure of firm size in this study. As it is an internally determined measure of

[63] Sales data had zero values for 44 observations and employee data had 310 missing observations at patent filing date. Market capitalisation had 988 missing observations at patent filing date, research and development expense had 20 missing observations and cash had 1518 missing observations. Total assets had no missing or zero values at patent filing date.

a firm's size as opposed to an externally determined measure effected by the volatility of the market price of the firm's stock (Wallace and Naser, 1995). Total assets might proxy for size, yet as any accounting number, total assets can be distorted. Nevertheless, this measure creates fewer econometric problems and does not require elimination of observations. I choose total assets as the primary deflator for the tests carried out in Chapter 5. This measure is represented by the variable $SIZE_{it}$ in the models outlined later in this chapter.

Due to the lack of consensus on the issue of scaling in prior literature, I perform crosschecks for sensitivity/robustness of the empirical results by using alternative scale proxies. These include the number of employees - measured by the natural log of the number of employees in the year the patent is filed and market capitalisation - measured by the natural log of the market capitalisation in the year the patent is filed.

Control Variable 2: Profitability (Return on Assets)

Profitability has been found to be an important firm-specific variable that affects patenting (Deng et al., 1999) and the measure reported here, return on assets (ROA), is a measure of current returns. While discussions concerning appropriate measures of firm profitability are prevalent (for example, Venkatraman and Ramanujam, 1986 and Scherer and Ross, 1990), ROA is a widely used profitability measure in innovation studies (Roberts and Amit, 2003; Sher and Yang, 2005) and captures the ability of a firm to develop profits from its asset or investment base. It is usually viewed as an indicator of the performance of the firm. It is one of the key characteristics that are associated with the extent of patenting. When profitability is high, firms tend to spend more on R&D and so patenting increases. If profitability is low, firms will have fewer resources to devote to R&D and to producing patented inventions. Firms that are more profitable may have more patents and therefore eventually more approved drugs.

One of the measures of a firm's profitability is the rate of return is the return on assets (ROA), ROA is calculated as:

$$ROA_{it} = \frac{Net\ Income\ (Loss)_{it}}{Total\ Assets_{it}}$$

Net income (loss) and total assets in the year of patent filing are collected from Compustat using variable codes NI (net income) and AT (Assets - Total). Compustat's definition of Net Income is as follows and is adopted in this study:

"This item represents the fiscal period income or loss reported by a firm after subtracting expenses and losses from all revenues and gains."

The U.S. and Canadian GAAP Definition of Total Assets in Compustat is:

"This item represents the total assets/liabilities of a firm at a point in time."

Control Variable 3: Firm Age

Prior literature has found a variety of findings in this area. Some studies found that patent counts tend to vary with firm age (McNamara and Baden-Fuller, 2007 and Khoury and Pleggenkuhle-Miles, 2011). Barnett (1990) found that firm age had a negative effect on firm innovation levels. However, Sorensen and Stuart (2000) found a positive effect, as increased experience seems to enhance firm innovation ability. There was also a negative effect, where older firms fail to keep up with recent advances in science and are more complacent in their outlook. Older firms are viewed as being less able to incorporate radical change in an efficient manner.

Age is measured in years from the date the firm was founded to the year of patent filing. I follow the previous literature and define firm age as the number of years (plus one) as this avoids zero values for age. The founding date was determined using multiple sources including the firm's website. IPO dates are collected from Compustat using the code IPODATE and the date of incorporation is collected from Datastream using variable code WC18273 (date of incorporation - represents the date the firm was incorporated). Using these sources, the earliest year is utilised as the firm-founding year. This measure is represented by the variable AGE_{it} in the models outlined later in this chapter.

Control Variable 4: Firm Experience of the FDA Drug Approval Process

Prior experience of the FDA regulatory drug approval process can increase the likelihood that a firm will have a successful application outcome. The firm will already be familiar with the approval system and may have an existing working relationship with FDA staff. The

stock of successful drug approvals a firm possesses can also be viewed as a proxy for accumulated knowledge capital. The variable indicates whether a firm has had any prior drug (NME) approvals in the sample period. This controls for the firm's experience in the drug discovery process and in managing the FDA approval process. Past approvals are collected for each firm from the FDA's website. This past approval experience is represented by the variable $FDAEXP_{it}$, which takes the value 1 if the firm has at least one prior approval and 0 otherwise.

4.5 Model Specification and Estimation Techniques

This section specifies the models for estimating the potential relationship of patent characteristics, firm-level publishing behaviour and drug approval success.

The discrete nature of the dependent variable requires a model that departs from the simple linear form. I am interested in the probability that a firm with certain publishing and patent characteristics will obtain drug approval from the FDA. As the probabilities are bounded between 0 and 1, a linear model may lead to inaccurate probabilities and the error term may suffer from heteroscedasticity (Greene, 2003). A conditional logistic model allows the function that links the dependent and independent variables to be distributed in a non-linear form. The data are analysed by logistic regressions and a conditional (fixed effect) logistic regression is applied. Logistic regression is an appropriate model to test such relationships, as it has the ability to predict the likelihood of certain events occurring. Further, logistic regression also requires less restrictive assumptions than OLS as homogeneity of variance and normality of errors are not assumed. Diagnostic tests are performed to ensure that the assumptions of logistic regression are fully met, and these are discussed in Chapter 5.

In a logistic model regression analysis, the framework is a little different in that the dependent variable is assumed to have a Bernoulli distribution with mean=$\pi_{\underline{x}}$ and variance= $\pi_{\underline{x}}$ $(1-\pi_{\underline{x}})$. The logistic function does not model the mean of Y = $\pi_{\underline{x}}$ linearly in $\underline{x} = [X_1 \ldots X_5]$, where X_i is a set of independent variables. Instead, it models the logit of the mean of Y = $\pi_{\underline{x}}$ linearly in $\underline{x} = [X_1 \ldots X_5]$.

The general relationship tested is presented in the form of a logistic function as:

$$\text{Logit } [E(Y)] = \text{logit } (\pi_{\underline{x}}) = \ln (\pi_x/(1-\pi_x)) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 X_{it}} + \boldsymbol{\beta_2 X_{it}} + \boldsymbol{\beta_3 X_{it}} + \boldsymbol{\beta_4 X_{it}} + \sum_{i=1}^{n} \boldsymbol{\beta_i X_{it}} + \boldsymbol{\varepsilon_{it}}$$

The logistic regression model focuses on the odds of an event occurring (in this case, drug approval). The data are presented in panel data format with variables being measured for the same firm, at multiple points in time. I therefore use Stata's xtlogit command that is devoted to working with Panel data, the fixed effect (fe) option is utilised to focus on only time variant factors. In order to use this command, the dataset needs to be properly structured. I therefore use the xtset facility within Stata to format the data and create a Panel. In this study, the Panel variable is the firm's id code and the time variable is either the patent grant date or the publication date. The latter date is selected if the patent was filed from 30th November 2000 onwards. This produces an unbalanced dataset as patent data are missing for some firm years.

These models are constructed with the variables defined earlier in this chapter and are empirically tested in the following chapters. The models are developed with reference to the hypotheses stated in sections 3.2 to 3.5 of Chapter 3. Once a logistic regression model has been fitted, the prediction equation can be used to estimate odds ratio (OR) measures of association, which are reported in the following chapter.

Diagnostic tests including examination for the normality of data using visual aids, variable descriptive statistics and correlation analysis, VIF and Eigen values, amongst others, are used to identify any problems with the data. Chapter 5 provides a detailed discussion of any data issues and measures employed to overcome these issues.

4.5.1 Baseline Model

The study first estimated a baseline model using the control variables alone. The hypotheses that inform this model are those outlined in Section 3.5 of Chapter 3. Hypotheses H_{4a}, H_{4b}, H_{4c} and H_{4d} relate to the firm-level characteristics that are available at the point in time that the patent application is filed and state that:

H_{4a}: The size of a firm influences the likelihood of receiving regulatory drug approval from the FDA.

H_{4b}: The age of a firm influences the likelihood of receiving regulatory drug approval from the FDA.

H_{4c}: If the firm experiences higher levels of profitability, it will produce greater numbers of patents. This increases the likelihood of producing patents of greater technological importance that are more likely to receive regulatory drug approval from the FDA.

H_{4d}: If the firm possess prior experience of the FDA regulatory drug approval process this increases the likelihood of receiving regulatory drug approval from the FDA.

Regulatory approval information is collected from the Original New Drug Approvals by Month section of the Drugs@FDA database on the FDA's website, patent data from Google Patents and the USPTO and financial data from Compustat. I estimate the following equation:

$$DRUG\ APPROVAL_{it} = \beta_0 + \beta_1 AGE_{\text{it}} + \beta_2 FDAEXP_{\text{it}} + \beta_3 SIZE_{\text{it}} + \beta_4 ROA_{\text{it}} + \varepsilon_{it} \quad (4.5.1)$$

Where:

DRUG APPROVAL$_{it}$ is 1 in the presence of a patent linked to an FDA drug approval or 0 otherwise

β_0 is a constant term

AGE$_{it}$ is firm age at the patent filing date, measured in years

FDAEXP$_{it}$ is an indicator of firm experience of the FDA regulatory drug approval process and takes the value 1 if the firm has at least one prior approval (at the patent filing date) and 0 otherwise

SIZE$_{it}$ is firm size measured by total assets at the patent filing date

ROA$_{it}$ is the firm's return on assets at the patent filing date

ε_{it} is the error term and subscript i denotes the firm where i = 1, 2… n and subscript t denotes time.

For the model above a sample of firms, whose patents are both related to and un-related to an approved drug is used.

Hypotheses H_{4a}, H_{4b}, H_{4c} and H_{4d} predict that β_1, β_2 (if prior experience exists), β_3 and β_4 are positive.

Next, the effects from the various hypotheses are added.

4.5.2 Model Specification: Patent Characteristics

This section specifies the models for estimating the effect of certain patent characteristics on the likelihood of regulatory drug approval. The hypotheses that inform these models are those outlined in sections 3.2 and 3.4 of Chapter 3. Hypotheses H_{1a}, H_{1b}, H_{1c}, $H_{1c(i)}$, $H_{1c(ii)}$, H_{1d} and H_{1e} relate to the patent characteristics that are available at the point in time that the patent is granted or published and state that:

H_{1a}: Firms whose patents have a larger inventor team size (number of inventors listed on the front page of the patent) will have an increased likelihood of receiving regulatory drug approval from the FDA.

H_{1b}: Firms with higher levels of science intensity (patent prior art references to scientific publications) will have an increased likelihood of receiving regulatory drug approval from the FDA.

H_{1c}: Firms whose patents contain higher numbers of claims will have an increased likelihood of receiving regulatory drug approval from the FDA.

$H_{1c(i)}$: Firms whose patents contain higher numbers of independent claims will have an increased likelihood of receiving regulatory drug approval from the FDA.

$H_{1c(ii)}$: Firms whose patents contain higher numbers of dependent claims will have an increased likelihood of receiving regulatory drug approval from the FDA.

H_{1d}: Firms whose patents have a shorter citation lag to non-patent prior art references (NPR) will have an increased likelihood of receiving regulatory drug approval from the FDA.

H_{1e}: Firms whose patents have a shorter citation lag to patent prior art references will have an increased likelihood of receiving regulatory drug approval from the FDA.

To empirically test the above hypotheses, the likelihood of drug approval is estimated as a function of certain patent characteristics, as specified below:

$$DRUG\ APPROVAL_{it} = \beta_0 + \beta_1 INVENTORS_{it} + \beta_2 NPR_{it} + \beta_3 ICLAIMS_{it} + \beta_4 DCLAIMS_{it} + \beta_5 NCL_{it} + \beta_6 PCL_{it} + \beta_7 AGE_{\mathrm{it}} + \beta_8 FDAEXP_{\mathrm{it}} + \beta_9 SIZE_{\mathrm{it}} + \beta_{10} ROA_{\mathrm{it}} + \varepsilon_{it} \quad (4.5.2a)$$

Where:

$INVENTORS_{it}$ is the number of inventors listed on the front of the patent document

NPR_{it} is the number of non-patent references cited in the references cited section of the patent document

$ICLAIMS_{it}$ is the number of independent claims listed at the end of the patent document

$DCLAIMS_{it}$ is the number of dependent claims listed at the end of the patent document

NCL_{it} is the period, in days, from the patent filing year to the publication year of the latest publication cited by the patent

PCL_{it} is the period, in days, from the patent filing date to the grant date of the latest patent cited by the patent

Hypotheses $H_{1a,}$ $H_{1b,}$ $H_{1c(i)}$ and $H_{1c(ii)}$ predict that β_1, β_2, β_3 and β_4 are positive, whereas hypotheses H_{1d} and H_{1e} predict that β_5 and β_6 are negative. The implicit null hypothesis is that the variation in non-patent reference numbers, inventors, claim numbers and average citation lag lengths is not systematically related to the firm's likelihood of approval success.

Hypotheses H_{3a} and H_{3b} relate to patent characteristics that become available during the period following a patent grant or publication date and state that:

H_{3a}: Firms whose patents receive citations from subsequent patents (forward citations), within a five-year time window, will have an increased likelihood of receiving regulatory drug approval from the FDA.

H_{3b}: Firms whose patents receive citations from subsequent patents (forward citations) at a faster rate will have an increased likelihood of receiving regulatory drug approval from the FDA.

Finally, patent variables available after patent publication (forward citations and forward citations lag length) are added.

$$DRUG\ APPROVAL_{it} = \beta_0 + \beta_1 INVENTORS_{it} + \beta_2 NPR_{it} + \beta_3 ICLAIMS_{it} + \beta_4 DCLAIMS_{it} + \beta_5 NCL_{it} + \beta_6 PCL_{it} + \beta_7 FC_{it} + \beta_8 FCL_{it} + \beta_9 AGE_{\mathrm{it}} + \beta_{10} FDAEXP_{\mathrm{it}} + \beta_{11} SIZE_{\mathrm{it}} + \beta_{12} ROA_{\mathrm{it}} + \varepsilon_{it} \quad (4.5.2b)$$

Where:

FC_{it} is the number of citations received from subsequent patents (forward citations) with a five-year period from the cited patent's grant date (or the publication date, if the patent is filed from 30th November 2000 onwards)

FCL_{it} is the period, in days, from the filing date of the citing patent to the grant date of the cited patent (or the publication date, if the patent is filed from 30th November 2000 onwards)

Hypotheses H_{3a} predicts that β_7 is positive and H_{3b} predicts that β_8 is negative. The implicit null hypothesis is that the variation in forward citation numbers and average citation lag lengths is not systematically related to the firm's likelihood of approval success.

4.5.3 Model Specification: Firm Publishing Characteristics

The following discussion specifies the models for estimating the effect of certain firm-level publishing characteristics on the likelihood of regulatory drug approval. The hypotheses that inform these models are those outlined in Section 3.3 of Chapter 3. Hypotheses H_{2a}, H_{2b}, H_{2c} and H_{2d} state that:

H_{2a}: Firms whose research teams publish scientific journal articles will produce patents of greater technological importance that will have an increased likelihood of receiving regulatory drug approval from the FDA.

H_{2b}: Firms whose research teams co-author scientific journal articles with academic researchers will produce patents of greater technological importance that will have an increased likelihood of receiving regulatory drug approval from the FDA.

H_{2c}: Firms whose research teams co-publish scientific journal articles with researchers from other firms will produce patents of greater technological importance that will have an increased likelihood of receiving regulatory drug approval from the FDA.

H_{2d}: Firms that fund research projects that result in co-authored scientific journal articles, will produce patents of greater technological importance that will have an increased likelihood of receiving regulatory drug approval from the FDA.

To empirically test the above hypotheses, the likelihood of drug approval is estimated as a function of certain firm-level publishing characteristics, as specified below:

$$DRUG\ APPROVAL_{it} = \beta_0 + \beta_1 PUB_{it} + \beta_2 COAUTHAC_{it} + \beta_3 COAUTHFIRM_{it} + \beta_4 FUNDER_{it} + \beta_5 AGE_{\mathrm{it}} + \beta_6 FDAEXP_{\mathrm{it}} + \beta_7 SIZE_{\mathrm{it}} + \beta_8 ROA_{\mathrm{it}} + \varepsilon_{it} \quad (4.5.3)$$

Where:

PUB_{it} is the firm's average annual publishing output of journal articles

$COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage

$COAUTHFIRM_{it}$ is 1 when the firm co-authors scientific journal articles with another firm or 0 otherwise

$FUNDER_{it}$ is 1 when the firm funds co-authored scientific journal articles or 0 otherwise

Hypotheses H_{2a} to H_{2d} predict that β_1, β_2, β_3 and β_4 are positive. The implicit null hypothesis is that the variation in average publication numbers, percentage of university co-authored articles, firm co-authoring and firm funding behaviour is not systematically related to the firm's likelihood of approval success.

Finally, the likelihood of drug approval is estimated as a function of all the selected patent characteristics and firm-level publishing characteristics, as specified below:

$$DRUG\ APPROVAL_{it} = \beta_0 + \beta_1 INVENTORS_{it} + \beta_2 NPR_{it} + \beta_3 ICLAIMS_{it} + \beta_4 DCLAIMS_{it} + \beta_5 NCL_{it} + \beta_6 PCL_{it} + \beta_7 FC_{it} + \beta_8 FCL_{it} + \beta_9 PUB_{it} + \beta_{10} COAUTHAC_{it} + \beta_{11} COAUTHFIRM_{it} + \beta_{12} FUNDER_{it} + \beta_{13} AGE_{it} + \beta_{14} FDAEXP_{it} + \beta_{15} SIZE_{it} + \beta_{16} ROA_{it} + \varepsilon_{it} \quad (4.5.4)$$

4.6 Model Evaluation

This section outlines the methods employed to evaluate model predictive performance.

4.6.1 Introduction

Whilst the explanatory power of the models developed using the techniques discussed in Section 4.5 is important, it is not necessarily their key characteristic. One of the primary goals of this study is the development of a model that can also predict the likelihood of a patent possessing a linkage to an approved drug. This section discusses the methodology employed to evaluate the model's performance. That is, does the model possess the ability to distinguish between approval-linked and non-approval linked patents within sample and can it predict future approval-linked patents? Predictive performance is therefore ascertained by conducting both in-sample and out-of-sample analyses. Section 4.6.2 discusses the use of in-sample and out-of-sample performance analysis.

4.6.2 Model Comparison Using Portfolio Approval Concentration Measures

To my knowledge, there is no prior literature that attempts to predict drug approval based on patent-level characteristics and firm publishing behaviour. I therefore use a common technique from other finance literature, in the areas of takeover and bankruptcy prediction, to assess the performance of selected models and to measure their ability to predict an event out-of-sample (see, for example, Ohlson (1980), Palepu (1986), Holthausen and Larcker (1992), Barnes (1998, 1999), and Powell (2001, 2004). Once a model formulation has been selected, based on the techniques outlined in Section 4.5, Stata is used to identify which patents are more likely to be linked to an approved drug. Approval-linkage probabilities are generated from the logit regression model's coefficients. These approval-linkage probabilities are then used to construct ten portfolios (deciles), which are sorted in

descending order of approval-linkage probability. Finally, these portfolios are assessed for predictive accuracy.

4.6.2.1 Identifying the Optimal Patent Portfolio using Cut-off Probabilities

The portfolio selection method employed is an important consideration in this study as this selection process determines the returns to patents predicted to have an approval-linkage (analysed in Chapter 7).

The logit model (for approval prediction) reports its output in terms of probability. That is, the model uses its coefficients to transform the independent variables for any patent into a probability value. This probability value represents the likelihood that the patent will be linked to an approval based on the publicly available information available for the patent and the patent-holding firm. The computed approval-linkage probabilities range from 0 to 1. A critical step in the predictive process is to determine a cut-off point for the approval-linkage probability, where the patent would be considered a potential approval-linked patent. In prior literature some studies employed an arbitrary cut-off point of 0.5 (see Ohlson (1980) and Palepu (1986)), as outcomes are either coded 1 or 0. Patents above this point would be classified as approval-linked and vice versa. However, this model is attempting to predict a relatively rare event, as approval-linked patents comprise 4.88% of the total patent sample. This means that any in-sample and out-of-sample datasets have unequal numbers of approval-linked and non-approval-linked patents, making the use off a cut-off value of 0.5 inappropriate.

In prior literature there are alternative approaches employed in selecting the optimal cut-off point for portfolio selection. The first, proposed by Palepu (1986) and extended in Barnes (1998), is based on an objective to minimise the total number of misclassifications, so minimising both Type I and Type II errors[64]. This approach assumes that the costs of committing both error types are equal. The second approach, proposed by Powell (2001), is based on an objective to maximise the proportion of accurately predicted approval-linked patents in the selected approval-linked patent portfolio. The cut-off point selected using this

[64] A Type I error, is a case where the selected cut-off probability (ex-ante) allows a patent to be incorrectly classified as non-approval linked (ex-post). A Type II error is a case where the selected cut-off probability (ex-ante) allows a non-approval linked patent to be incorrectly classified as an approval-linked patent (ex-post).

approach allows non-approval linked patents to be classified as approval-linked only if this increases the number of actual approval-linked patents within the approval-linked patent portfolio. This assumes that the cost of a Type I error is higher than the cost of a Type II error. Palepu's procedure leads to the selection of a lower cut-off probability when compared with Powell's procedure. This would mean that approval-linked patent portfolios developed using the Palepu (1986) approach are likely to have a higher number of approval-linked patents but also a higher number of non- approval linked patents misclassified as approval-linked. Powell's (2001) approach imposes a stricter rule for including each prospective approval-linked patent into the approval-linked patent portfolio.

The Powell (2001) procedure for computing optimal cut-off probabilities is adopted in this study. This decision assumes that in the drug development process the costs of a Type I error, not proceeding with a research project that would result in an approved drug, exceed the costs of a Type II error, proceeding with a research project that would not result in an approved drug. With a Type I error the firm forgoes the future revenues from a successful drug product which can amount to billions of dollars. Type II errors cause the firm to incur unnecessary drug clinical trial costs. See Chapter 6, Section 6.3, for a more in-depth discussion of these costs. The portfolios constructed using Powell's decile method are examined to gauge any differences in the portfolios' approval concentration ratio (a measure of predictive ability) and is computed as follows.

$$Approval\ Concentration\ Ratio\ (ACR) = \frac{Number\ of\ patents\ corrrectly\ predicted\ to\ have\ an\ approval\ linkage}{Total\ number\ of\ patents\ predicted\ to\ have\ an\ approval\ linkage}$$

Following Powell (2001), this technique identifies an optimal portfolio within which the proportion of correctly predicted approval-linked patents is highest and the number of misclassifications of non-approval linked patents is lowest. The optimal cut-off point selected maximises the portfolio's approval concentration ratio (ACR). The cut-off value selected is the first approval probability in a portfolio that produces the highest ACR.

4.6.3 In-Sample and Out-of-Sample Testing

Once cut-off points have been selected based on this approach, I proceed to test the predictive ability of the model. I firstly conduct in-sample tests of the model's predictive ability. I initially estimate the model using data from the entire sample period but only including patent characteristic variables that are available at the date of patent publication, firm-

publishing behaviour variables and control variables. I then introduce an additional forward citations variable as a patent characteristic variable. Firm-publishing behaviour variables and control variables are still present. The models' predictive abilities are assessed using cut-off points determined by a Powell (2001) approach and are reported at the upper and lower 10% and 20% cut-off values and at the 50% cut-off.

I then conduct a series of robustness and sensitivity checks on the results of the model in-sample testing. The objective of these checks is to establish whether the results reported for the in-sample analysis are robust. As this is an in-sample analysis robustness test, the sample cannot be partitioned but its composition can be altered, this is done by varying the firms in the sample. As a robustness check, I examined firms with differing levels of patenting activity to observe if this influences the predictive ability of the model with and without forward citations. Samples are created based on firms who have patenting levels above and below median levels, producing four sample groups.

Good predictive ability, as assessed by in-sample analysis, is very useful in assessing the goodness of fit of the model but it does not exclude the possibility that the model may suffer from overfitting and it does not necessarily provide the most accurate predictions. To assist in excluding these possibilities, an out-of-sample analysis is also conducted to assess the model's predictive ability. To conduct this type of analysis, I withhold a portion of the sample data from the model estimation process, then using the resultant model formulation I make predictions for those patents that are excluded from the estimation sample. This process assesses how accurate these predictions are and determines whether they are similar to those made using the in-sample analysis approach.

According to Kuhn and Johnson (2013), in datasets with binary dependent variables where the frequency of positive outcomes is substantially different from 50%, such as this one where approval-linked patents comprise 4.88% of the sample, it can be advisable to choose an estimation sample that closely represents the characteristics of the whole original data sample. For my first out-of-sample tests, I split my data based on time using only the observations for patents with grant years up to and including 1996. The model's predictive ability (with and without forward citations) is assessed using similar techniques to those employed for in-sample testing.

Next, I test the out-of-sample analysis to observe if it is robust to the selection of alternative time periods for the model estimation sub-sample and the model test sub-sample. To validate

my results, I tested a range of estimation and test samples. I introduce a shorter six-year estimation sample (patents granted during the period 1976-81) and then expand this estimation window, keeping the same 1976 start point, to 12 years (patents granted during the period 1976-87) and then to 18 years (patents granted during the period 1976-93).

As another robustness check, I also move the starting point of the six-year estimation sample to test if the patent composition of the estimation sample influences the model's predictive ability. I estimate the model using a sample of patents with grant years of 1976 to 1981, 1988 to 1993 and 1994 to 1999.

4.7 Evaluating Model Investment Potential

A key research question in this study seeks to explore is whether approval prediction can form the basis of a successful investment strategy. This builds on prior research findings that firms who receive an FDA drug approval gain substantial abnormal returns during the period surrounding the approval announcement (see for an example Dedman et al. (2008), Hamill et al. (2013) and Hamill et al. (2018)). Chapter 6 focuses on developing a portfolio of patents that are predicted to be linked to an approved drug. This subsection discusses the methods used in Chapter 7 to compute the abnormal returns earned by this portfolio.

I employ an event study methodology, with the market model used to generate expected returns, to quantify the impact on shareholder wealth of a firm's patent being linked to a drug approval. The first step in this process is computing the portfolio returns from stock returns. Daily discrete (or simple) returns are computed using the Center for Research in Security Prices (CRSP) Holding Period Returns that represents share prices adjusted for dividend payments. For this study, it is assumed that for approval-linked patents, the stock related to that patent is disposed once the drug to which it is linked is approved. Daily discrete returns for each patent-holding firm are then computed for the period from the publication date of the patent to the approval date of the drug to which it is linked. This approach results in the potential holding period differing between firms and within firms. For non-approval linked patents, there is no drug approval date for calculation purposes. Instead the approval-linked patents' median publication date to approval date period (in calendar days) is used as a proxy for the stock holding period, this gives a holding period of 1,078 calendar days. Simple firm returns are computed as follows.

$R_{it} = (R_{it} - R_{it-1})/ R_{it-1}$

R_{it} is the day t simple return on a sample firm and R_{it-1} is the prior trading day's return. These returns are adjusted for dividends and other capital changes. If there are any bankrupt firms in the stock portfolio, they are identified and removed. The abnormal return for a sample firm on day t is computed using the Mean Adjusted Return (MAR) Model. This model can be viewed as a restricted market model with alpha equal to zero and beta equal to one for each stock (see MacKinlay (1997)). The parameters are predefined, so a separate estimation window is not necessary, the model is computed as follows.

$$AR_{it} = R_{it} - R_{mt}$$

R_{mt} is the day t simple market return and is the drug industry return listed within the 48-industry portfolio in the Kenneth French Data Library[65].

For additional information I also cumulate returns across the holding period, this yields a cumulative abnormal return (CAR):

$$CAR_{it} = \sum_{t=1}^{\tau} AR_{it}$$

However, Barber and Lyon (1997) found that cumulative abnormal returns can be a biased predictor of long-run buy-and hold abnormal returns. I therefore compute buy-and-hold abnormal returns as I wish to detect the presence of long-run buy-and-hold abnormal returns over the holding period for each stock.

$$BHAR_{it} = \prod_{t=1}^{\tau}[1 + R_{it}] - \prod_{t=1}^{\tau}[1 + E(R_{it})]$$

$E(R_{it})$ is the day t expected return for the sample firm, estimated using the MAR Model.

Returns are computed for both equal-weighted portfolios and value-weighted portfolios. Equal-weighted portfolios assume that an equal amount is invested in each firm in the portfolio at the beginning of the portfolio-holding period and the portfolio is held until the end of this period. The value-weighted portfolios assume that an investor allocates his

[65] Available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed on 12th January 2018.

investment to the stocks in the portfolio in proportion to their market value at the beginning of the portfolio-holding period.

The second step in the process of computing portfolio abnormal returns is to adjust portfolio returns for risk. Several approaches (with different strengths and weaknesses) have been employed in the literature in a bid to adjust returns for the risks involved. To allow for robustness and comparison of the results with previous literature, portfolio daily returns are adjusted for risk by using the Capital Asset Pricing Model (CAPM) as the risk adjustment model.

$$R_{it} - R_{fit} = \alpha_i + \beta_{mkt} (R_{mt} - R_{ft}) + \varepsilon_{it}$$

Where, R_{it} is the discrete return on portfolio *i* on day *t*, R_{ft} is the risk free rate on day *t,* α_i is the abnormal (excess) daily return or portfolio alpha in the period, R_{mt} is the market return on day *t*. The data for the daily risk-free rate (R_f) are the daily return on the U.S. three year Treasury Bill obtained from Datastream and the daily market return (R_m) are again obtained from the Kenneth French Data Library.

As per the equation, I fit daily excess portfolio returns ($R_{it} - R_{ft}$) to excess market returns ($R_{mt} - R_{ft}$). The intercept or constant term (α_i) from this regression provides an estimate of returns that cannot be explained by common risk factors. Abnormal returns are computed for a test period that runs from the patent publication date to the drug approval date. The model estimation period from 250 days to 30 days prior to the publication date. This analysis is done in Stata.

4.8 Summary

This chapter covers the definitions of the independent, dependent and control variables. It explains the sample selection procedure, data sources and collection, the methodology and research design of this study. The first half of this chapter discusses the data included in this study, data definitions, data collection, the sample composition and time frame. Three main types of data are used in this study: firm patent data, firm publication and collaboration data and regulatory drug approval data. These were mainly collected from Google Patents and the USPTO, Web of Science database and the FDA Drugs@FDA database. The final sample consists of 12,957 patents (both associated and not associated with an approved drug) and 81 firms over a period of 40 years. The second half of the chapter outlines the specification

of models used to test the hypotheses as outlined in Chapter 3. The predictive performance of the model is evaluated by testing its ability to predict approval-linked patents out-of-sample (predictive ability). The model's ability to generate abnormal returns for investors is also tested using the methodology discussed in Section 4.7.

Overall, the focus of the methodology employed across the three empirical chapters of this study (Chapters 5, 6 and 7 that follow) is to ensure that a robust process is followed in both the development and testing of the approval prediction model. In the next chapter, the descriptive statistics and the univariate analytical results of the study's variables are reported.

Chapter 5 Hypotheses Validation

5.1 Introduction

This chapter discusses the data and the fixed effects (conditional likelihood) logistic regression assumptions. It sets out to present and discuss a detailed empirical analysis of the potential impact of firm patent and publishing characteristics on the probability of drug approval. Since the seminal paper by Trajtenberg in 1990, there has been a growing theoretical and empirical literature on the importance of particular patent characteristics to the success of a firm. In addition, the importance of a firm's relationship to basic scientific research has been recognised as a driver of innovative performance. This chapter specifically provides detailed univariate and multivariate logistic analyses to determine if the hypothesised relationship, discussed in Chapter 3, between firm patent and publishing characteristics and the probability of drug approval success holds.

The present chapter is divided into 5 main sections. Section 5.2 provides a detailed description of the sample utilised to analyse the impact of firm patent and publishing characteristics on the probability of drug approval success. It also discusses potential issues with the estimation techniques employed and any limitations in the methodologies utilised, with possible solutions to assuage these concerns. Section 5.3 presents a univariate analysis of the impact of firm patent and publishing characteristics on the probability of a successful drug approval. Section 5.4 presents a multivariate analysis of firm patent and publishing characteristics on the probability of drug approval success to confirm or disprove the hypotheses outlined in Chapter 3. Section 5.5 presents the robustness and sensitivity checks including the use of alternative variable measures and estimation techniques while Section 5.6 presents concluding remarks.

5.2 Sample Characteristics

This section presents the descriptive statistics for the selected variables that are proposed for the models testing the impact of firm patent and publishing characteristics on the probability of drug approval success. The section is divided into two main subsections. Subsection 5.2.1 presents the descriptive statistics of the variables before and after any necessary data transformation. Section 5.3 presents diagnostic tests for the sample suitability of the specific estimation techniques suggesting possible remedial measures, if the data violate any assumptions for the estimation techniques used. Firstly, the next section outlines how the

distributional characteristics of the data were assessed and where necessary how data outliers were treated.

5.2.1 Sample Statistics and Outlier Treatment

As discussed in Chapter 4, Section 4.2.4, the final sample consists of 12,957 firm patents drawn from a drug approval period of 40 years (1974 to 2014). All patents for which no approved-drug link was found are considered as non-approval-linked (further discussed in Section 4.2.4.2). Out of the 12,957 observations, 632 approval-linked patents are recorded leaving a sample of 12,325 non-approval-linked observations.

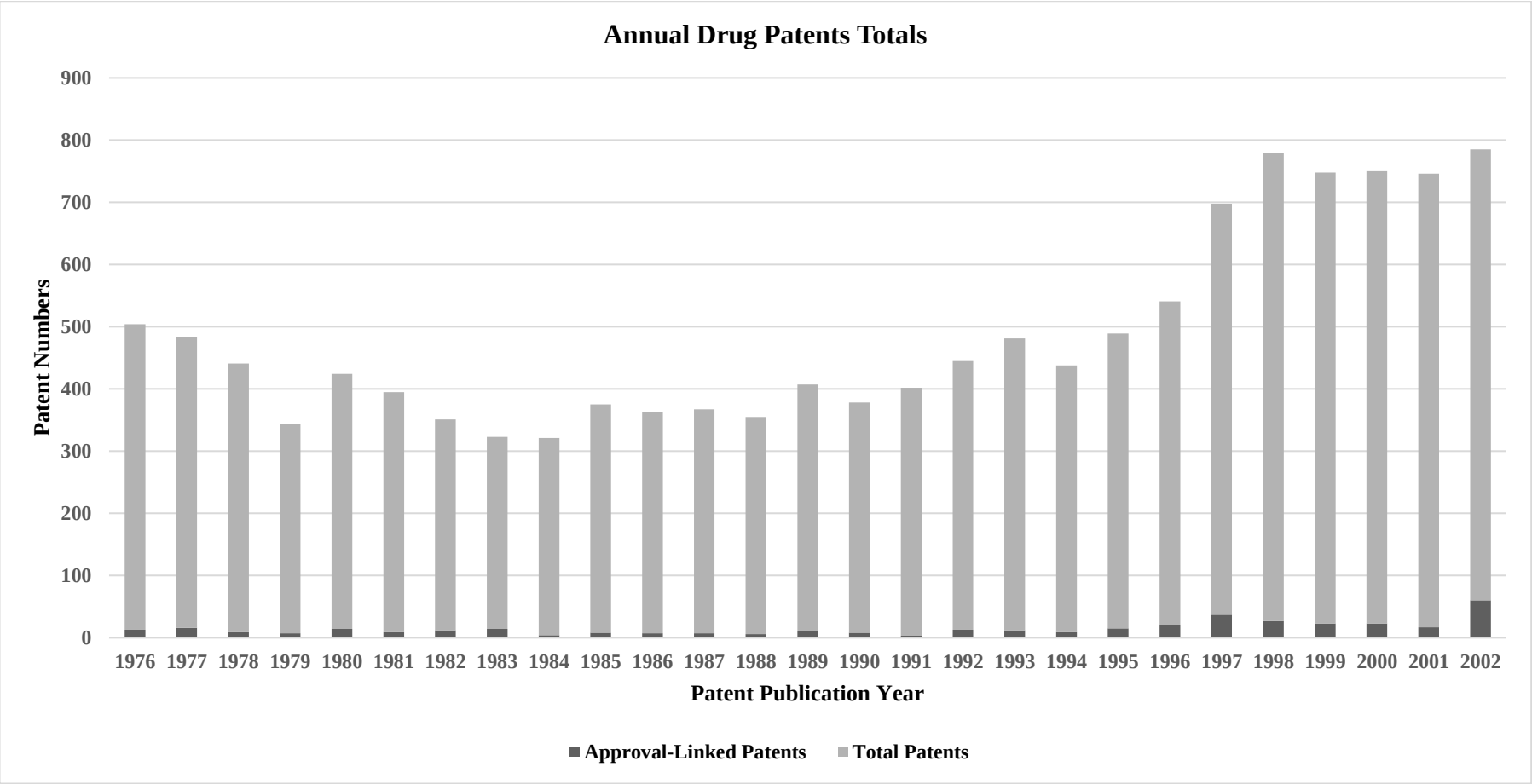
Figure 5-1 shows the distribution of patents for each year throughout the period, that has both approval and non-approval patent data (1976–2002). The data show that out of an average of 471 patents per year, 15 patents are linked to an approved drug (on average) in each year. The implication is that on average 3.18% of drug patents are linked to an approved drug in each year. However, this ratio varies from year to year. The lowest number of approval-linked patents is recorded in 1991 with just 3 patents out of a population of 399 (0.75%) being linked to an approved drug. Highs of 37 and 56 are recorded in 1997 and 2002 (5.60% and 7.77% respectively), marking peaks in U.S. drug approval activity.

5.2.1.1 Dependent, Independent and Control Variables Descriptive Statistics

This section presents detailed descriptive statistics for the study variables and a comparison of the firm-level patent and publication characteristics of approval-linked and non-approval-linked patents. The definitions and derivation of the variables are discussed in Chapter 4, Section 4.4.

In Panel A (Table 5.1), the descriptive statistics for the raw data are presented. As discussed later in this chapter, a key observation from Panel A is the presence of extreme and, seemingly, unlikely values. Some analysis on why such extreme values are observed is conducted and the results are presented in Panel B of Table 5.1. In general, I find that the extreme values are not primarily due to data errors but are in fact true data values as these data were checked and confirmed using original patent documents obtained from USPTO patent database. This is further discussed below.

Figure 5-1 Annual Drug Patent Totals



Notes: Figure 5-1 plots the annual variations in patenting levels for approval-linked and non-approval linked patents for the period 1976 to 2002. Approval-linked patents are indicated in black and total patent numbers in grey.

In Table 5.1, I consider different outlier treatment procedures given the presence of extreme values in my dataset. There are several extreme values in the raw data (Panel A) as evidenced by the skewness statistics, minimum and maximum values and mean and median values. The skewness statistics for measures of total assets, return on assets, NPR, patent claims, independent patent claims, dependent patent claims, NPR citation lag length, patent citation lag length and forward citations received are all above the acceptability threshold of 3. While skewness measures whether the data are normally distributed it also indicates the presence of extreme values. The large difference between the mean and median values (e.g., patent citation lag length), the very low (high) minimum (maximum) values, and the spread between the maximum and minimum values (e.g. total assets) are further indications of the existence of extreme values.

These results (Table 5.1, Panel A) raise some questions about data integrity. I attempt to evaluate data integrity by manually checking a sample of observations. Here, I compare the patent data obtained from Google Patents to the data in original patent documents obtained from the USPTO patent database. I compare the firm financial data obtained from Compustat to the data in original annual reports obtained from the Perfect Information database. The approach I adopt for this review process involves identifying a sample of firms with the highest (50 firm observations) and lowest values (50 firm observations) for each 'problem' variable. I then obtain the firms' original annual reports from the Perfect Information database and compare the data from Compustat to the data presented in the reports. For patent data, I retrieve patent documents from the USPTO patent database and compare the data from Google Patents to the data presented in the original patent.

The main variables with extreme observations are total assets, return on assets, NPR and forward citations received. After a review, I find that the "problem" variables generating the extreme values are net income, which can be highly volatile (hence a very noisy measure for a given year) and total assets. Total assets display a wide range of values due to the range of firm ages in the sample. The extreme observations for return on assets (net income compared to total assets) are due to negative reported net income for some firms and high levels of total assets for others.

Table 5.1 Descriptive Statistics of Variables Before Transformation

Panel A: Descriptive Statistics of Variables Before Transformation									
	N	**Mean**	**Std. Dev.**	**Range**	**Minimum**	**Maximum**	**Median**	**Skewness**	**Kurtosis**
Firm Age (AGE_{it})	12,957	80.966	45.148	165	1	166	85	-0.179	2.071
Firm FDA Experience ($FDAEXP_{it}$)	12,957	0.868	0.338	1	0	1	1	-2.177	5.738
Total Assets ($SIZE_{it}$)	12,957	7480.145	15531.080	49022.916	0.084	495023	3183	13.964	311.253
Return on Assets (ROA_{it})	12,957	0.037	0.301	5.278	-4.881	0.397	0.098	-7.398	76.724
Number of Inventors ($INVENTORS_{it}$)	12,957	2.760	1.972	17	1	18	2	2.745	15.479
Non-patent References (NPR_{it})	12,957	11.052	26.141	515	0	515	3	7.092	84.170
Patent Claims ($CLAIMS_{it}$)	12,957	13.934	14.311	357	1	358	10	5.258	67.148
Independent Patent Claims ($ICLAIMS_{it}$)	12,957	2.353	2.614	82	1	83	2	6.926	111.358
Dependent Patent Claims ($DCLAIMS_{it}$)	12,957	11.581	13.488	343	0	343	8	5.385	70.243
NPR Citation Lag Length (NCL_{it})	9,142	1527.042	2270.208	390181.750	-5844	33237.750	730.500	3.392	22.212
Patent Citation Lag Length (PCL_{it})	11,847	676.161	1581.332	35382	-6637	28745	272	4.506	40.053
Forward Citations (FC_{it})	12,957	7.285	14.119	499	0	499	3	7.389	146.054
Forward Citation Lag Length (FCL_{it})	11,649	807.314	1866.497	18774	-4428	14346	225	2.546	11.133
Publishing (PUB_{it})	12,957	226.988	256.645	817.377	0.184	817.561	59.683	0.882	2.590
University Co-authoring ($COAUTHAC_{it}$)	12,957	32.341	20.565	100	0	100	28.474	1.244	4.287
Co-authoring with Firms ($COAUTHFIRM_{it}$)	12,957	0.975	0.156	1	0	1	1	-6.104	38.265
Firm Funds Co-authored Publications ($FUNDER_{it}$)	12,228	0.800	0.400	1	0	1	1	-1.497	3.240

Notes to Panel A: This table presents the descriptive statistics of the key variables prior to transformation. The first column shows the independent and control variables and associated names. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is total assets at the date of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets) at the date of patent filing, $INVENTORS_{it}$ is the number of inventors listed on the front of the patent document, NPR_{it} is the number of NPR cited in the references cited section of the patent document, $CLAIMS_{it}$ is the total number of claims listed on the front page of the patent document, $ICLAIMS_{it}$ is number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the number of dependent claims listed at the end of the patent document, NCL_{it} is the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent, PCL_{it} is the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent, FC_{it} is the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, FCL_{it} is the time period, in days, from the filing date of the citing patent to the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles, $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing, $COAUTHFIRM_{it}$ is 1 when the firm co-authors scientific journal articles with another firm or 0 otherwise and $FUNDER_{it}$ is 1 when the firm funds co-authored scientific journal articles or 0 otherwise. The second column shows the number of observations and the remaining columns display sample descriptive statistics.

Panel B: Descriptive Statistics of Variables After Transformation (Continuation of Table 5.1)									
	N	**Mean**	**Std. Dev.**	**Range**	**Minimum**	**Maximum**	**Median**	**Skewness**	**Kurtosis**
Firm Age (AGE_{it})	12,957	80.966	45.148	165	1	166	85	-0.179	2.071
Firm FDA Experience ($FDAEXP_{it}$)	12,957	0.868	0.338	1	0	1	1	-2.177	5.738
Total Assets ($SIZE_{it}$)	12,957	7.808	1.859	15.589	-2.477	13.112	8.066	-0.986	4.023
Return on Assets (ROA_{it}) (Winsorised)	12,957	0.063	0.139	0.575	-0.363	0.212	0.098	-1.916	5.985
Number of Inventors ($INVENTORS_{it}$)	12,957	0.825	0.602	2.89	0	2.890	0.693	0.281	2.751
Non-patent References (NPR_{it})	12,957	1.479	1.331	6.246	0	6.246	1.386	0.607	2.510
Patent Claims ($CLAIMS_{it}$)	12,957	2.253	0.919	5.881	0	5.881	2.303	-0.404	3.190
Independent Patent Claims ($ICLAIMS_{it}$)	12,957	0.575	0.670	4.419	0	4.419	0.693	1.021	3.656
Dependent Patent Claims ($DCLAIMS_{it}$)	12,957	2.103	0.990	5.841	0	5.841	2.197	-0.442	2.933
NPR Citation Lag Length (NCL_{it})	7,244	7.128	0.925	4.511	5.901	10.411	6.999	0.388	2.495
Patent Citation Lag Length (PCL_{it})	8,039	6.287	1.367	10.266	0	10.266	6.415	-0.719	4.426
Forward Citations (FC_{it})	12,957	1.396	1.139	6.215	0	6.215	1.386	0.489	2.494
Forward Citation Lag Length (FCL_{it})	7,314	6.499	1.480	9.571	0	9.571	6.613	-0.685	3.886
Firm Publishing (PUB_{it})	12,957	226.988	256.645	817.377	0.184	817.561	59.683	0.882	2.590
$COAUTHAC_{it}$ (Sqrt)	12,957	0.540	0.179	1	0	1	0.534	0. 199065	3.625
Co-authoring with Firms ($COAUTHFIRM_{it}$)	12,957	0.975	0.156	1	0	1	1	-6.104	38.265
Firm Funds Co-authored Publications ($FUNDER_{it}$)	12,228	0.800	0.400	1	0	1	1	-1.497	3.240

Notes to Panel B: This table presents the descriptive statistics of the key variables after to transformation. The first column shows the independent and control variables and associated names. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets at the date of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5^{th} and 95^{th} percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $CLAIMS_{it}$ is the natural log of the total number of claims listed on the front page of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, NCL_{it} is the natural log of the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent, PCL_{it} is the natural log of the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, FCL_{it} is the natural log of the time period, in days, from the filing date of the citing patent to the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles, $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing, $COAUTHFIRM_{it}$ is 1 when the firm co-authors scientific journal articles with another firm or 0 otherwise and $FUNDER_{it}$ is 1 when the firm funds co-authored scientific journal articles or 0 otherwise. The second column shows the number of observations and the remaining columns display sample descriptive statistics.

In general, the review suggests that the data from Compustat and Google Patents are consistent with the data in source documents. The results suggest that the data available on Compustat reflect the data in annual reports and the data on Google Patents reflect the data in patent documents. Extreme observations appear are genuine and not due to data integrity issues. Nonetheless, these extreme observations cannot be left in the sample as they can potentially distort statistical inferences. Panel B of Table 5.1 explores different techniques for eliminating outliers.

The normal probability, histogram plots (which for conciseness are not presented here), the skewness, and kurtosis statistics all indicate that many of the variables in the study sample have non-normal distributions, containing several outliers. This may affect the reliability of the results derived from models employing these variables. Their existence could be due to several possible factors, incorrect data entry, the nature of the variable itself or the data exhibits accurate values that are quite remote from the rest of the observations. In this study, the data outliers are accurate but remote observations. These outliers may cause potential problems in the analyses by influencing the results, generating bias and suggesting a relationship between some of the variables when none exists. Some of the variables in the study are in the form of proportions or percentages, these variable types tend to be skewed and so they could affect the precision of the study. Excluding the dichotomous variables (FDAEXP, COAUTHFIRM and FUNDER), it can be observed from Table 5.1, Panel A that nearly all the firm-level variables in the sample, with the exception of firm age at patent filing (AGE) and firm publishing (PUB), exhibit significant skewness and kurtosis and some variables exhibit several data outliers.

There were also extreme values detected in the patent-related variables, namely the prior art citation variables, for the patent citation lag length (PCL), non-patent reference citation lag length (NCL) and forward citation lag length (FCL). For example, the minimum (maximum) value for the PCL is -6,637 days (28,745 days), -5,844 days (33,237.75 days) for the NCL and -4428 days (14346 days) for the FCL.

The NCL is the period, in days, from the citing patent filing year to the publication year of the latest cited publication. The PCL is the period, in days, from the citing patent filing date to the grant/publication date of the latest cited patent. FCL is the period, in days, from the filing date of the citing patent to the grant/publication date of the cited patent. The negative values for these citation-related variables are created by the nature of the patent examining

process in the U.S.. In the U.S. a patent is filed with the USPTO and prior to being granted is reviewed by a patent examiner to determine if its content is patentable. Inventors, patent attorneys and examiners are responsible for ensuring that all relevant patent and non-patent prior art material is referenced within the patent under review. If a relevant patent or piece of non-patent material is published after the filing date, but prior to the grant date of the patent under examination, under U.S. patent law this reference must be added to the patent's references cited section as relevant prior art. As the reference cited occurs after the filing date of the patent being reviewed this creates a negative value for these two lag measures (NCL and PCL). For the FCL variable negative values are also created by the patent examination process. Negative lag values are created where the citing patent is filed prior to the grant date of the cited patent. This occurs if during the examination process for the citing patent either the inventor, patent attorney or examiner recognises that a patent application granted after the citing patent's filing date is relevant prior art. This patent has to be referenced under patent law. For example patent A is granted on 28th May 1996. Another patent, patent B, has a filing date of 13th April 1994 but is not granted until 3rd July 2000. During the examination period for patent B, patent A is determined to be relevant prior art and has to be cited. This creates a negative FCL measure as the citing patent's filing date occurs before the cited patent's grant date.

The means and standard deviations of the citation lag-related variables are seriously affected by such observations. These distributional characteristics and extreme data values necessitate consideration of applying some form of outlier treatment.

Various techniques can be employed to deal with non-normality issues. In the literature on innovation, studies do not generally use winsorising or trimming, but instead tend to transform data using a logarithmic transformation technique (see for example Bollinger and Chandra, 2001 and Kotharia, Sabino and Zach, 2005). Previous studies do not suggest any consistent procedure for the detection and treatment of outliers. Therefore, to limit the effect of outliers and distributional non-normality, specifically following Howell (2007, pp. 318-324) and Tabachnick and Fidell (2007, pp. 86-89), positively skewed variables were transformed using the natural logarithmic function. Not all variables require outlier treatment, only those that show extreme observations are logged or winsorised depending on the nature of the extreme observations. Following prior studies (e.g. Fabrizio, 2009; Fischer and Ringler, 2014 and Gao and Chu, 2015) the firm age variable was not transformed. In fact this variable, along with firm publishing, generally has fewer extreme

values and does not require transformation. The dichotomous variables, prior FDA experience, co-authoring with firms and funder variables, were not transformed. The remaining patent-related and control variables were transformed as their distributions departed from normality.

Panel A in Table 5.1 reports the statistics prior to any data transformation and Panel B displays the statistics after any necessary transformation. The statistics that are reported and discussed for the control, patent and publishing variables are values before and after logarithmic transformation. For each variable, the mean, standard deviation, skewness, kurtosis, minimum and maximum values are reported. With the exception of firm age, return on assets (profitability measure), firm publishing and a firm's university co-authoring percentage, the distributions and distributional statistics based on a natural logarithmic transformation of the variables produced variable distributions closer to normality than those based on the actual levels of all the variables. The firm age, prior FDA experience, firm publishing, co-authoring with firms and funder variables were not subject to any form of data transformation as no apparent extreme values are observed. This approach is supported by prior literature (Coombs and Deeds, 2000; Gittelman and Kogut, 2003; Nagaoka, 2005; Fabrizio, 2009; Hirshleifer, Hsu and Li, 2013 and Novelli, 2015). Prior FDA experience, co-authoring with firms and funder variables were not subject to any form of data transformation as they are binary indicator variables and therefore have no extreme values. Whilst the natural logarithmic transformation of the profitability variable (return on assets) produces an improvement in the variable's distribution, there are still issues with skewness. I therefore adopt a winsorising approach (which involves the replacement of extreme values with threshold values) as opposed to an elimination approach (which eliminates extreme values) as it does not result in the loss of data. I firstly applied a 1st and 99th percentile winsorising approach, but this still yields a data distribution that exhibits some extreme values. I therefore winsorised at 5% at each tail (affecting 1,218 observations). As shown in Panel B, this approach appears to improve the distribution of the profitability variable. I also considered subjecting the university co-authoring variable to a square root transformation (reported in Panel B of Table 5.1) to bring its distribution closer to normality but, following prior literature this variable was not subjected to such an operation (see Fabrizio, 2009; Subramanian, Lim and Soh, 2013 and Soh and Subramanian, 2014).

Following the necessary transformations, the overall statistics of the variables have considerably improved. When comparing Panels A and B and viewing the mean, standard

deviation, skewness and kurtosis statistics of the independent and control variables, it appears to indicate that the outlier and non-normality issues are now minimal. Variables such as the total assets and NPR exhibited extreme values, which after transformation reduces their standard deviation from 15,531.08 and 26.14 to 1.85 and 1.33, respectively. In addition, the mean values are less extreme and much nearer to the median values. This process has substantially reduced the level of skewness in the distribution of the variables and has reduced the occurrence of extreme values. All transformed variables now have skewness statistics below 3. The distribution of the variables is likely to be closer to normal leading to results, from any empirical analysis that are more reliable. However, while steps have been taken to eliminate extreme values, their presence does not necessarily pose any problems to the multivariate analysis as the model specification (i.e., the logit probability model) does not assume that the independent variables are normally distributed (Cox, 1970; Press and Wilson, 1978; and Lo, 1986). The data transformation results in Panel B are therefore adopted for the rest of the analysis in this study.

As can be seen in Table 5.1 (Panel A), the average percentage of publications co-authored with universities is 32.34% for the sample of firms, which is comparable to previous studies by Fabrizio (2009), who reports the mean percentage for U.S. pharmaceutical and biotechnology firms of 30.93%. However, the average university co-authoring percentage is lower than the study by Gittelman and Kogut (2003) that reported 66% co-authorship rates for U.S. biotechnology firms. The minimum and maximum values for this variable indicate that there are firms that do not have any university co-authored publications in their publication output, whilst other firms have all their publications in the form of co-authored works. Other variables, including number of claims (CLAIMS), citations to non-patent references (NPR), forward citations (FC) and number of inventors (INVENTORS), show considerable within-variable differences. On average, patents include approximately 14 claims, cite over 11 NPR and receive approximately 7 forward citations. These results are generally consistent with the reported statistics by previous U.S. studies for pharmaceutical and biotechnology firms (Gittelman and Kogut, 2003; Liu, Arthurs, Cullen and Alexander, 2008; Fabrizio, 2009 and Arts and Veugelers, 2012). Thus, the results of the present study can be compared to prior studies investigating the patent characteristics of U.S. firms.

To examine the effect of firm, patent and publishing characteristics on the likelihood of drug approval, two sample subsets are used to construct the final sample. The first dataset consists of only those patents that are linked to approved drugs whilst the second dataset contains

only patents with no approval link. The reasons underlying the combination of the two sample sets are discussed in detail in the previous chapter. Table 5.2 contains data without the application of any data transformation techniques. Table 5.3 displays the two datasets after the application of data transformation techniques. Panel C compares the descriptive statistics (including, mean, standard deviation, skewness, minimum, maximum and median) for each firm-level hypothesis (and proxy) for patents with an approval-linkage (denoted by 1) and patents without such a link (denoted by 0) to highlight any differences, if any exist. Panel A consists of the descriptive statistics of patents linked to approved drugs, Panel B displays the descriptive statistics for the patents with no link to an approved drug and Panel C reports the independent sample t-tests for the differences in mean values of the patent level characteristics of approval and non-approval linked patents.

As can be seen in Table 5.3, approval-linked patents cite on average 27 NPR, which is nearly three times higher (as indicated in Panel C) than the sample of non-approval-linked patents, having an average NPR citation level of 10. The higher non-patent referencing levels indicate that these patents and more closely linked to basic scientific research. These statistics are comparable to the mean of just over 24 citations reported in a previous U.S. study by Arts and Veugelers (2012) for biotechnology firms. Approval-linked patents also utilise NPR that are over 60% younger than those references cited in non-approval-linked patents (the youngest reference cited is 1,004 days old versus 1,555 days old). These patents also cite patent references that are just over 50% younger than those cited by non-approval-linked patents (the youngest reference cited is 331 days old versus 694 days old). This result suggests that approval-linked patents are more likely to use a patent knowledge base that is younger and may be an indicator that these patents are more innovative, which is consistent with the discussions detailed in Chapter 2. Other patent-level characteristics also differ between the two groups of patents. For approval-linked patents the number of total claims is just over 30% higher (19.39 versus 13.65 claims) and the difference for independent claims is even higher with just over 50% more claims listed within the patent (3.50 versus 2.29 claims). The total number of claims is comparable with those reported for U.S. patents by Lanjouw and Schankerman (2001), Palomeras (2003), Su, Chen and Lee (2012) and Arts and Veugelers (2012). Where differences in the two groups are even more marked in the number of citations received. Approval-linked patents receive over double the numbers of citations within a five-year window of the patent grant/publication date (14.30 versus 6.93 citations) and receive their first citation in a just over a quarter of the time experienced by non-approval-linked patents (230 days versus 839 days).

Table 5.2 Comparison of Approval-Linked Patents and Non-Approval-Linked Patents Before Transformation

Panel A: Patents Linked to Drug Approvals									
	N	**Mean**	**Std. Dev.**	**Range**	**Minimum**	**Maximum**	**Median**	**Skewness**	**Kurtosis**
Number of Inventors ($INVENTORS_{it}$)	632	3.483	2.793	17	1	18	3	2.759	12.873
Non-patent References (NPR_{it})	632	27.571	62.444	515	0	515	5	4.274	24.402
Patent Claims ($CLAIMS_{it}$)	632	19.389	18.136	119	1	120	14	2.297	9.768
Independent Patent Claims ($ICLAIMS_{it}$)	632	3.503	4.334	40	1	41	2	3.755	22.095
Dependent Patent Claims ($DCLAIMS_{it}$)	632	15.886	17.181	108	0	108	11	2.373	10.259
NPR Citation Lag Length (NCL_{it})	470	1004.049	2077.039	19723.500	-5844	13879.500	730.500	2.577	14.187
Patent Citation Lag Length (PCL_{it})	574	330.774	1271.400	16121	-4129	11992	97.500	2.602	19.603
Forward Citations (FC_{it})	632	14.297	19.968	160	0	160	7	2.931	14.506
Forward Citation Lag Length (FCL_{it})	607	229.812	1218.790	11848	-3080	8768	-1	1.977	12.198
Panel B: Patents Not Linked to Drug Approvals									
	N	**Mean**	**Std. Dev.**	**Range**	**Minimum**	**Maximum**	**Median**	**Skewness**	**Kurtosis**
Number of Inventors ($INVENTORS_{it}$)	12,325	2.723	1.913	17	1	18	2	2.652	14.822
Non-patent References (NPR_{it})	12,325	10.205	22.451	437	0	437	3	6.283	73.778
Patent Claims ($CLAIMS_{it}$)	12,325	13.654	14.031	357	1	358	10	5.569	74.977
Independent Patent Claims ($ICLAIMS_{it}$)	12,325	2.294	2.480	82	1	83	2	7.350	132.557
Dependent Patent Claims ($DCLAIMS_{it}$)	12,325	11.360	13.235	343	0	343	8	5.692	78.164
NPR Citation Lag Length (NCL_{it})	8,672	1555.387	2276.891	35794.500	-2556.750	33237.750	730.500	3.433	22.523
Patent Citation Lag Length (PCL_{it})	11,273	693.747	1593.548	35382	-6637	28745	282	4.549	40.261
Forward Citations (FC_{it})	12,325	6.925	13.656	499	0	499	2	8.019	171.734
Forward Citation Lag Length (FCL_{it})	11,042	839.060	1890.626	18774	-4428	14346	238	2.528	10.893

Continuation: Table 5.2 Comparison of Approval-Linked Patents and Non-Approval-Linked Patents Before Transformation				
Panel C: Independent Sample t-test for comparing of means Status N Sample Mean				
	Patent Type	**N**	**Sample Mean**	**Mean Difference**
Number of Inventors ($INVENTORS_{it}$)	Approval-Linked	632	3.483	0.763***
	Non-Approval-Linked	12,325	2.720	
Non-patent References (NPR_{it})	Approval-Linked	632	27.571	17.366***
	Non-Approval-Linked	12,325	10.205	
Patent Claims ($CLAIMS_{it}$)	Approval-Linked	632	19.389	5.735***
	Non-Approval-Linked	12,325	13.654	
Independent Patent Claims ($ICLAIMS_{it}$)	Approval-Linked	632	3.503	1.209***
	Non-Approval-Linked	12,325	2.294	
Dependent Patent Claims ($DCLAIMS_{it}$)	Approval-Linked	632	15.886	4.526***
	Non-Approval-Linked	12,325	11.360	
NPR Citation Lag Length (NCL_{it})	Approval-Linked	470	1004.049	-551.338***
	Non-Approval-Linked	8,672	1555.387	
Patent Citation Lag Length (PCL_{it})	Approval-Linked	574	330.774	-362.973***
	Non-Approval-Linked	11,273	693.747	
Forward Citations (FC_{it})	Approval-Linked	632	14.297	7.372***
	Non-Approval-Linked	12,325	6.925	
Forward Citation Lag Length (FCL_{it})	Approval-Linked	607	229.812	-609.248***
	Non-Approval-Linked	11,042	839.060	

Notes: Panels A and B of Table 5.2 display sample descriptive statistics over the sample period from 1974-2014 for the approval and non-approval linked U.S. patents respectively. $INVENTORS_{it}$ is the number of inventors listed on the front of the patent document, NPR_{it} is the number of NPR cited in the references cited section of the patent document, $CLAIMS_{it}$ is the total number of claims listed on the front page of the patent document, $ICLAIMS_{it}$ is number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the number of dependent claims listed at the end of the patent document, NCL_{it} is the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent, PCL_{it} is the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent, FC_{it} is the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent and FCL_{it} is the time period, in days, from the filing date of the citing patent to the grant date of the cited patent. Panel C reports the independent t-test of mean values for firm's approval and non-approval linked patents. *** denotes that p-values are significant at the 1% level.

Table 5.3 Comparison of Approval-Linked Patents and Non-Approval-Linked Patents After Transformation

Panel A: Patents Linked to Drug Approvals									
	N	**Mean**	**Std. Dev.**	**Range**	**Minimum**	**Maximum**	**Median**	**Skewness**	**Kurtosis**
Number of Inventors ($INVENTORS_{it}$)	632	1.023	0.652	2.890	0	2.890	1.099	0.266	3.078
Non-patent References (NPR_{it})	632	1.967	1.660	6.246	0	6.246	1.792	0.415	2.146
Patent Claims ($CLAIMS_{it}$)	632	2.575	0.947	4.787	0	4.787	2.079	-0.495	3.305
Independent Patent Claims ($ICLAIMS_{it}$)	632	0.843	0.828	3.714	0	3.714	0.693	0.793	3.051
Dependent Patent Claims ($DCLAIMS_{it}$)	632	2.326	1.115	4.691	0	4.691	2.326	-0.586	2.860
NPR Citation Lag Length (NCL_{it})	326	7.012	0.872	3.638	5.901	9.538	6.999	0.587	2.857
Patent Citation Lag Length (PCL_{it})	319	6.172	1.375	9.392	0	9.392	6.319	-0.937	5.152
Forward Citations (FC_{it})	632	2.065	1.192	5.081	0	5.081	2.079	-0.008	2.307
Forward Citation Lag Length (FCL_{it})	301	6.232	1.396	7.980	1.099	9.079	6.471	-0.961	4.177
Panel B: Patents Not Linked to Drug Approvals									
	N	**Mean**	**Std. Dev.**	**Range**	**Minimum**	**Maximum**	**Median**	**Skewness**	**Kurtosis**
Number of Inventors ($INVENTORS_{it}$)	12,325	0.815	0.597	2.890	0	2.890	0.693	0.273	2.708
Non-patent References (NPR_{it})	12,325	1.454	1.307	6.082	0	6.082	1.386	0.593	2.460
Patent Claims ($CLAIMS_{it}$)	12,325	2.236	0.915	5.881	0	5.881	2.303	-0.409	3.192
Independent Patent Claims ($ICLAIMS_{it}$)	12,325	0.561	0.658	4.419	0	4.419	0.693	1.013	3.610
Dependent Patent Claims ($DCLAIMS_{it}$)	12,325	2.092	0.982	5.841	0	5.841	2.197	-0.443	2.942
NPR Citation Lag Length (NCL_{it})	6,918	7.133	0.928	4.511	5.901	10.411	6.999	0.379	2.482
Patent Citation Lag Length (PCL_{it})	7,720	6.291	1.367	10.266	0	10.266	6.417	-0.710	4.393
Forward Citations (FC_{it})	12,325	1.362	1.125	6.215	0	6.215	1.099	0.512	2.530
Forward Citation Lag Length (FCL_{it})	7,013	6.510	1.482	9.571	0	9.571	6.620	-0.679	3.874

Continuation: Table 5.3 Comparison of Approval-Linked Patents and Non-Approval-Linked Patents After Transformation				
Panel C: Independent Sample t-test for comparing of means Status N Sample Mean				
	Patent Type	**N**	**Sample Mean**	**Mean Difference**
Number of Inventors ($INVENTORS_{it}$)	Approval-Linked	632	1.023	0.208***
	Non-Approval-Linked	12,325	0.815	
Non-patent References (NPR_{it})	Approval-Linked	632	1.967	0.514***
	Non-Approval-Linked	12,325	1.454	
Patent Claims ($CLAIMS_{it}$)	Approval-Linked	632	2.575	0.339***
	Non-Approval-Linked	12,325	2.236	
Independent Patent Claims ($ICLAIMS_{it}$)	Approval-Linked	632	0.843	0.282***
	Non-Approval-Linked	12,325	0.561	
Dependent Patent Claims ($DCLAIMS_{it}$)	Approval-Linked	632	2.326	0.234***
	Non-Approval-Linked	12,325	2.092	
NPR Citation Lag Length (NCL_{it})	Approval-Linked	326	7.012	-0.121**
	Non-Approval-Linked	6,918	7.133	
Patent Citation Lag Length (PCL_{it})	Approval-Linked	319	3.172	-0.120
	Non-Approval-Linked	7,720	6.291	
Forward Citations (FC_{it})	Approval-Linked	632	2.065	0.703***
	Non-Approval-Linked	12,325	1.362	
Forward Citation Lag Length (FCL_{it})	Approval-Linked	301	6.232	0.278***
	Non-Approval-Linked	7,013	6.510	

Notes: Panels A and B of Table 5.3 display sample descriptive statistics over the sample period from 1974-2014 for the approval and non-approval linked U.S. patents respectively. $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $CLAIMS_{it}$ is the natural log of the total number of claims listed on the front page of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, NCL_{it} is the natural log of the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent, PCL_{it} is the natural log of the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent and FCL_{it} is the natural log of the time period, in days, from the filing date of the citing patent to the grant date of the cited patent. Panel C reports the independent t-test of mean values for firm's approval and non-approval linked patents. *** denotes that p-values are significant at the 1% level.

In some areas the differences between the two groups is not so marked. The number of inventors listed for approval-linked patents is just under 30% higher than for the non-approval group (average of 3.48 versus 2.72 inventors). This is not unexpected, as firms will have individual research teams that produce multiple patents, some of which lead to approved drugs and some that do not. Reported inventor numbers are comparable with other U.S. pharmaceutical and biotechnology patent studies, such as Gittelman and Kogut (2003) and Arts and Veugelers (2012).

Although both patent groups contain the same sample firms there are still some observed differences. In the approval-linked patent sample, more patents are held by younger firms reducing the average firm age by 15% (69.9 years old versus 81.5 years old). In addition, many of the patents in this sample are held by larger firms, 60% larger in terms of total assets ($7,264.43m versus $11,686.96m). However, for some variables the average firm-level differences are not as notable. The mean values for firm publishing levels and university co-authoring activity are virtually identical with the approval-linked group having average publications of 226.5 publications per year, of which 32.8% are co-authored with university-based authors. For the non-approval-linked group these figures are 227 and 32.3% respectively.

In Table 5.3, Panel C the results (after data transformation) also show that approval patents are significantly different from non-approval patents in all but one of the patent characteristics, patent citation lag length (PCL). Although the means of the two groups differ, this difference is statistically insignificant. This indicates that both the groups have a somewhat similar patent citation pattern in terms of the age of the patents cited. The logarithmic transformation of this data item transforms its extreme values, which may reduce the differences in the age of patents cited within the two groups. Patent citation has a more legal role than non-patent citation, this would mean that patents for work carried out in similar areas would have to reference the same patent prior art. If they did not do so the patent cannot be granted as it would not be providing legal proof of its novelty, one of the three characteristics a patent application must possess to be granted.

5.3 Checking for Collinearity

When independent variables in a regression model are correlated with each other, this condition is referred to as collinearity. An important assumption for logistic regression is that multi-collinearity does not exist (Gujarati and Porter, 2009). Severe multi-collinearity

can cause any estimates to be sensitive to small alterations in the model. The standard errors of the regression coefficients are also large if multi-collinearity is an issue. The result is that the odds ratio estimates can be unstable and more difficult to interpret, making it more difficult to specify the model correctly.

To test the assumption of variable independence and detect any multi-collinearity issues within the sample, different tests are conducted such as examining correlation coefficient matrices, computing Eigenvalues[66] and variance inflation factors (VIF) and these are discussed later in this chapter. Standard correlation methods, such as Pearson product-moment correlation, assume that variables are continuous and follow a multivariate normal distribution. As the non-binary variables in this study exhibit a non-normal distribution Spearman non-parametric correlation coefficients are also calculated for these variables

The matrix with Pearson product moment correlation coefficients (parametric) for transformed non-binary independent and control variables is displayed in Table 5.4, with the matrix of Spearman's rho coefficients (non-parametric) displayed in Table 5.5 and for approval-linked patents in Table 5.6. As can be seen in Tables 5.4 and 5.5, a few of the coefficients indicate the existence of collinearity issues and there is a significant correlation between most of the variables. However, in most cases, the level of correlation appears to be modest (i.e., the Pearson and Spearman correlation coefficients are close to zero). This level of correlation is unlikely to create substantial issues of multi-collinearity. There are exceptions, with total assets having a highly positive and significant correlation with firm age (that is values of 0.6086[67] and 0.5545 for Pearson and Spearman coefficients, respectively). Return on assets is moderately correlated with age, having a Pearson coefficient value of 0.5350 (0.3687 Spearman coefficient). This is not unexpected as total assets and profitability tend to increase with firm age.

Total claims and dependent claims (DCLAIMS) are nearly perfectly positively correlated at 0.9513 and 0.9676 (Pearson and Spearman coefficients, respectively) but this can be attributed to the fact that dependent claims are a subset of total claims, with all but one of the total claims being a dependent claim. This does not create problems in the analysis, as

[66] Eigenvalues measure of the amount of variance contained in the correlation matrix so that the sum of the eigenvalues is equal to the number of variables.
[67] Prior literature uses various cut-off values for collinearity e.g. Liu (2014) uses -/+0.70 in line with that suggested by Cohen, Cohen, West and Aiken (2003).

these two variables are not jointly included in any of the models discussed later in this chapter. The two variables measuring forward citations and their related lags are highly negatively correlated. This negative correlation implies that patents that receive higher than average levels of forward citations also start to receive them sooner than the average patent. For some of the variables exhibiting high levels of correlation, this is because they are linked by calculation or derived from similar sources e.g. return on assets, total assets, claims-related, forward citation-related and the publication-related variables. In Table 5.4, it can be observed that the variables return on assets and total assets show a moderately high correlation value of 0.6667, this is not an unexpected result as the return on assets variable is computed using the total assets variable.

In the Spearman correlation matrix, there are two instances of high correlations, these involve claims and dependent claims (coefficient of 0.9676) and forward citations and its related lag (coefficient of -0.7543) and one instance of moderate correlation involving firm size and age (coefficient 0.5545). All of these correlations are significant at the 1% level. Most of the other variables in the Spearman matrix are correlated at the 1% or 5% level, but the coefficients mostly lie within the -/+0.30 range, which seems an acceptable range of values.

Pearson and Spearman correlation coefficients were also calculated for all the variables, including the binary variables, although not displayed. The results are similar to those of Tables 5.4 and 5.5 with high correlations for firm size and age, claims and dependent claims, forward citations and their related lag. A firm's prior experience of the FDA approval process is significant and highly correlated with firm size (Spearman coefficient of 0.5066 and a Pearson coefficient of 0.7070). It is also highly correlated with profitability, having a Pearson coefficient of 0.7470. These results are not unexpected as industry experience and profitability often tend to become greater as a firm ages and increases in size.

As a robustness check, Pearson and Spearman correlation coefficients were calculated for non-binary variable pairs. Coefficients were similar in sign, magnitude and significance to those in Tables 5.4 and 5.5 with age and total assets, patent claims and dependent claims, forward citations and the related cite lag being greater than -/+0.5. Return on assets and age and return on assets and totals assets had Pearson coefficients greater than -/+0.5.

Table 5.4 Pearson Correlation Matrix of Dependent, Independent and Control Variables (Transformed)

	AGE_{it}	$SIZE_{it}$	ROA_{it}	$INVENTORS_{it}$	NPR_{it}	$CLAIMS_{it}$	$ICLAIMS_{it}$	$DCLAIMS_{it}$	NCL_{it}	PCL_{it}	FCit	FCL_{it}	$PUB_{it/}$	$COAUTHAC_{it}$
AGE_{it}	1													
$SIZE_{it}$	0.6086**	1												
ROA_{it}	0.5350**	0.6667**	1											
$INVENTORS_{it}$	-0.0564*	0.0800**	-0.0259	1										
NPR_{it}	-0.2927**	-0.1660**	-0.3553**	0.1843**	1									
$CLAIMS_{it}$	-0.0226	0.0295	-0.0661**	0.0536*	0.1033**	1								
$ICLAIMS_{it}$	-0.1110**	-0.1038**	-0.1361**	0.0884**	0.0964**	0.3676**	1							
$DCLAIMS_{it}$	-0.0002	0.0499*	-0.0504*	0.0377	0.0942**	0.9513**	0.1471**	1						
NCL_{it}	0.1443**	0.0648**	0.1322**	-0.1173**	-0.3801**	-0.0845**	-0.0513*	-0.0732**	1					
PCL_{it}	-0.0195	-0.0467	-0.0189	-0.0591*	-0.0438	-0.0477	-0.0075	-0.0545*	0.1452**	1				
FCit	-0.1274**	-0.0507*	-0.1916**	0.1440**	0.2417**	0.2147**	0.1105**	0.1982**	-0.1582**	-0.0736**	1			
FCL_{it}	0.0502*	0.0096	0.0922**	-0.0550*	-0.1259**	-0.1521**	-0.0498*	-0.1472**	0.1139**	0.0757**	-0.6654**	1		
$PUB_{it/}$	0.4427**	0.3288**	0.3735**	0.0199	-0.0315	-0.0661**	-0.0873**	-0.0497*	-0.0203	-0.0508*	-0.0386	0.0029	1	
$COAUTHAC_{it}$	-0.0675**	-0.0867**	-0.0242**	-0.0286	0.1020**	0.0126	0.0817**	-0.0050	0.0090	0.0086	0.0708**	-0.0273	-0.2460**	1

**Correlation is significant at the 0.01 level * Correlation is significant at the 0.05 level

Notes: The table shows Pearson product-moment correlation coefficients between the independent variables in the study. Variables are defined as Firm Age (AGE_{it}), Total Assets ($SIZE_{it}$), Return on Assets (ROA_{it}), Number of Inventors ($INVENTORS_{it}$), NPR (NPR_{it}), Patent Claims ($CLAIMS_{it}$), Independent Patent Claims ($ICLAIMS_{it}$), Dependent Patent Claims ($DCLAIMS_{it}$), NPR Citation Lag Length (NCL_{it}), Patent Citation Lag Length (PCL_{it}), Forward Citations Received (5 year Window) (FC_{it}), Forward Citation Lag Length (FCL_{it}), Firm Publishing (PUB_{it}) and University Co-authoring Percentage ($COAUTHAC_{it}$).

Table 5.5 Spearman's Correlation Matrix of Non-Binary Independent and Control Variables (Transformed)

	AGE_{it}	$SIZE_{it}$	ROA_{it}	$INVENTORS_{it}$	NPR_{it}	$CLAIMS_{it}$	$ICLAIMS_{it}$	$DCLAIMS_{it}$	NCL_{it}	PCL_{it}	FCit	FCL_{it}	PUB_{it}	$COAUTHAC_{it}$
AGE_{it}	1													
$SIZE_{it}$	0.5545**	1												
ROA_{it}	0.3687**	0.4090**	1											
$INVENTORS_{it}$	-0.0455*	0.1427**	0.0576**	1										
NPR_{it}	-0.1851**	-0.0013	-0.1678**	0.1737**	1									
$CLAIMS_{it}$	-0.0065	0.0414*	-0.0622**	0.0546**	0.1067**	1								
$ICLAIMS_{it}$	-0.1040**	-0.0573**	-0.0930**	0.0807**	0.0746**	0.3469**	1							
$DCLAIMS_{it}$	0.0143	0.0571*	-0.0498**	0.0405*	0.1025**	0.9676**	0.1690**	1						
NCL_{it}	0.1181**	0.0279	0.0387*	-0.1185**	-0.4042**	-0.0826**	-0.0512**	-0.0737**	1					
PCL_{it}	-0.0212	-0.0626*	-0.0490**	-0.0647**	-0.0616**	-0.0582**	-0.0110	-0.0625**	0.1796**	1				
FC_{it}	-0.0852**	0.0422*	-0.0666**	0.1164**	0.2281**	0.2167**	0.0972**	0.2065**	-0.1564**	-0.0790**	1			
FCLit	0.0384*	-0.0416*	0.0301	-0.0579**	-0.1392**	-0.1619**	-0.0554**	-0.1598**	0.1221**	0.0863**	-0.7543**	1		
PUB_{it}	0.4182**	0.3208**	0.4188**	0.0625**	-0.0135	-0.0756**	-0.0638**	-0.0648**	-0.0095	-0.0374*	-0.0158	0.0021	1	
$COAUTHAC_{it}$	0.0496**	0.0691**	-0.2714**	-0.0557**	0.1132	0.0255	0.0679**	0.0167	0.0152	0.0024	0.0750**	-0.0471**	-0.0618**	1

**Correlation is significant at the 0.01 level * Correlation is significant at the 0.05 level

Notes: The table shows Spearman's rho correlation coefficients between the independent variables in the study. Variables are defined as Firm Age (AGE_{it}), Total Assets ($SIZE_{it}$), Return on Assets (ROA_{it}), Number of Inventors ($INVENTORS_{it}$), NPR (NPR_{it}), Patent Claims ($CLAIMS_{it}$), Independent Patent Claims ($ICLAIMS_{it}$), Dependent Patent Claims ($DCLAIMS_{it}$), NPR Citation Lag Length (NCL_{it}), Patent Citation Lag Length (PCL_{it}), Forward Citations Received (5 year Window) (FC_{it}), Forward Citation Lag Length (FCL_{it}), Firm Publishing (PUB_{it}) and University Co-authoring Percentage ($COAUTHAC_{it}$)

Table 5.6 Spearman's Correlation Matrix of Non-Binary Independent and Control Variables (Transformed): Patents Linked to Drug Approvals Only

	AGE_{it}	$SIZE_{it}$	ROA_{it}	$INVENTORS_{it}$	NPR_{it}	$CLAIMS_{it}$	$ICLAIMS_{it}$	$DCLAIMS_{it}$	NCL_{it}	PCL_{it}	FC_{it}	FCLit	PUB_{it}	$COAUTHAC_{it}$
AGE_{it}	1													
$SIZE_{it}$	0.5775**	1												
ROA_{it}	0.5299**	0.5939**	1											
$INVENTORS_{it}$	0.1101	0.3220**	0.2649**	1										
NPR_{it}	-0.3974**	-0.1674	-0.4090**	0.2141**	1									
$CLAIMS_{it}$	-0.1704	-0.0920	-0.1722	-0.0447	-0.0425	1								
$ICLAIMS_{it}$	-0.2116*	-0.0061	-0.1383	-0.0207	0.1270	0.3627**	1							
$DCLAIMS_{it}$	-0.1190	-0.0584	-0.1701	0.0224	-0.0081	0.9179**	0.1522	1						
NCL_{it}	0.1885	0.1248	0.1905	-0.1178	-0.4380**	0.1965	0.1394	0.1539	1					
PCL_{it}	-0.1137	-0.2000	-0.0983	-0.1313	-0.1661	0.0694	0.1680	0.0334	0.1023	1				
FC_{it}	-0.0408	0.0124	-0.0423	-0.0185	-0.0826	0.0424	0.1052	0.0197	0.0032	0.1186	1			
FCLit	-0.0450	-0.1138	0.0014	0.0981	0.0434	0.0045	-0.0200	-0.0032	0.1590	0.0337	-0.6420**	1		
PUB_{it}	0.7648**	0.3999**	0.5025**	-0.0063	-0.2916**	-0.2944**	-0.1782	-0.2365**	0.0468	-0.0089	0.1394	-0.2240**	1	
$COAUTHAC_{it}$	0.0206	0.0520	-0.0240	-0.0035	-0.0640	-0.0307	0.1518	-0.1114	0.1301	0.0692	0.1696	0.0487	-0.0243	1

**Correlation is significant at the 0.01 level * Correlation is significant at the 0.05 level

Notes: The table shows Spearman's rho correlation coefficients between the independent variables in the study. Variables are defined as Firm Age (AGE_{it}), Total Assets ($SIZE_{it}$), Return on Assets (ROA_{it}), Number of Inventors ($INVENTORS_{it}$), NPR (NPR_{it}), Patent Claims ($CLAIMS_{it}$), Independent Patent Claims ($ICLAIMS_{it}$), Dependent Patent Claims ($DCLAIMS_{it}$), NPR Citation Lag Length (NCL_{it}), Patent Citation Lag Length (PCL_{it}), Forward Citations Received (5 year Window) (FC_{it}), Forward Citation Lag Length (FCL_{it}), Firm Publishing (PUB_{it}) and University Co-authoring Percentage ($COAUTHAC_{it}$).

Correlation matrices do not reveal higher order collinearity. Multi-collinearity can also be detected with the help of tolerance[68] and its reciprocal, called the variance inflation factor (VIF) and are viewed as superior measures of multi-collinearity. These measures allow for the possibility that one independent variable can be a function of two or more other independent variables (Brooks, 2008). The tolerance assesses how much multi-collinearity can be tolerated in the model, whilst the VIF measures the proportion of the inflation in standard errors results from the presence of multi-collinearity. To assess the potential model implications of multi-collinearity, I calculated the tolerance and variance inflation factors (VIFs) for each coefficient.

Table 5.7 Tolerance and VIF

Variable Name	Tolerance	VIF
Firm Age (AGE_{it})	0.415	2.410
Firm FDA Experience Prior to Patent Filing ($FDAEXP_{it}$)	0.321	3.120
Total Assets ($SIZE_{it}$)	0.316	3.160
Return on Assets (ROA_{it})	0.353	2.830
Number of Inventors ($INVENTORS_{it}$)	0.925	1.080
Non-patent References (NPR_{it})	0.665	1.500
Patent Claims ($CLAIMS_{it}$)	0.040	24.930
Independent Patent Claims ($ICLAIMS_{it}$)	0.423	2.370
Dependent Patent Claims ($DCLAIMS_{it}$)	0.045	22.130
NPR Citation Lag Length (NCL_{it})	0.830	1.200
Patent Citation Lag Length (PCL_{it})	0.962	1.040
Forward Citations Received (five-year Window) (FC_{it})	0.509	1.970
Forward Citation Lag Length (FCL_{it})	0.557	1.790
Firm Publishing (PUB_{it})	0.629	1.590
University Co-authoring Percentage ($COAUTHAC_{it}$)	0.804	1.240
Firm co-authors with other firms ($COAUTHFIRM_{it}$)	0.881	1.140
Firm funds co-authored publications ($FUNDER_{it}$)	0.718	1.390

Notes: The table shows tolerance and variance inflation factors (VIFs) for all the variables in this study. The results show that all VIFs, except for CLAIMS and DCLAIMS, are below the threshold of 10 (see Gujarati and Porter, 2009). AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets at the date of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5th and 95th percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $CLAIMS_{it}$ is the natural log of the total number of claims listed at the end of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, NCL_{it} is the natural log of the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent, PCL_{it} is the natural log of the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, FCL_{it} is the natural log of the time period, in days, from the filing date of the

[68] Tolerance values approaching zero indicate that the variable is highly collinear with the other independent variables.

citing patent to the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles, $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing, $COAUTHFIRM_{it}$ is 1 when the firm co-authors scientific journal articles with another firm or 0 otherwise and $FUNDER_{it}$ is 1 when the firm funds co-authored scientific journal articles or 0 otherwise.

Condition indices are another measure employed to assist in the detection of multi-collinearity. Condition indices between 30 and 100 indicate moderate to strong collinearity. Condition indices between 10 and 30 are acceptable (Gujarati and Porter, 2009, pp.343-44). These formal tests of multi-collinearity, VIF and eigenvalues, for the sample are displayed in Tables 5.7 and 5.8 respectively. The VIFs for the majority of variables are low (below 3.00), with tolerances also displaying high values (above 0.30) suggesting that these variables do not suffer from multi-collinearity. These values indicate that the level of multi-collinearity is, perhaps, not a problem and suggests that the level of multi-collinearity in the model is modest. Only the patent claims (from now on referred to as claims) and dependent patent claims (from now on referred to as dependent claims) variables exhibit inflated VIF values. The VIFs for these variables are very high, exhibiting values that are well over 10[69], with tolerances approaching zero and eigenvalues close to zero (Gujarati and Porter, 2009 p.340). The standard errors and the variances of the estimated coefficients of a regression model are inflated when multi-collinearity exists, which reduces the significance of the independent variables. The VIF for an estimated coefficient quantifies how much the variance is inflated. The VIF for the predictor claims, for example, indicates that the variance of the estimated coefficient of claims is inflated by a factor of 24.93; this is because the claims variable is highly correlated with at least one of the other predictors in the model. Also, eigenvalues near 0 and Condition Index above 30 (Hall et al., 2010 p.7) also suggest that claims and dependent claims have high multi-collinearity issues.

One solution to multi-collinearity is to delete one or more of the independent variables from the regression equation. As discussed earlier in this chapter there is no model formulation in this study that would have included both claims and dependent claims as this would involve the double-counting of data as the dependent claims variable is a subset of claims.

[69] If the VIF is equal to 1 there is no multi-collinearity among factors, but if the VIF is greater than 1, the predictors may be moderately correlated. The output above shows that the VIF for most of the variables are below 3, which indicates some correlation, but not grounds for concern. A VIF between 5 and 10 indicates high correlation that may be problematic. A VIF above 10 indicates that the regression coefficients may be poorly estimated due to multi-collinearity. Pan and Jackson (2008) recommended a VIF of less than 5, whilst Rogerson (2001) recommended a VIF of less than 4.

Table 5.8 Eigenvalues and Conditional Index

Dimension	Eigenvalue	Condition Index
1	14.2070	1.0000
2	1.1773	3.4738
3	0.5141	5.2569
4	0.4361	5.7077
5	0.3823	6.0963
6	0.3366	6.4969
7	0.2520	7.5080
8	0.1857	8.7463
9	0.1628	9.3424
10	0.1349	10.2620
11	0.0967	12.1217
12	0.0396	18.9465
13	0.0310	21.3950
14	0.0182	27.9097
15	0.0116	34.9471
16	0.0085	40.8760
17	0.0033	65.7843
18	0.0023	79.0851

Notes: The table shows multicollinearity diagnostics for all the variables in this study. Eigenvalues near 0 and Condition Indices greater than 15 suggests that the variables CLAIMS and DCLAIMS have high multicollinearity issues.

I therefore made the decision to remove the claims variable as this study is concerned with observing the two different types of claims (independent and dependent claims) data separately. When this was implemented the VIF, see Table 5.9, for dependent claims dropped to 1.08 with most of the other variables having VIFs below 2, the four remaining having values below 3.2. Condition indices are another measure used to detect multi-collinearity. Once the claims variable was removed the condition index, see Table 5.10, had a value of 76[70] which is higher than would be desirable.

Overall, the collinearity statistics (including the parametric and non-parametric correlation matrices, condition indices, eigenvalues and VIF statistics) suggest that the levels of multi-collinearity in the variables appear to be statistically acceptable. Overall, the analyses suggest that any remaining multi-collinearities and non-normalities in the variables are not so severe as to cause serious issues in the regression models. The following subsections

[70] Condition indices between 30 and 100 indicate moderate to strong collinearity. Condition indices between 10 and 30 are acceptable (Gujarati and Porter, 2009, pp.343-44).

report the empirical results from the regressions of these variables on the likelihood of receiving an FDA drug approval.

Table 5.9 Revised Tolerance and VIF

Variable Name	Tolerance	VIF
Firm Age (AGE_{it})	0.415	2.410
Firm FDA Experience Prior to Patent Filing ($FDAEXP_{it}$)	0.321	3.110
Total Assets ($SIZE_{it}$)	0.316	3.160
Return on Assets (ROA_{it})	0.353	2.830
Number of Inventors ($INVENTORS_{it}$)	0.925	1.080
Non-patent References (NPR_{it})	0.665	1.500
Independent Patent Claims ($ICLAIMS_{it}$)	0.945	1.060
Dependent Patent Claims ($DCLAIMS_{it}$)	0.927	1.080
NPR Citation Lag Length (NCL_{it})	0.831	1.200
Patent Citation Lag Length (PCL_{it})	0.962	1.040
Forward Citations Received (five-year Window) (FC_{it})	0.509	1.960
Forward Citation Lag Length (FCL_{it})	0.557	1.790
Firm Publishing (PUB_{it})	0.629	1.590
University Co-authoring Percentage ($COAUTHAC_{it}$)	0.805	1.240
Firm co-authors with other firms ($COAUTHFIRM_{it}$)	0.881	1.140
Firm funds co-authored publications ($FUNDER_{it}$)	0.718	1.390

Notes: The table shows tolerance and variance inflation factors (VIFs) for all the variables in this study. The results show that all VIFs are below the 10 threshold (see Gujarati and Porter, 2009). AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets at the date of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5th and 95th percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, NCL_{it} is the natural log of the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent, PCL_{it} is the natural log of the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, FCL_{it} is the natural log of the time period, in days, from the filing date of the citing patent to the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles, $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing, $COAUTHFIRM_{it}$ is 1 when the firm co-authors scientific journal articles with another firm or 0 otherwise and $FUNDER_{it}$ is 1 when the firm funds co-authored scientific journal articles or 0 otherwise.

Table 5.10 Revised Eigenvalues and Conditional Index

Dimension	Eigenvalue	Condition Index
1	13.3223	1.0000
2	1.1594	3.3899
3	0.5072	5.1253
4	0.4343	5.5384
5	0.3809	5.9142
6	0.3297	6.3570
7	0.2110	7.9470
8	0.1685	8.8918
9	0.1437	9.6293
10	0.1349	9.9393
11	0.0964	11.7530
12	0.0396	18.3477
13	0.0310	20.7229
14	0.0182	27.0274
15	0.0116	33.8444
16	0.0085	39.5967
17	0.0023	76.1360

Notes: The table shows multicollinearity diagnostics for all the variables in this study. Eigenvalues near 0 and Condition Indices greater than 15 suggests the continuing presence of multicollinearity issues but at a reduced level.

5.4 Drug Approval Prediction Model

As the dependent variable in this study is binary in nature, it is analysed using a logistic regression technique. The model postulates that the probability of a patent being linked to an FDA approved drug is a (logit) function of a vector of 16 patent-related and firm-related variables. The hypotheses and independent variables employed in the approval prediction model are developed in Chapter 3.

The general relationship tested is presented in logistic function as:

$$Log\left(\frac{\rho}{1-\rho}\right) = Logit(\rho) = \sum_{i=1}^{n} \beta_i X_{\text{ i, t}}$$

The above equation shows the general relationship of a set of patent and firm characteristics (X) with the likelihood of a successful FDA drug approval (ρ). In the above logistic function, the probability of success represents the probability that the firm will receive an approval depending on the formulation of the model, $\frac{\rho}{1-\rho}$ represents the odds of ρ (receipt of an approval).

ρ is the drug approval likelihood and $X_{i,t}$ is a vector of 16 variables for each patent. ρ can be computed as the inverse of the logit function – i.e., the logistic function – as shown below.

$$\rho = \text{Logit}^{-1}(\beta_i X_{i,t}) = \frac{1}{1+e-\beta iXi,t}$$

The model coefficients (β) are estimated using the maximum likelihood estimation technique with Stata software. Here, β shows the direction of the relationship and e^{β} shows the odds ratio for each variable in the model. The odds ratio is "the increase (or decrease if the ratio is less than one) in the odds of being in one outcome category when the value of the predictor increases by one unit" (Tabachnick and Fidell, 2001, p. 548).

The interpretation of the odds ratio is straightforward. An odds ratio greater than 1 indicates that patents with such a ratio on an explanatory variable, which is assumed to have a positive relationship with the dependent variable, will have an increased likelihood of being associated with a dependent variable coded 1 for success (approval). An odds ratio of less than 1 indicates that the chances that patents with such a ratio on an explanatory variable will have a reduced likelihood of being associated with a dependent variable coded 1 for success.

5.4.1 Univariate Analysis

A conditional logistic regression with firm fixed effects is employed as the main estimation technique to analyse the factors suspected of influencing the likelihood of receiving an FDA drug approval. When modelling, I consider the influence of selected variables by testing alternative models. The first step in this process is identifying the most significant patent and firm-level characteristics, observing their relationship to the approval outcome and to establish which variables could be possible predictors of drug approval success. This is accomplished by performing a univariate logistic regression for each variable in turn[71] and thereafter applying a multivariate logistic regression analysis approach. Variable definitions are discussed in detail in Chapter 4, Section 4.4.1. Consideration must be given to the fact that statistically significant variables in a univariate model may not necessarily be so in a multivariate model. This can be attributed to the presence of some other key variables in the multivariate model, which may prevent a variable that is significant in the univariate analysis

[71] Statistical analysis was performed with the Stata (version 13.0) statistical package.

from displaying any incremental information content in the multivariate context. This section discusses how individual variables influence the likelihood of a successful drug approval and the univariate regression results are reported in Table 5.11. The values reported in the table represent the coefficients, odds ratios and their statistical significance (standard errors are in parentheses) obtained by regressing each independent variable against the dependent variable (drug approval probability). Section 5.4.2 discusses the multivariate logistic regression process.

Table 5.11 Binary Logistic Univariate Regression Analysis of Patent and Firm-Level Characteristics on the likelihood of Receiving a Drug Approval

Variable	Odds Ratio (eβ)	Coefficient (β)	Log-Likelihood	Coefficient Expected Sign	Sample Size
AGE_{it}	1.034(0.007)***	0.033(0.007)***	-1,956.1	+	12,957
$FDAEXP_{it}$	2.340(0.584)***	0.850(0.250)***	-1,963.5	+	12,957
$SIZE_{it}$	1.528(0.067)***	0.424(0.044)***	-1,919.9	+	12,957
ROA_{it} (Winsorised)	29.13(16.92)***	3.372(0.581)***	-1,952	+	12,957
$INVENTORS_{it}$	1.454(0.112)***	0.375(0.077)***	-1,957.2	+	12,957
NPR_{it}	1.272(0.051)***	0.241(0.040)***	-1,950.8	+	12,957
$ICLAIMS_{it}$	1.512(0.093)***	0.413(0.062)***	-1,947.4	+	12,957
$DCLAIMS_{it}$	1.263(0.061)***	0.233(0.048)***	-1,956.9	+	12,957
NCL_{it}	0.946(0.068)	-0.056(0.072)	-1,024.3	-	6,890
PCL_{it}	0.943(0.042)	-0.058(0.045)	-1,032	-	7,857
FC_{it}	1.585(0.067)***	0.461(0.042)***	-1,908.5	+	12,957
FCL_{it}	0.916(0.039)**	-0.088(0.042)**	-954.2	-	6,981
PUB_{it}	0.925(0.014)***	-0.078(0.015)***	-1,957.3	+	12,957
$COAUTHAC_{it}$	0.973(0.011)**	-0.027(0.012)**	-1,966.2	+	12,957
$COAUTHFIRM_{it}$	$4.60e^{-19}$(0.000)	-42.224(2.07e+08)	-1,961.4	+	12,957
$FUNDER_{it}$	Omitted			+	12,957

Notes: The table presents the results of a conditional logit regression analysis with fixed effects where the dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate) and the control and independent variables are as defined in Chapter 3. Each variable is regressed on the dependent variable individually. For example, the coefficients of total assets (a proxy for firm size) is obtained from regressing total assets as the sole independent variable with patent drug approval linkage probability as the binary dependent variable (with no control variables). AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets at the date of patent filing, ROA_{it} is the firm's return on assets (net

income divided total assets), winsorised at the 5th and 95th percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, NCL_{it} is the natural log of the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent, PCL_{it} is the natural log of the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, FCL_{it} is the natural log of the time period, in days, from the filing date of the citing patent to the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles, $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing, $COAUTHFIRM_{it}$ is 1 when the firm co-authors scientific journal articles with another firm or 0 otherwise and $FUNDER_{it}$ is 1 when the firm funds co-authored scientific journal articles or 0 otherwise. Firm Funds Co-authored Publication (FUNDER) is omitted, as there is no within-group variance. Column 2 displays odds ratios and column 3 coefficients. Standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Column 4 displays the log likelihood of each univariate regression and column 5 the expected sign of the coefficient based on the relevant hypothesis.

The univariate regression results demonstrate that all the control variables (age, FDA experience, size and profitability) have statistically significant relationships with the likelihood of approval at the 1% level with odds ratios greater than 1, this is in line with expectations. As a firm ages, it tends to grow in size and profitability, potentially creating more resources available for investment into its drug development programme. With increased investment, the likelihood is that more research projects will receive funding and that more drug candidates will be created from this expanded range of projects. With a larger suite of drug candidates to select from, more of these candidates are likely to proceed to the final FDA approval stage of the drug development process. As a firm gains more experience of this final approval stage it will learn from its successes and failures. This will inform subsequent research projects and in so doing potentially improve the probability of these projects being successful and gaining FDA approval.

The patent-related variables, number of inventors, NPR, independent patent claims, dependent patent claims and forward citations also have statistically significant relationships with the likelihood of approval at the 1% level with odds ratios greater than 1. Inventor numbers are regarded as an indicator of the level of resources employed in a research project. Larger inventor teams can indicate a research project in receipt of superior funding; indicating the level of importance a firm places in such a project and may indicate the firm's belief that this project is more likely to result in a successful approval outcome. The number of non-patents references listed in a patent document can be an indicator of the patent's link to scientific knowledge and a measure of the innovativeness of a patent. (see, for example, Gittelman and Kogut, 2003 and Arts and Veugelers, 2012). This may indicate that the patent

is more likely to be linked to a more innovative research project, increasing the likelihood of a linkage to a new drug approval.

Some observations are dropped by the logistic regression process because an approval/non-approval (1/0) outcome is perfectly predicted or there is no within-group variance. This results in the funding behaviour of the firm variable being omitted from the regression output. Also, note that co-authoring with other firms has an odds ratio of virtually zero.

5.4.2 Multivariate Analysis: Testing the Impact of Patent and Firm-level Characteristics on the Likelihood of FDA Drug Approval

Table 5.12 presents the logit regression results; Models 1 to 6 are multivariate logit regression models. In Model 1, I first estimated a baseline model using control variables only. Next, I added the effects from hypotheses discussed in Chapter 3. Models 2 to 6 contain only those patent and firm-level characteristics that are available at the patent grant date (or the publication date for patents filed from 30th November 2000 onwards), along with control variables (Models 5 and 6 only).

Initially, forward citations (citations received from subsequent patents) are excluded from the analysis, despite evidence from prior literature (Harhoff et al., 1999; Palomeras, 2003; Gambardella et al., 2008 and Arts and Veugelers, 2012) supporting their importance as an indicator of patent value and/or technological importance. These citations are usually not available until quite some time after the patent grant/publication date. Since this study aims to identify factors that link patent characteristics to the likelihood of a successful approval outcome that can be applied to both older and newer patents, forward citations and their related lag measure are initially excluded from the analysis.

Models 7 to 10 contain all available patent-level, firm-level characteristics and control variables (Model 7 excludes the control variables).

All the models are tested for the sample containing both approval and non-approval linked patents.

5.4.2.1 Analysis of the Influence of the Patent and Firm-level Publishing Characteristics, Available at Patent Grant/Publication Date, on the Likelihood of FDA Drug Approval

This section investigates the influence of patent and firm-level publishing characteristics on the likelihood of drug approval. It was hypothesised in Chapter 3 that certain patent and firm-level publishing characteristics would increase the likelihood of a patent being linked to an approved drug. Models 2 to 6 are formulated to test these hypothesised relationships and the main regression results are detailed in Table 5.12. In order to measure the impact of patent and firm-level characteristics on approval likelihood, I begin with a simple baseline specification that includes only firm-level control variables. Model 1 therefore comprises control variables only. Models 2, 3 and 4 have no control variables. Model 2 explores the relationship between certain patent document characteristics and approval likelihood. It combines all of the hypotheses derived from the patent characteristics specified within the patent document as model independent variables. The model excludes prior art citation lag variables. These variables are excluded as not all patents reference both patents and NPR and missing values significantly reduce the sample size. In Model 3 the independent variables are derived from all the hypotheses linked to the firm's publishing characteristics. Model 4 combines all the independent variables from Models 2 and 3. Model 5 adds control variables to the Model 4 specification. Finally, Model 6 adds independent variables derived from the hypotheses linked to patent backward citation lag characteristics.

5.4.2.1.1 Model 1: Testing the Influence of the Control Variables

Model 1 includes only the 4 non-lag-related firm-level control variables and is formulated to test hypotheses H_{4a}, H_{4b}, H_{4c} and H_{4d}. These hypotheses are discussed in detail in Chapter 3, Section 3.5.

$$\text{DRUG APPROVAL} = \beta_1 AGE_{it} + \beta_2 FDAEXP_{it} + \beta_3 SIZE_{it} + \beta_4 ROA_{it}$$

(Model 1)

The subscript i denotes firm and t denotes the patent filing year. There is no constant term as the intercept of this model is time invariant so remains un-estimated.

Model 1 captures the effect of controls for firm age, a firm's prior experience of the FDA approval process, a firm's size (proxied by total assets) and return on assets (firm profitability) on the likelihood of approval. The results from univariate analyses (Table 5.11) show that a firm's profitability and prior experience of the FDA approval process have the highest influence on the likelihood of approval but all control variables have a positive (odds ratio >1) and significant influence, supporting hypotheses H_{4a}, H_{4b}, H_{4c} and H_{4d}. However, when these controls are combined in a multivariate analysis these findings change.

I find that firm age has an odds ratio below 1 that is statistically significant at the 1% level. This indicates that an increase in firm age will reduce the likelihood of approval. This is consistent with a prior study of U.S. biotechnology firms by Pisano (1997), who found that the age of a firm reduces the likelihood that an R&D project will reach the market. My results also indicate that larger firms are more likely to receive an approval, with total assets having an odds ratio greater than 1 that is statistically significant at the 1% level. The size of the firm is an influential factor, more than doubling the likelihood of approval. Larger firms can more easily finance riskier research projects and can benefit from economies of scale and scope. This is consistent with Guler and Nerkar (2012), who found that the size of the firm is significantly and positively related to the number of new drug products launched by a firm. As a robustness test, an alternate measure of firm size is estimated later in this chapter. Return on assets has an odds ratio greater than 1, indicating that an increase in this variable increases the likelihood of approval. However, the relationship is statistically insignificant. This result is consistent with that of Kapoor and Klueter (2015) who found that return on assets had a positive but insignificant relationship with the likelihood of a drug entering clinical trials. Prior experience of the approval process has an odds ratio below 1, indicating that prior experience reduces the likelihood of approval. However, it is an insignificant predictor of the likelihood of approval. This insignificant result is consistent with the findings of Stone and Flamm (2009) who found that firm experience of the FDA approval process has an insignificant and positive effect on drug development time.

These results imply that older firms with more experience and more funds available to invest in research projects are less likely than younger firms to receive an approval, although increasing age by one year only reduces approval likelihood by around 6%. From the small p-value from the likelihood-ratio (LR) test, <0.00001, it can be concluded that at least one of the regression coefficients in this model are not equal to zero. This suggests that the model with only control variables is significant in explaining the likelihood of approval.

5.4.2.1.2 Model 2: Testing the Influence of Patent Document Characteristics

Models 2 includes only the 4 non-lag-related patent variables and is formulated to test hypotheses H_{1a}, H_{1b}, $H_{1c(i)}$ and $H_{1c(ii)}$. These hypotheses are discussed in detail in Chapter 3, Section 3.2.

$$\text{DRUG APPROVAL} = \beta_1 \text{INVENTORS}_{it} + \beta_2 \text{NPR}_{it} + \beta_3 \text{ICLAIMS}_{it} + \beta_4 \text{DCLAIMS}_{it}$$

(Model 2)

The inventor team size hypothesis (hypothesis 1_a, discussed in Section 3.2.1) predicts that patents with a larger inventor team have an increased likelihood of being linked to an approved drug. As discussed in Section 3.2.1, the inventor team size can be used as a proxy of the level of resources invested in a drug research project. Also, according to Singh and Fleming (2010), teams of inventors are more likely to generate breakthrough inventions. This hypothesis is supported in Model 2, which shows that approval probability is positively related to inventor team size. The odds ratio of the variable is greater than 1, as I predicted, and significant at the 1% level. Having an additional inventor added to the research team increases the odds of an approval linkage by 30.5%. The non-patent reference hypothesis (hypothesis 1_b, discussed in Section 3.2.2.1) predicts that patents with more citations to NPR will increase the likelihood of a patent being linked to an approved drug. This patent characteristic indicates that the patent is linked to more exploratory research that may lead to a scientific breakthrough. Allison and Tiller (2003) found that patents with these references were often considered more valuable than those that did not. This hypothesis is also supported in Model 2 as the odds ratio of the variable is greater than 1 and significant ($p < 0.01$). Having an additional non-patent reference in a patent document increases the odds of an approval linkage by 20.1%. Hypothesis $1_{c(i)}$ posits that if a patent contains higher numbers of independent claims, the patent is more likely to be linked to an approved drug. This is based on the argument that patents that contain higher numbers of independent claims are more technologically important and of higher potential economic value. Barney (2002) observed that the number of independent claims indicated the novelty and quality of a patent. He also found that the greater the number of these claims the more valuable the patent. This importance and value will lead to an increased likelihood that these patents will be linked to an FDA drug approval. This hypothesis is also supported in Model 2 as the odds ratio for independent claims is greater than 1 and significant ($p < 0.01$), indicating that an additional

independent claim on a patent document increases the odds of an approval linkage by 39%. Hypothesis $1_{c(ii)}$ posits that if a patent contains higher numbers of dependent claims then the patent is more likely to be linked to an approved drug. Pioneering inventions can make greater numbers of claims and this increased numbers of claims acts as an indicator of a larger inventive step. This may indicate that the patent is linked to a piece of innovative research that is more likely to result in a new drug. This hypothesis is also supported in Model 2 as the odds ratio for dependent claims is greater than 1 and significant ($p < 0.01$), indicating that an additional dependent claim in a patent document increases the odds of an approval linkage by 16.2%.

The odds ratios of inventors (research team size), citations to NPR, independent claims and dependent claims have values above 1 and are statistically significant at the 1% level, with independent claims being the most influential factor, increasing the odds of an approval-linkage by nearly 40%. Dependent claims have the least positive influence, increasing the odds of an approval-linkage by just over 16%. Although not displayed, this model with the addition of the firm-level control variables produced very similar results in terms of both odds ratio values and statistical significance. These results provide strong support for the acceptance of hypotheses H_{1a}, H_{1b}, $H_{1c(i)}$ and $H_{1c(ii)}$.

5.4.2.1.3 Model 3: Testing the Influence of Firm-Level Publishing Characteristics

Model 3 includes only the 4 firm-level publishing variables and is formulated to test hypotheses H_{2a}, H_{2b}, H_{2c} and H_{2d}. These hypotheses are discussed in detail in Chapter 3, Section 3.3.

$$\text{DRUG APPROVAL} = \beta_1 PUB_{it} + \beta_2 COAUTHAC_{it} + \beta_3 COAUTHFIRM_{it} + \beta_4 FUNDER_{it}$$

(Model 3)

The firm publishing hypothesis (hypothesis 2_a, discussed in Section 3.3.1) predicts that patents held by firms who publish scientific journal articles will have an increased likelihood of being linked to an approved drug. As discussed in Section 3.3.1, this publication measure can be used as a proxy for a firm's scientific expertise (Gittelman, 2007). It follows from this argument that a firm with greater levels of scientific expertise will be more likely to produce patents that will result in an approved drug. The co-authoring behaviour hypotheses

(hypotheses 2_b and 2_c, discussed in Section 3.3.1) predict that patents held by firms who publish scientific journal articles co-authored with universities or other firms will have an increased likelihood of being linked to an approved drug. As discussed in Section 3.3.1, these publication relationships can improve a firm's innovative performance (Jong and Slavova, 2014) and make it more likely for such a firm to produce patents that will result in an approved drug. However, as can be observed in Table 5.12 publishing and university co-authoring behaviour by the firm have odds ratios of less than 1, marginally reducing the odds of an approval linkage. Only firm publishing is statistically significant at the 1% level, with university co-authoring being insignificant. Although not displayed, this model with the addition of the firm-level control variables produced very similar results in terms of odds ratio values for firm publishing and university co-authoring although university co-authoring is statistically significant at the 5% level. Co-authoring with other firms has an odds ratio almost equal to zero and the funder variable is excluded from the regression by Stata as there is no within firm variation. The co-authoring with firms and funder variables are excluded from any further analyses. These findings reject hypotheses H_{2a}, H_{2b}, H_{2c} and H_{2d}.

5.4.2.1.4 Model 4: Testing the Influence of Patent Document and Firm-Level Publishing Characteristics

Model 4 includes the four patent document variables (Model 2) and two firm-level publishing variables in the regression. Co-authoring with firms and the firm funder variables have been excluded based upon the regression results from Model 3. It is formulated to test hypotheses H_{1a}, H_{1b}, $H_{1c(i)}$, $H_{1c(ii)}$, H_{2a} and H_{2b}. These hypotheses are discussed in detail in Chapter 3, Sections 3.2 and 3.3.

$$\text{DRUG APPROVAL} = \beta_1 \text{INVENTORS}_{it} + \beta_2 \text{NPR}_{it} + \beta_3 \text{ICLAIMS}_{it} + \beta_4 \text{DCLAIMS}_{it} + \beta_5 \text{PUB}_{it} + \beta_6 \text{COAUTHAC}_{it}$$

(Model 4)

The odds ratios of inventors, citations to NPR, independent patent claims and dependent patent claims, as in Model 2, have values greater than 1 and are statistically significant at the 1% level. Odds ratio values are almost identical to those for Model 2. Publishing and co-authoring behaviour by the firm have odds ratios below 1, again marginally reducing the odds of an approval-linkage, with publishing being significant at the 1% level and co-

authoring at the 10% level. Again, the odds ratio values are almost identical to those for Model 3. The results for the patent-related variables are consistent with hypotheses H_{1a}, H_{1b}, H_{1c} and H_{1d} but those for the publishing-related variables are contrary to the proposed hypotheses H_{2a} and H_{2b} and confirm the results for model formulations 2 and 3.

5.4.2.1.5 Model 5: Testing the Influence of Patent Document and Firm-Level Publishing Characteristics, with Control Variables

Model 5 includes the four patent document variables (Model 2), two firm-level publishing variables (reduced form of Model 3) and now includes the control variables in the regression (Model 1). To test whether the observed relationships could be invalid and produced by the omission of some variables, the control variables are included in the Model 5 regression. It is formulated to test hypotheses H_{1a}, H_{1b}, $H_{1c(i)}$, $H_{1c(ii)}$, H_{2a}, H_{2b}, H_{4a}, H_{4b}, H_{4c} and H_{4d}. These hypotheses are discussed in detail in Chapter 3, Sections 3.2, 3.3 and 3.5.

$$\text{DRUG APPROVAL} = \beta_1 AGE_{it} + \beta_2 FDAEXP_{it} + \beta_3 SIZE_{it} + \beta_4 ROA_{it} + \beta_5 INVENTORS_{it} + \beta_6 NPR_{it} + \beta_7 ICLAIMS_{it} + \beta_8 DCLAIMS_{it} + \beta_9 PUB_{it} + \beta_{10} COAUTHAC_{it}$$

(Model 5)

Consistent with the hypotheses above, the odds ratio of inventors, citations to NPR, independent patent claims and dependent patent claims have values above 1 and are again statistically significant at the 1% level. As in Model 2, this indicates that increases in these patent characteristics increase the odds of a patent being linked to drug approval, with the number of patent independent claims having the strongest influence, with one additional claim increasing the odds of an approval-linkage by 44.5%. Previous studies of patent value and quality such as Barney (2002), Reitzig (2004), Allison and Sager (2007) and Trappey, Trappey, Wu and Lin (2012) found that patent value and quality increases with the number of independent claims. Having one additional inventor on the research team increases the odds of an approval-linkage by 25.6% and this is in line with empirical findings in studies of the value and importance of U.S. patents, such as Singh and Fleming (2010), Liu (2014) and Breitzman and Thomas (2015).

The addition of one citation to a non-patent reference or one dependent patent claim increases the odds of an approval-linkage by just over 14% in both instances. Publishing and

co-authoring behaviour by the firm marginally reduces the likelihood of receiving an approval, with a one-unit increase in either variable reducing the odds of an approval-linkage by between 3% to 6%. In the area of firm publishing behaviour there have been mixed results from prior literature. Kapoor and Klueter (2015), in their global pharmaceutical study, also found that a firm's scientific publications have a negative impact on the likelihood of a drug first entering clinical trials, but this relationship is statistically insignificant. Whilst Cockburn and Henderson (1998) found that publication volume also has a generally negative and insignificant effect on a patent's importance but that co-authoring with the public sector has a positive and significant effect. However, Gittelman and Kogut (2003) found, for U.S. biotechnology firms, that publishing volumes significantly raised the level of patent citations received (forward citations) but that co-authoring (includes firms/universities) has a positive but insignificant effect.

5.4.2.1.6 Model 6: Testing the Influence of All Patent and Firm-Level Publishing Characteristics, with Control Variables

Model 6 includes all the control variables, all six patent-level variables and the two firm-level publishing variables. It is formulated to test the hypotheses of Model 5 with the addition of hypotheses H_{1d} and H_{1e}. These hypotheses are discussed in detail in Chapter 3, Sections 3.2, 3.3 and 3.5.

$$DRUGAPPROVAL = \beta_1 AGE_{it} + \beta_2 FDAEXP_{it} + \beta_3 SIZE_{it} + \beta_4 ROA_{it} + \beta_5 INVENTORS_{it} + \beta_6 NPR_{it} + \beta_7 ICLAIMS_{it} + \beta_8 DCLAIMS_{it} + \beta_9 PUB_{it} + \beta_{10} COAUTHAC_{it} + \beta_{11} NCL_{it} + \beta_{12} PCL_{it}$$

(Model 6)

The prior art citation lag length hypotheses, hypotheses 1_d and 1_e, predict that patents that reference younger prior art will have an increased likelihood of being linked to an approved drug. As discussed in Section 3.2.4, shorter prior art citation lags indicate that the patent is built upon new or novel knowledge. Nagaoka (2007) found that short citation lags indicated a patent of higher quality.

In Model 6, it can be observed that when patent and non-patent citation lag lengths are added to the model, the age of non-patent citations is only significant at the 10% level, with an additional day in lag length increasing the odds of an approval-linkage by 21.3%. This result

is not consistent with hypothesis H_{1d} where I would expect that a shorter citation lag, indicating the use of newer knowledge by the patent, would produce a more innovative patent, increasing the odds of an approval-linkage and producing an odds ratio of less than one. This does occur for the patent citation lag, but the result is statistically insignificant. This result contradicts hypothesis H_{1e} and also results from prior studies such as Schoenmakers and Duysters (2010) and Nemet and Johnson (2012) who found that usefulness of knowledge depreciates, so older prior art is viewed as less valuable than more recent prior art. It is however consistent with the findings of Stone and Flamm (2009) who, in their study of U.S. drug development times, found that development times decreased with shorter patent citation lag lengths but the effect was statistically insignificant. Although not reported here, when each of these citation lag measures is included individually neither is statistically significant. This suggests that patents are more likely to receive a drug approval if they are based on a combination of newer patented knowledge and older basic scientific knowledge. This result whilst unexpected is partly consistent with a prior study by Jibu, Osabe and Börner (2015) which found that, for global pharmaceutical firms, patents that were linked to drugs cited older NPR as well as older patents. Studies by Ahuja and Morris Lampert (2001) and Schoenmakers and Duysters (2010) also suggested that older knowledge is also employed in creating radical inventions. Firms may find older knowledge more accessible, easier to understand, assimilate and recombine to produce innovative drug products. Newer scientific knowledge may also present a more challenging task to translate it into a new commercial product, as found by Chesbrough and Rosenbloom (2002) in their case study-based paper. Although not displayed, this model, without publication-related variables, produced very similar results in terms of both odds ratio values and statistical significance for these citation lag variables. These results reject hypotheses H_{1d} and H_{1e}.

NPR play no significant role in the model, indicating that it is the age of the scientific knowledge used, not the quantity that influences the likelihood of approval. Return on assets is a significant predictor in the model at the 10% level but with an odds ratio close to zero, indicating that higher levels of firm profitability significantly reduce the likelihood of approval. This is counter-intuitive as a profitable firm has more funds to invest in its research projects. The significance level for inventors has reduced from 1% to 5% and its odds ratio indicates that for each additional inventor listed on the patent document the odds of an approval-linkage increase by 35%. The statistical significance of university co-authoring has fallen from 1% to 10% but with little change in its odds ratio. The odds ratio for dependent claims indicates that one additional claim now increases the odds of an approval-linkage by

28% compared to the 16% noted for Model 2. Differences in the statistical significance of variables and their odds ratios, could to some extent be attributable to the different sample size being analysed. The observations available for Model 6 have reduced to just under 4000, only 30% of the total sample and including only 50 firms out of the 81 in the original sample.

With reference to the control variables, patent and publishing variables, the values of all the odds ratios remain unchanged for the complete sample. They remain either consistently above one or below one. In summary, several interesting observations emerge from Table 5.12. First, irrespective of the sample size, the odds ratios for firm age and publishing are always highly significant at the 1% level, with a value below 1. This provides sufficient evidence to conclude that these variables are relevant in determining approval likelihood. An increase in either of these variables appears to reduce the likelihood of approval. The odds ratios for firm size, number of inventors, patent independent claims and patent dependent claims are always highly significant at the 1% level, but with a value of greater than 1. This provides evidence that these variables are also relevant in determining approval likelihood, with increases in these variables increasing the likelihood of approval.

Table 5.12 Conditional Logistic Regression of Likelihood of Approval on Patent and Firm-level Characteristics (Available at Patent Grant Date)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Firm Age (AGE_{it})	0.939(0.012)***				0.923(0.013)***	0.902(0.027)***
FDA Experience ($FDAEXP_{it}$)	0.718(0.215)				0.765(0.238)	0.382(0.245)
Total Assets ($SIZE_{it}$)	2.183(0.215)***				2.209(0.223)***	2.972(0.659)***
Return on Assets (ROA_{it})	1.625(1.133)				1.623(1.153)	0.071(0.099)*
Number of Inventors ($INVENTORS_{it}$)		1.305(0.102)***		1.305(0.103)***	1.256(0.104)***	1.353(0.207)**
Non-patent References (NPR_{it})		1.201(0.049)***		1.206(0.049)***	1.145(0.050)***	1.048(0.117)
Independent Patent Claims ($ICLAIMS_{it}$)		1.390(0.088)***		1.409(0.089)***	1.445(0.093)***	1.450(0.177)***
Dependent Patent Claims ($DCLAIMS_{it}$)		1.162(0.056)***		1.163(0.056)***	1.146(0.056)***	1.279(0.122)***
Firm Publishing (PUB_{it})			0.925(0.014)***	0.935(0.015)***	0.938(0.015)***	0.923(0.023)***
University Co-authoring Share ($COAUTHAC_{it}$)			0.995(0.013)	0.978(0.013)*	0.964(0.013)***	0.951(0.029)*
Firm Co-authoring ($COAUTHFIRM_{it}$)			0.000(0.000)			
Funder of Co-authored Articles ($FUNDER_{it}$)			1.000(Omitted)			
NPR Citation Lag Length (NCL_{it})						1.213(0.128)*
Patent Citation Lag Length (PCL_{it})						0.922(0.059)
Observations	12,957	12,957	12,228	12,957	12,957	3,915
Number of Firms	81	81	68	81	81	50
Firm and Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Log Likelihood	-1,905.5	-1,921.4	-1,862.8	-1,908.5	-1,855	-520.5
McFaddens-R^2	0.032	0.024	0.010	0.031	0.058	0.078
Prob>Chi^2	0.000	0.000	0.000	0.000	0.000	0.000

Notes: The table presents the results of a conditional logit regression analysis with fixed effects where the dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate) and the independent and control variables are as defined in Chapter 3. Models 1 to 6 are multivariate logit models that use various combinations of the control variables, the variables available at a patent's grant date as independent variables and then regresses them on patent drug approval linkage probability. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets at the date of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5^{th} and 95^{th} percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, NCL_{it} is the natural log of the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent, PCL_{it} is the natural log of the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent, PUB_{it} is the firm's average annual publishing output of journal articles, $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing, $COAUTHFIRM_{it}$ is 1 when the firm co-authors scientific journal articles with another firm or 0 otherwise and $FUNDER_{it}$ is 1 when the firm funds co-authored scientific journal articles or 0 otherwise. Odds ratios are displayed with their standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

5.4.2.2 Testing the Impact of All Patent and Firm-level Characteristics on the Likelihood of FDA Drug Approval

Forward citation measures that were initially excluded to reflect their lack of availability at the patent grant date are now introduced into the model formulation. Hypotheses H_{3a} and H_{3b} relate to forward citation-related measures and are discussed in detail in Chapter 3, Section 3.4.

Table 5.13 presents the results of the fixed effects logistic regressions for Models 7 to 10 based on the likelihood of FDA drug approval.

5.4.2.2.1 Models 7 to 10: Testing the Influence of Forward Citations

Model 7 is formulated to test hypotheses H_{1a}, H_{1b}, $H_{1c(i)}$, $H_{1c(ii)}$, H_{2a}, H_{2b} and H_{3a}. It has the same formulation as Model 4 but with the addition of the forward citations (citations received). As discussed in Section 3.2 and 3.4 of Chapter 3, this explores the relationship between all patent-level characteristics (excluding any lag measures) and the likelihood of FDA drug approval. The model is restated as:

$$APPROVAL = \beta_1 INVENTORS_{it} + \beta_2 NPR_{it} + \beta_3 ICLAIMS_{it} + \beta_4 DCLAIMS_{it} + \beta_5 FC_{it} + \beta_6 PUB_{it} + \beta_7 COAUTHAC_{it}$$

(Model 7)

In Model 7, consistent with hypothesis H_{3a}, the odds ratio of forward citations has a value of 1.464 indicating that one additional citation received increases the odds of an approval-linkage by 46.4%. The effects and significance of the other variables in this model are mostly similar to those of Model 4. However, dependent claims now have an insignificant odds ratio that is close to 1. The odds ratios for inventors and NPR have fallen in value increasing the odds of an approval-linkage by 20.6% for one additional inventor and 15.1% for an additional non-patent reference compared to the 30.5% and 20.6% reported for Model 4. The significance for inventors falls to the 5% level, compared to 1% in Model 4. Firm publishing and university co-authoring odds ratio are virtually identical to those of Model 4. Firm publishing remains statistically significant at the 1% level, but university co-authoring is now also significant at the 1% level.

Model 8 is formulated to test hypotheses H_{1a}, H_{1b}, $H_{1c(i)}$, $H_{1c(ii)}$, H_{2a}, H_{2b}, H_{3a}, H_{4a}, H_{4b}, H_{4c} and H_{4d}. It has the same formulation as Model 5 but with the addition of the forward citations.

$$\text{DRUG APPROVAL} = \beta_1 AGE_{it} + \beta_2 FDAEXP_{it} + \beta_3 SIZE_{it} + \beta_4 ROA_{it} + \beta_5 INVENTORS_{it} + \beta_6 NPR_{it} + \beta_7 ICLAIMS_{it} + \beta_8 DCLAIMS_{it} + \beta_9 FC_{it} + \beta_{10} PUB_{it} + \beta_{11} COAUTHAC_{it}$$

(Model 8)

In Model 8, consistent with hypothesis H_{3a}, the odds ratio of forward citations has a value of 1.548 indicating that one additional citation received increases the odds of an approval-linkage by 54.8%. The effects and significance of some of the other variables in this model differ markedly from Model 5. Dependent claims again have an insignificant odds ratio of close to 1. The odds ratios for inventors and NPR have fallen in value increasing the odds of an approval-linkage increase by 18% for one additional inventor and 11%, for an additional non-patent reference compared to the 25% and 14% reported for Model 5. The results are significant at the 5% level, compared to 1% in Model 5. University co-authoring remains statistically significant at the 1% level, as found in Model 5, with the value of its odds ratio being virtually identical.

Model 9 is formulated to test hypotheses H_{1a}, H_{1b}, $H_{1c(i)}$, $H_{1c(ii)}$, H_{2a}, H_{2b}, H_{3a}, H_{3b}, H_{4a}, H_{4b}, H_{4c} and H_{4d}. It has the same formulation as Model 5 but with the addition of the forward citations and the forward citation lag length variable.

$$\text{DRUG APPROVAL} = \beta_1 AGE_{it} + \beta_2 FDAEXP_{it} + \beta_3 SIZE_{it} + \beta_4 ROA_{it} + \beta_5 INVENTORS_{it} + \beta_6 NPR_{it} + \beta_7 ICLAIMS_{it} + \beta_8 DCLAIMS_{it} + \beta_9 FC_{it} + \beta_{10} FCL_{it} + \beta_{11} PUB_{it} + \beta_{12} COAUTHAC_{it}$$

(Model 9)

In Model 9, the speed with which forward citations are received has an odds ratio that is statistically significant, but only at the 10% level and with a value greater than 1. This indicates a positive relationship with the likelihood of approval, with the likelihood of approval increasing with the time taken for a citation to be received. The odds of an approval-linkage increase by just over 11%. This rejects hypothesis H_{3b}, that there is a statistically significant and negative relationship between the speed at which patent citations are received and the likelihood of approval. Although not reported here, when the quantity of forward

citations received is excluded and only the speed of these citations is observed then shorter citation lags do increase the odds of an approval-linkage but by just under 9%. The quantity of citations received is the dominant factor in influencing the likelihood of approval and not the speed at which they are received. There is no evidence that shorter forward citation lags are a key factor in increasing the likelihood of approval. The U.S. patent study of Palomeras (2003) suggested that the forward citation lag could be regarded as a measure of the innovativeness of the patent, with longer lags indicating that other inventors took longer to recognise the patent's potential. The odds ratios for NPR and co-authoring with academia have fallen in value and are also statistically insignificant in the model, having no role in explaining the factors influencing the likelihood of approval. The odds ratio for inventors has increased but it too is also an insignificant variable within the model.

Model 10 is formulated to test hypotheses H_{1a}, H_{1b}, $H_{1c(i)}$, $H_{1c(ii)}$, H_{1d}, H_{1e} H_{2a}, H_{2b}, H_{3a} and H_{3b}. It has the same formulation as Model 6 but with the addition of the forward citations and the forward citation lag length variable.

$$DRUGAPPROVAL = \beta_1 AGE_{it} + \beta_2 FDAEXP_{it} + \beta_3 SIZE_{it} + \beta_4 ROA_{it} + \beta_5 INVENTORS_{it} + \beta_6 NPR_{it} + \beta_7 ICLAIMS_{it} + \beta_8 DCLAIMS_{it} + \beta_9 FC_{it} + \beta_{10} FCL_{it} + \beta_{11} PUB_{it} + \beta_{12} COAUTHAC_{it} + \beta_{13} NCL_{it} + \beta_{14} PCL_{it}$$

(Model 10)

In Model 10, it can be observed that both independent claims and firm publishing cease to be statistically significant when patent and non-patent citation lag lengths are added to the model. Dependent claims display a significantly increased odds ratio, with one additional claim increasing the odds of an approval-linkage by nearly 30% that is significant at the 10% level. FDA experience becomes significant for the first time with its odds ratio indicating that it decreases the odds of an approval-linkage by just below 80% and this result is significant at the 10% level. Such differences, in which variables are statistically significant, could also be attributable to the different sample size. The observations available to the model have reduced significantly to just over 2,000 representing only 16% of the total sample and only includes 33 firms out of the 81 in the original sample.

The addition of the citation lag measures (both backward and forward) to model formulations significantly reduces both the number of observations and firms included in these regressions. The inclusion of the patent citation lag measure is not statistically significant in

Table 5.13 Conditional Logistic Regression of Likelihood of Approval on All Patent and Firm-level Characteristics

	Model 7	Model 8	Model 9	Model 10
AGE_{it}		0.901(0.013)***	0.834(0.022)***	0.799(0.045)***
$FDAEXP_{it}$		0.793(0.249)	0.579(0.265)	0.203(0.182)*
$SIZE_{it}$		2.429(0.252)***	4.082(0.844)***	5.850(2.571)***
ROA_{it}		1.675(1.199)	0.548(0.635)	0.042(0.095)
$INVENTORS_{it}$	1.202(0.095)**	1.184(0.099)**	1.252(0.151)*	1.387(0.312)
NPR_{it}	1.151(0.047)***	1.112(0.049)**	1.096(0.074)	1.005(0.171)
$ICLAIMS_{it}$	1.361(0.087)***	1.375(0.089)***	1.332(0.126)***	1.262(0.223)
$DCLAIMS_{it}$	1.082(0.053)	1.064(0.053)	1.043(0.074)	1.282(0.186)*
FC_{it} (five-year window)	1.464(0.064)***	1.548(0.071)***	1.729(0.159)***	1.968(0.319)***
FCL_{it}			1.115(0.067)*	0.999(0.106)
PUB_{it}	0.939(0.015)***	0.943(0.015)***	0.908(0.024)***	0.007(6643.30)
$COAUTHAC_{it}$	0.974(0.013)**	0.961(0.013)***	0.952(0.029)	0.536(240330.6)
NCL_{it}				1.254(0.189)
PCL_{it}				0.957(0.088)
Observations	12,957	12,957	6,981	2,091
Number of Firms	81	81	63	33
Firm and Time Fixed Effects	Yes	Yes	Yes	Yes
Log Likelihood	-1,870.4	-1,808.5	-883.9	-234.6
McFaddens-R^2	0.050	0.082	0.076	0.158
Prob>Chi^2	0.0000	0.0000	0.0000	0.0000

Notes: The table presents the results of a conditional logit regression analysis with fixed effects where the dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate) and the independent variables and control variables are as defined in Chapter 3. Models 7 to 10 are multivariate logit models which use various combinations of the control variables, the patent and publishing variables defined in Chapter 3 as independent variables except for $COAUTHFIRM_{it}$ (firm co-authors scientific journal articles with another firm) and $FUNDER_{it}$ (firm funds co-authored scientific journal articles) and then regresses them on patent drug approval linkage probability. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets at the date of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5th and 95th percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, FCL_{it} is the natural log of the time period, in days, from the filing date of the citing patent to the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles, $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing, NCL_{it} is the natural log of the time period, in days, from the patent filing date to the publication year of the latest publication cited by the patent and PCL_{it} is the natural log of the time period, in days, from the patent filing date to the grant date of the latest patent cited by the patent. Odds ratios are displayed with their standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

any of the models, which rejects the proposition of hypothesis H_{1e}. At the point of patent publication (Model 6) only the non-patent citation lag measure is significant at the 10% level, but with an odds ratio 1.213. This is contrary to hypothesis H_{1d} which predicted an odds ratio of less than 1. In Model 10 the non-patent citation lag and forward citation lag measures play no significant role in the model when all variables are included thereby rejecting hypotheses H_{1d} and H_{3b}.

5.4.2.3 The Use of Logged Explanatory Variables

I have included multiple logged explanatory variables within my models based on variable treatment techniques employed by prior literature (see for example Harhoff, Scherer and Vopel, 2003; Fabrizio, 2009 and Wagner and Wakeman, 2014). However, the inclusion of multiple logged explanatory variables within the regression models can imply a multiplicative relationship between variables due to the product rule of logarithms. This is not the nature of the relationships outlined in the hypotheses discussed in Chapter 4. Whilst the odds ratios for the logged patent-related variables in Models 5 and 8 are lower than those in the univariate regressions displayed in Table 5.11, they are broadly similar in magnitude. As a check of the integrity of the results observed in Tables 5.12 and 5.13 I ran the regressions for Models 5 and 8 using one explanatory logged variable at a time with the remainder of the previously logged variables winsorised. The previously logged variables remained significant and odds ratios remained above 1, supporting the multivariate results.

5.5 Robustness and Sensitivity Analysis

This section discusses results based on a series of robustness and sensitivity analyses.

5.5.1 Regression Results based on Alternative Measures of Firm Size

As noted in Section 4.4.1.4 of Chapter 4 there is no consistency in the use of any single proxy for size. Different size proxies capture different aspects of firm size and no measure captures all the characteristics of the firm. This leaves proxy choice to the discretion of the researcher. The total assets size proxy measures the total resources available to the firm and this attribute combined with superior data availability influenced the choice of total assets as the firm size proxy for this research. For Models 1, 5, 6, and 8 to 10, firm size has the largest influence on the likelihood of approval. As firm size plays such a significant role in the results, it is important to test the robustness of my results to changes in firm size proxies. The logarithm of total employees in the year of patent filing is utilised as an alternative measure of firm size. This measure has been used in many prior studies, such as Fabrizio

(2009), Stone and Flamm (2009) and Subramanian et al. (2013). The regressions in Tables 5.12 (Model 5) and 5.13 (Model 8) were run using this proxy for firm size and the results are displayed in Table 5.14.

Table 5.14 Conditional Logistic Regression of Likelihood of Approval Results: Models 5 and 8 with Total Employees as an Alternate Firm Size Measure

	Model 5a	**Model 8a**
Firm Age (AGE_{it})	0.981(0.008)**	0.963(0.008)***
FDA Experience ($FDAEXP_{it}$)	0.763(0.251)	0.786(0.263)
Total Employees ($SIZE_{it}$)	3.157(0.457)***	3.710(0.546)***
Return on Assets (ROA_{it})	4.723(3.296)**	5.290(3.711)**
Number of Inventors ($INVENTORS_{it}$)	1.243(0.105)***	1.166(0.099)*
Non-patent References (NPR_{it})	1.141(0.051)***	1.105(0.049)**
Independent Patent Claims ($ICLAIMS_{it}$)	1.464(0.096)***	1.392(0.092)***
Dependent Patent Claims ($DCLAIMS_{it}$)	1.136(0.056)***	1.045(0.052)
Forward Citations Received (five-year window) (FC_{it})		1.595(0.075)***
Firm Publishing (PUB_{it})	0.941(0.015)***	0.947(0.016)***
University Co-authoring Share ($COAUTHAC_{it}$)	0.965(0.014)**	0.961(0.014)***
Observations	12,642	12,642
Number of Firms	79	79
Firm and Time Fixed Effects	Yes	Yes
Log Likelihood	-1,812.2	-1,761.4
McFaddens-R^2	0.062	0.088
Prob>Chi^2	0.0000	0.0000

Notes: The table presents the results of a conditional logit regression analysis with fixed effects where the dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate) and the control and the independent variables are as defined in Chapter 3. Models 5a and 8a are multivariate logit models, Model 5a uses the independent and control variables of Model 5 (Table 5.12). Model 8a uses the independent and control variables of Model 8 (Table 5.13) and then regresses them on patent drug approval-linkage probability. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total employees in the year of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5^{th} and 95^{th} percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles and $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing. Odds ratios are displayed with their standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Model 5a displays the regression results for patent and firm-level variables available at the patent grant date whilst Model 8a displays the results for all patent and firm-level variables;

both exclude citation lag measures to maintain sample size. The results are consistent with those produced when using total assets as a measure for firm size. Interestingly, the patent-related and publishing variables have very similar odds ratios to those reported in Tables 5.12 and 5.13. For the control variables the odds ratios are larger for the firm size proxy and return on assets. All but two of the significance levels are identical for models using total assets and total employees as firm size proxies. The exceptions are the odds ratio of return on assets and number of inventors. Return on assets was statistically insignificant when the size proxy used is total assets but becomes statistically significant at the 5% level and increases the likelihood of approval by more than 400% when total employees is the size proxy. This outcome may be a result of the fact that return on assets is derived from the total assets measure and when they are both included in the regression the total assets relationship with the dependent variable dominates that of the return on assets measure. The number of inventors still has an odds ratio greater than one, but at a slightly reduced level. It is now only significant at the 10% rather than 5% level in the model including forward citations.

As a further robustness check I use firm market capitalisation as market-based measure of firm size. The logarithm of market capitalisation in the year of patent filing is utilised as a measure of firm size. The regressions for Models 5 and 8 were run using this alternate proxy for firm size, results are displayed in Table 5.15.

Model 5b displays the regression results for patent and firm-level variables available at the patent grant date whilst Model 8b displays the results for all patent and firm-level variables. The results are similar in size and significance to those produced when using total assets and total employees as size measures. The odds ratios for the patent-related and publishing variables are similar in magnitude to those reported in Tables 5.12, 5.13 and 5.14. For the control variables the odds ratios are very similar for firm age and lower for the firm size proxy. The insignificant odds ratio for the return on assets is lower than for total employees size proxy but is similar to that displayed in Table 5.12. The odds ratio on prior FDA experience now increases the likelihood of approval by 37% but remains statistically insignificant. All but two of the significance levels are identical for models using total assets, total employees and market capitalisation as firm size proxies. In Model 5 all patent-related variables, firm publishing and firm size are significant at the 1% level and prior FDA experience is insignificant. Return on assets is statistically insignificant when total assets or market capitalisation is the size proxy but statistically significant at the 5% level with total employees. In Model 8 the results are similar in size and significance to those produced when using total assets and total employees as firm size proxies. The main differences can be

observed in prior FDA experience, which has an odds ratio greater than one, and the reduced influence of firm size. The findings suggest that the results are robust regardless of the firm size measure used[72].

Table 5.15 Conditional Logistic Regression of Likelihood of Approval Results: Models 5 and 8 with Market Capitalisation as an Alternate Firm Size Measure

	Model 5b	**Model 8b**
Firm Age (AGE_{it})	0.943(0.015)***	0.933(0.015)***
FDA Experience ($FDAEXP_{it}$)	1.371(0.429)	1.501(0.473)
Market Capitalisation ($SIZE_{it}$)	1.601(0.142)***	1.599(0.142)***
Return on Assets (ROA_{it})	1.702(1.247)	2.146(1.568)**
Number of Inventors ($INVENTORS_{it}$)	1.271(0.109)***	1.206(0.104)***
Non-patent References (NPR_{it})	1.161(0.050)***	1.135(0.051)***
Independent Patent Claims ($ICLAIMS_{it}$)	1.426(0.095)***	1.366(0.092)***
Dependent Patent Claims ($DCLAIMS_{it}$)	1.161(0.059)***	1.084(0.056)
Forward Citations Received (five-year window) (FC_{it})		1.456(0.069)***
Firm Publishing (PUB_{it})	0.940(0.015)***	0.944(0.016)***
University Co-authoring Share ($COAUTHAC_{it}$)	0.972(0.013)**	0.971(0.014)**
Observations	11,965	11,965
Number of Firms	74	74
Firm and Time Fixed Effects	Yes	Yes
Log Likelihood	-1,725.5	-1,693.2
McFaddens-R^2	0.049	0.066
Prob>Chi^2	0.0000	0.0000

Notes: The table presents the results of a conditional logit regression analysis with fixed effects where the dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate) and the control and the independent variables are as defined in Chapter 3. Models 5a and 8a are multivariate logit models, Model 5a uses the independent and control variables of Model 5 (Table 5.12). Model 8a uses the independent and control variables of Model 8 (Table 5.13) and then regresses them on patent drug approval-linkage probability. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of market capitalisation in the year of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5th and 95th percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, FC_{it} is the natural log of the number of citations

[72] Regressions were also run for Models 5 and 8 using filing date research and development expense (R&D) and cash. For R&D results were similar in size/significance. Firm age, all patent and publishing variables had odds ratios of similar size and significance. Prior FDA experience had an insignificant odds ratio greater than one. Return on assets had an odds ratio similar in size to that for the regression using total employees as a size proxy. For cash the Model 5 results were similar in size/significance, although firm age became insignificant. Prior FDA experience was significant at the 10% level, increasing the likelihood of approval by 74% and the cash variable increased approval likelihood by 16%. Results were similar for Model 8 with prior FDA experience being significant at the 5% level, increasing the likelihood of approval by 95%. In addition the odds ratio for dependent claims, although similar in size, remained statistically significant at the 10% level.

received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles and $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing. Odds ratios are displayed with their standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

5.5.2 Regression Results based on Alternative Time Windows for Citations Received

The selection of the time period for which citations received are counted was based on prior literature and timeliness of data availability. Some prior studies have chosen alternatives to the five-year period utilised in this study. For instance, Liu et al. (2008) and Aggarwal and Hsu (2014) chose a four-year window from the patent grant date for their U.S. studies, whilst Liu (2014) chose a six-year window from the patent application date. In Models 7 to 10 the forward citations variable is the patent variable that has the largest influence on the likelihood of approval. It is therefore important to test the robustness of the results to changes in the time window for which these citations are counted.

The results are tested for robustness using alternative time windows for the receipt of citations from other patents. Time windows from one to seven years from the cited patent grant date were tested and the results are displayed in Table 5.16. Windows beyond seven years are not tested, as it is desirable that the data used in the regressions are available without a significant time lag. For all windows greater than two years, the results are very similar to those for the five-year window. All variables apart from prior FDA experience, return on assets and dependent patent claims are statistically significant. Firm age, publishing and university co-authoring have odds ratios of less than 1, the control variable firm size, and all the remaining patent-related variables have odds ratios greater than 1. It can also be observed that the values of these odds ratios are broadly similar across all the chosen time windows. The odds ratios for NPR vary very little with the time window selected, with values ranging from 1.120 to 1.104. Their statistical significance changes from 1% for one, two, and three-year windows to 5% for windows four years and longer.

The results for the one and two-year windows are very similar to those for the regressions excluding the forward citation measure with dependent claims being statistically significant but now significant at the 10% level. This is not unexpected as it can take several years for patents to receive citations with Omland (2011) observing that patents only received around 10% of their total citations in the first three years post grant. This suggests that the majority

of the results based on the five-year window are not sensitive to changes in time window length.

I also ran regressions for varying time windows including the forward citation lag measure and again the results (not displayed) for all windows greater than two years were broadly similar to those for the five-year window. All variables, apart from prior FDA experience, return on assets, citations to NPR, dependent patent claims and co-authoring with universities (not for the seven-year window) are statistically significant. However, for windows up to four years in length the number of inventors is significant at the 5% level, which reduces to 10% for windows five years and longer. The forward citation and forward citation lag measure's significance levels do vary, with both being significant for the one-year window with an odds ratio of less than 1 but then insignificant for the two-year window. Forward citations then become significant for the three-year window, with an odds ratio of greater than 1, but the related lag measure is not significant until the five-year window where its odds ratio is also greater than 1 and remains so for the six and seven-year windows. This may indicate that in the first year after patent grant it is not the number of citations a patent receives that increases the odds of an approval-linkage but the fact that it receives citations in this early period that influences the odds of an approval-linkage. As time passes and the patent receives more citations then the quantity of these citations becomes influential. As for the five-year window, firm age and publishing have odds ratios of less than 1. Firm size and all the remaining patent-related variables have odds ratios greater than 1.

5.5.3 Regression Results based on Alternative Measures of Prior Art Citations

The selection of the time period for which prior art citations are counted is based on prior literature and to reflect citations made by firm researchers. For that reason, the citation period was measured, for both patent and non-patent citations, from the grant date/publication year of the cited reference to the filing date/year of the citing patent. However, some prior studies have chosen alternatives to this period. Palomeras (2003) used the filing date of the citing patent to the filing date of the cited patent, Criscuolo and Verspagen (2008) used the priority date of the citing patent to the priority date of the cited patent and Nemet and Johnson (2012) used the grant date of the citing patent to the grant date of the cited patent.

The results are tested for robustness using the time elapsed, in days, from the grant date of the citing patent to the grant date of the latest cited patent and also for non-patent citations

from the grant year of the citing patent to the publication year of the latest non-patent reference (expressed in days). This doubles the observations available to the regression and allows 69 firms to be included rather than the 50 previously. The results are displayed in Table 5.17. Again, only the non-patent citation lag measure is significant but now with an odds ratio of less than 1. The statistical significance of the odds ratios of firm age, number of inventors, NPR and university co-authoring has changed. Specifically, the odds ratio of firm age and return on assets that were statistically significant at the 1% and 10% level respectively are no longer statistically significant. NPR and number of inventors are now significant at the 10% level, with university co-authoring being significant at the 1% level.

Table 5.16 Conditional Logistic Regression Results for Model 8 with Various Forward Citation Time Windows

	1 Year	2 Year	3 Year	4 Year	6 Year	7 Year
AGE_{it}	0.917(0.013)***	0.912(0.013)***	0.909(0.013)***	0.906(0.013)***	0.898(0.013)***	0.895(0.013)***
$FDAEXP_{it}$	0.744(0.232)	0.727(0.227)	0.748(0.239)	0.765(0.240)	0.831(0.262)	0.877(0.278)
$SIZE_{it}$	2.256(0.229)***	2.316(0.236)***	2.344(0.240)***	2.374(0.244)***	2.472(0.257)***	2.503(0.261)***
ROA_{it}	1.686(1.206)	1.676(1.193)	1.698(1.210)	1.709(1.219)	1.620(1.161)	1.623(1.163)
$INVENTORS_{it}$	1.210(0.101)**	1.204(0.100)**	1.196(0.100)**	1.196(0.100)**	1.181(0.098)**	1.175(0.098)**
NPR_{it}	1.120(0.049)***	1.120(0.049)***	1.119(0.049)***	1.117(0.049)**	1.108(0.048)**	1.104(0.048)**
$ICLAIMS_{it}$	1.413(0.092)***	1.406(0.091)***	1.398(0.091)***	1.384(0.090)***	1.369(0.089)***	1.363(0.089)***
$DCLAIMS_{it}$	1.097(0.054)*	1.089(0.054)*	1.080(0.054)	1.075(0.053)	1.054(0.052)	1.046(0.052)
FC_{it}	1.484(0.088)***	1.473(0.077)***	1.460(0.072)***	1.482(0.070)***	1.583(0.071)***	1.614(0.071)***
PUB_{it}	0.940(0.015)***	0.941(0.015)***	0.942(0.015)***	0.943(0.015)***	0.943(0.016)***	0.946(0.015)***
$COAUTHAC_{it}$	0.964(0.013)***	0.962(0.014)***	0.962(0.013)***	0.961(0.013)***	0.961(0.013)***	0.959(0.013)***
Observations	12,957	12,957	12,957	12,957	12,957	12,957
Number of Firms	81	81	81	81	81	81
Firm and Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Log Likelihood	-1,833.3	-1,827.8	-1,825.3	-1,819.7	-1,800.9	-1,793.8
McFaddens-R^2	0.069	0.072	0.073	0.076	0.085	0.089
Prob>Chi^2	0.000	0.000	0.000	0.000	0.000	0.000

Notes: The table presents the results of a conditional logit regression analysis with fixed effects where the dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate) and the control and the independent variables are those of Model 8 (see Table 5.13) and as defined in Chapter 3. FC_{it} is now the natural log of the number of citations received from subsequent patents (forward citations) within a 1,2,3,4,6 or 7 year period from the grant date of the cited patent, AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets in the year of patent filing, ROAit is the firm's return on assets (net income divided total assets), winsorised at the 5th and 95th percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, PUB_{it} is the firm's average annual publishing output of journal articles and $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing. Odds ratios are displayed with their standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

The maximum citation lag lengths for both measures is also measured from the filing date of the citing patent to the oldest cited patent and also the filing year of the citing patent to the oldest non-patent citation. The results are broadly similar, with total assets, independent claims, dependent claims, firm publishing and university co-authoring being statistically significant, whilst prior FDA experience and return on assets are insignificant within the regression. A limited number of sensitivities can be identified when compared with the original results. First, firm age and return on assets are no longer significant. Specifically, the odds ratio of firm age and return on assets, which were statistically significant at the 1% and 10% levels respectively, are no longer significant. By contrast, the odds ratio of NPR, which was statistically insignificant, is now significant at the 1% level. The maximum patent citation lag measure is statistically significant, at the 10% level, with an odds ratio greater than one, whilst the non-patent citation lag measure is insignificant with an odds ratio below one.

This result is similar to those of the European study by Criscuolo and Verspagen (2008) and the U.S. study by Nemet and Johnson (2012) who found that increased citation lags have a small but significant role in determining forward citation levels. Generally, the results suggest that the evidence regarding the sensitivity or robustness of the variables to a difference in citation lag measures is varied. Specifically, the odds ratios on the majority of variables are robust to the change in lag measure; however, the odds ratios on a limited number of variables (i.e., firm age, return on assets, number of inventors, NPR and citation lag measures) are sensitive to the use of alternate lag measures. The variation in the results may also be explained by the 100% increase in the number of patent observations available (i.e., from just under 4,000 to nearer to 8,000).

Table 5.17 Conditional Logistic Regression of Likelihood of Approval with Alternative Citation Lag Measures

	Grant Date to Grant Date	Maximum Citation Lag Length
Firm Age (AGE_{it})	0.974(0.018)	0.985(0.019)
FDA Experience ($FDAEXP_{it}$)	0.566(0.202)	0.557(0.201)
Total Assets ($SIZE_{it}$)	2.458(0.307)***	2.365(0.299)***
Return on Assets (ROA_{it})	0.272(0.227)	0.304(0.256)
Number of Inventors ($INVENTORS_{it}$)	1.180(0.114)*	1.159(0.112)
Non-patent References (NPR_{it})	1.129(0.072)*	1.207(0.082)***
Independent Patent Claims ($ICLAIMS_{it}$)	1.370(0.107)***	1.347(0.107)***
Dependent Patent Claims ($DCLAIMS_{it}$)	1.188(0.072)***	1.216(0.075)***
Firm Publishing (PUB_{it})	0.947(0.016)***	0.942(0.017)***
University Co-authoring Share ($COAUTHAC_{it}$)	0.933(0.015)***	0.934(0.017)***
NPR Citation Lag Length (NCL_{it})	0.801(0.051)***	0.941(0.047)
Patent Citation Lag Length (PCL_{it})	1.026(0.083)	1.154(0.077)**
Observations	8,076	7,726
Number of Firms	69	68
Firm and Time Fixed Effects	Yes	Yes
Log Likelihood	-1,211.8	-1,178.2
McFaddens-R^2	0.107	0.110
Prob>Chi^2	0.0000	0.0000

Notes: The table presents the results of a conditional logit regression analysis with fixed effects. The dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate). The control and the independent variables are as defined in Chapter 3. The models are multivariate logit models that use the independent and control variables of Model 6 (Table 5.12), with differing measures of NCL_{it} and PCL_{it}, and then regresses them on patent drug approval linkage probability. In column 2, NCL_{it} is the natural log of the period, in days, from the patent grant year to the publication year of the latest publication cited by the patent and PCL_{it} is the natural log of the period, in days, from the patent grant date to the grant date of the latest patent cited by the patent. In column 3, NCL_{it} is the natural log of the period, in days, from the patent grant date to the publication year of the oldest publication cited by the patent and PCL_{it} is the natural log of the period, in days, from the patent grant date to the filing date of the oldest patent cited by the patent. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets in the year of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5^{th} and 95^{th} percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, PUB_{it} is the firm's average annual publishing output of journal articles and $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing. Odds ratios are displayed with their standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

5.5.4 Regression Results based on Sample Spilt by Firm Age at Patent Filing

The robustness of the results is also tested with reference to firm age by splitting the sample into two sub-samples. One sample contains patents whose firm age (at patent filing) is above the median value of 85 years (6470 patents) and the other includes those at or below the median (6487 patents). Regressions are run on these sub-samples for Models 5 and 8, including control variables but excluding lag measures to maintain a reasonable sample size, and the results are displayed in Table 5.18.

Similar to the results for the sample as a whole, FDA experience and return on assets are statistically insignificant for Model 5 (excluding forward citations), with firm age, total assets, independent claims and university co-authoring all being statistically significant for both sub-samples. Firm age and university co-authoring have odds ratios of less than 1 and total assets and independent claims have odds ratios greater than 1 and are broadly similar to those for the sample as a whole. The significance levels are also at the 1% level for the below median age sub-sample but at the 5% or 10% levels for the above median sample for all significant variables apart from university co-authoring which remains at the 1% level. These results also hold when forward citations are added into the regression.

In contrast to the results for the sample as a whole, the number of inventors is no longer a statistically significant variable in either of the two sub-samples. NPR results differ in significance to those for the sample as a whole. These citations are significant at the 5% level for Model 5 and the 10% level for Model 8 in the below median sample, representing a reduction from the 1% and 5% levels for the sample as a whole. For the above median sample they are statistically significant only at the 10% level for Model 5 (significant at the 1% level for the whole sample) and statistically insignificant in Model 8 (significant at the 5% level for the whole sample). Dependent claims are statically significant for both Models 5 and 8, at the 1% and 5% levels respectively for the above median sample but insignificant for both models in the below median sample and with odds ratios below 1. This differs from the whole sample where dependent claims are statistically significant at the 1% level for Model 5 and insignificant for Model 8, with odds ratios greater than 1 in both cases. Firm publishing has similar results for the whole sample and the below median sample, but results differ in the above median sample, where firm publishing is insignificant in both Models 5 and 8.

Table 5.18 Conditional Logistic Regression of Likelihood of Approval Results: Models 5 and 8

	Below Median Firm Age		Above Median Firm Age	
Firm Age (AGE_{it})	0.919(0.018)***	0.906(0.018)***	0.937(0.035)*	0.912(0.034)**
FDA Experience ($FDAEXP_{it}$)	0.783(0.257)	0.785(0.261)	0.383(0.418)	0.421(0.460)
Total Assets ($SIZE_{it}$)	2.249(0.257)***	2.393(0.278)***	2.142(0.719)**	2.337(0.787)**
Return on Assets (ROA_{it})	2.537(2.219)	2.774(2.335)	0.345(0.477)	0.301(0.423)
Number of Inventors ($INVENTORS_{it}$)	1.180(0.144)	1.132(0.138)	1.152(0.135)	1.074(0.126)
Non-patent References (NPR_{it})	1.140(0.072)**	1.123(0.070)*	1.129(0.072)*	1.081(0.069)
Independent Patent Claims ($ICLAIMS_{it}$)	1.562(0.140)***	1.514(0.136)***	1.266(0.128)**	1.191(0.120)*
Dependent Patent Claims ($DCLAIMS_{it}$)	0.998(0.067)	0.944(0.064)	1.332(0.102)***	1.205(0.093)**
Forward Citations Received (five-year window) (FC_{it})		1.392(0.090)***		1.681(0.113)***
Firm Publishing (PUB_{it})	0.856(0.033)***	0.858(0.034)***	0.970(0.021)	0.970(0.020)
University Co-authoring Share ($COAUTHAC_{it}$)	0.819(0.064)***	0.812(0.064)***	0.891(0.024)***	0.900(0.025)***
Observations	6,470	6,470	6,461	6,461
Number of Firms	64	64	22	22
Firm and Time Fixed Effects	Yes	Yes	Yes	Yes
Log Likelihood	-856.7	-843.6	-915.8	-885.3
McFaddens-R^2	0.098	0.112	0.040	0.072
Prob>Chi^2	0.0000	0.0000	0.0000	0.0000

Notes: The table presents the results of a conditional logit regression analysis with fixed effects where the dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate) and the control and the independent variables for columns 1 and 3 are those of Model 5 (see Table 5.12). Columns 2 and 4 use Model 8 (see Table 5.13). All variables are as defined in Chapter 3. Columns 1 and 2 use a sample of patents belonging to firms with a below median age, whilst columns 3 and 4 use a sample of patents belonging to firms with an above median age. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets in the year of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5^{th} and 95^{th} percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles and $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing. Odds ratios are displayed with their standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

5.5.5 Regression Results based on Sample Split by Firm Sales at Patent Filing

I test the robustness of my results with reference to firm size by splitting the sample into two sub-samples based on firm sales (at patent filing). As discussed in Section 4.2.3 the sample firms vary in size and age. One sample contains patents whose firm sales (at patent filing) have an annual value of $100 million or less and the other includes those with sales exceeding $5 billion. Regressions are run on these sub-samples for Models 5 and 8, results are displayed in Table 5.19.

For firms with sales above $5 billion, firm age, FDA experience, return on assets, dependent claims and firm publishing are statistically insignificant for Model 5, with odds ratios below one apart from dependent claims. Similar to the results for the whole sample total assets, inventor numbers, NPR, independent claims and university co-authoring are significant with odds ratios of a similar magnitude. However, total assets odds ratio is twice the size. The sub-sample's significance levels are also at the 1% level for all significant variables apart from inventors which is significant at the 5% level. The majority of these results hold for Model 8, although inventors becomes statistically insignificant. This sub-sample shows broadly similar results to those for the whole sample. Firm publishing and dependent claims no longer play an explanatory role, FDA experience has a significant negative influence on the likelihood of approval and firm size has an increasingly important positive influence.

For firms with sales below $100 million the results for Model 5 differ to those for the sample as a whole. Total assets, inventor numbers, NPR, dependent claims, firm publishing and university co-authoring are no longer a statistically significant variables. FDA experience now has a significant negative influence and firm age has a significant positive influence on the likelihood of approval. These results hold for Model 8, with the addition that total assets becomes statistically significant at the 10% level. The patent characteristics that have a significant influence on approval likelihood are independent claims and forward citations.

This subsection has sought to establish the extent to which the results of Models 5 and 8 reported in Tables 5.12 and 5.13 are robust or sensitive to changes in variable measurement techniques. Generally, apart from a limited number of changes in the magnitude and statistical significance levels that are observed and described above, the evidence is that the model results are essentially robust to changes in variable measurement techniques.

Table 5.19 Conditional Logistic Regression of Likelihood of Approval Results: Models 5 and 8

	Below $100 Million Sales		Above $5 Billion Sales	
Firm Age (AGE_{it})	1.422(0.078)***	1.405(0.018)***	0.963(0.050)	0.956(0.050)
FDA Experience ($FDAEXP_{it}$)	0.421(0.167)**	0.449(0.261)**	0.024(0.021)***	0.025(0.021)***
Total Assets ($SIZE_{it}$)	1.329(0.247)	1.407(0.278)*	4.656(1.912)***	4.839(1.991)***
Return on Assets (ROA_{it})	2.051(2.447)	1.649(2.335)	0.109(0.198)	0.119(0.219)
Number of Inventors ($INVENTORS_{it}$)	0.943(0.192)	0.904(0.138)	1.254(0.168)*	1.213(0.163)
Non-patent References (NPR_{it})	1.122(0.100)	1.113(0.070)	1.281(0.091)***	1.266(0.090)***
Independent Patent Claims ($ICLAIMS_{it}$)	1.326(0.199)*	1.292(0.136)*	1.552(0.178)***	1.513(0.174)***
Dependent Patent Claims ($DCLAIMS_{it}$)	1.070(0.117)	1.013(0.064)	1.139(0.100)	1.101(0.100)
Forward Citations Received (five-year window) (FC_{it})		1.454(0.090)***		1.193(0.087)**
Firm Publishing (PUB_{it})	0.000(0.000)	0.000(0.000)	0.969(0.027)	0.969(0.027)
University Co-authoring Share ($COAUTHAC_{it}$)	0.253(12.558)	0.266(0.064)	0.868(0.031)***	0.870(0.032)***
Observations	1,665	1,665	4,230	4,230
Number of Firms	42	42	21	21
Firm and Time Fixed Effects	Yes	Yes	Yes	Yes
Log Likelihood	-284.8	-278.7	-566.2	-553.3
McFaddens-R^2	0.192	0.210	0.127	0.131
Prob>Chi^2	0.0000	0.0000	0.0000	0.0000

Notes: The table presents the results of a conditional logit regression analysis with fixed effects where the dependent variable is the probability that a patent will be linked to an FDA drug approval (bivariate) and the control and the independent variables for columns 1 and 3 are those of Model 5 (see Table 5.12). Columns 2 and 4 use Model 8 (see Table 5.13). All variables are as defined in Chapter 3. Columns 1 and 2 use a sample of patents belonging to firms with sales below $x billion, whilst columns 3 and 4 use a sample of patents belonging to firms with sales above $y billion. AGE_{it} is the number of years since firm founding, $FDAEXP_{it}$ is 1 when the firm has experience of the FDA approval process prior to patent filing or 0 otherwise, $SIZE_{it}$ is the natural log of total assets in the year of patent filing, ROA_{it} is the firm's return on assets (net income divided total assets), winsorised at the 5^{th} and 95^{th} percentile, at the date of patent filing, $INVENTORS_{it}$ is the natural log of the number of inventors listed on the front of the patent document, NPR_{it} is the natural log of the number of NPR cited in the references cited section of the patent document, $ICLAIMS_{it}$ is the natural log of the number of independent claims listed at the end of the patent document, $DCLAIMS_{it}$ is the natural log of the number of dependent claims listed at the end of the patent document, FC_{it} is the natural log of the number of citations received from subsequent patents (forward citations) within a five-year period from the grant date of the cited patent, PUB_{it} is the firm's average annual publishing output of journal articles and $COAUTHAC_{it}$ is the firm's average annual publishing output of journal articles co-authored with academia, expressed as a percentage of annual publishing. Odds ratios are displayed with their standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

5.6 Conclusion

This chapter presented empirical results to investigate the relationship between patent and firm-level publishing characteristics and the likelihood of FDA drug approval for US-listed pharmaceutical firms. See Tables 5.20 and 5.21 for a summary of hypothesised and actual univariate and multivariate results. It was hypothesised that increases in certain patent-level characteristics, the number of inventors, non-patent prior art citations, independent patent claims, dependent patent claims and citations received, would increase the likelihood of approval. Whilst increases in others, specifically non-patent citation lag length, patent citation lag length and forward citation lag length, would reduce the likelihood of receiving an approval. It was also hypothesised that higher levels of firm publishing and co-authoring would also increase the likelihood of approval. Results are presented using different measures of firm size, prior art citation lags and citations received. The results of the present chapter find strong support for the hypothesised relationship between firm non-lag-related patent characteristics and the likelihood of approval. The results for citation lag length measures are found to be either insignificant or in a direction contrary to that hypothesised. As discussed, this finding could be attributable to the reduced sample size over which the model can be tested. Higher levels of firm publishing and university co-authoring reduce the likelihood of receiving an approval, which is also contrary to the relationships hypothesised. The results remain robust to alternative estimation techniques.

The publication-related results are contrary to prior studies such as Gittelman and Kogut (2003) and Fabrizio (2009) which found that publishing volume positively affects innovation. My findings are instead, more consistent with those of Kapoor and Klueter (2015) who found, in their global pharmaceutical study, that a firm's scientific publications have a negative impact on the likelihood of a drug entering clinical trials, but this relationship was statistically insignificant. The results for university co-authoring could be driven by the nature of the co-authoring relationship. Guler and Nerkar (2012) found that global collaborations reduced innovation whilst local collaborations improved innovation outcomes (new drug launch). This study measures university co-authoring levels but did not analyse the location of the collaborators.

In the next chapter (Chapter 6), the main aim is to test the extent to which the relationships found in this chapter can be utilised to predict the likelihood of FDA drug approval.

Table 5.20 A Summary Table of All Hypotheses and Univariate Results

Independent Variable	Hypothesis Number	Hypothesised Odds Ratio	Actual Result	Statistical Significance of Result	Conclusion (Hypothesis)
Variables Available at Patent Grant/Publication Date					
Number of Inventors ($INVENTORS_{it}$)	1a	>1	>1	Significant (1%)	Accepted
Non-patent References (NPR_{it})	1b	>1	>1	Significant (1%)	Accepted
Independent Patent Claims ($ICLAIMS_{it}$)	1c(i)	>1	>1	Significant (1%)	Accepted
Dependent Patent Claims ($DCLAIMS_{it}$)	1c(ii)	>1	>1	Significant (1%)	Accepted
NPR Citation Lag Length (NCL_{it})	1d	<1	<1	Insignificant	Rejected
Patent Citation Lag Length (PCL_{it})	1e	<1	<1	Insignificant	Rejected
Publishing (PUB_{it})	2a	>1	<1	Significant (1%)	Rejected
University Co-authoring Percentage ($COAUTHAC_{it}$)	2b	>1	<1	Significant (5%)	Rejected
Firm Co-authors with Other Firms ($COAUTHFIRM_{it}$)	2c	>1	<1	Significant (1%)	Rejected
Firm Funds Co-authored Publications ($FUNDER_{it}$)	2d	>1	Omitted	Insignificant	Rejected
Firm Size ($SIZE_{it}$)	4a	>1	>1	Significant (1%)	Accepted
Firm Age (AGE_{it})	4b	>1	>1	Significant (1%)	Accepted
Firm Profitability (ROA_{it})	4c	>1	>1	Significant (1%)	Accepted
Firm Prior Experience of the FDA Drug Approval Process ($FDAEXP_{it}$)	4d	>1	>1	Significant (1%)	Accepted
Variables Available After Patent Grant/Publication Date					
Forward Citations Received (five-year Window) (FC_{it})	3a	>1	>1	Significant (1%)	Accepted
Forward Citation Lag Length (FCL_{it})	3b	<1	<1	Significant (5%)	Accepted

Notes: The dependent variable in all regressions is FDA regulatory drug approval.

Table 5.21 A Summary Table of All Hypotheses and Multivariate Results

Independent Variable	Hypothesis Number	Hypothesised Odds Ratio	Actual Result	Statistical Significance of Result	Conclusion (Hypothesis)
Variables Available at Patent Grant/Publication Date (No Lagged Variables)					
Number of Inventors ($INVENTORS_{it}$)	1a	>1	>1	Significant (1%)	Accepted
Non-patent References (NPR_{it})	1b	>1	>1	Significant (1%)	Accepted
Independent Patent Claims ($ICLAIMS_{it}$)	1c(i)	>1	>1	Significant (1%)	Accepted
Dependent Patent Claims ($DCLAIMS_{it}$)	1c(ii)	>1	>1	Significant (1%)	Accepted
Publishing (PUB_{it})	2a	>1	<1	Significant (1%)	Rejected
University Co-authoring Percentage ($COAUTHAC_{it}$)	2b	>1	<1	Significant (1%)	Rejected
Firm Co-authors with Other Firms ($COAUTHFIRM_{it}$)	2c	>1	<1	Significant (1%)	Rejected
Firm Funds Co-authored Publications ($FUNDER_{it}$)	2d	>1	Omitted	Insignificant	Rejected
Firm Size ($SIZE_{it}$)	4a	>1	>1	Significant (1%)	Accepted
Firm Age (AGE_{it})	4b	>1	<1	Significant (1%)	Rejected
Firm Profitability (ROA_{it})	4c	>1	>1	Insignificant	Rejected
Firm Prior Experience of the FDA Drug Approval Process ($FDAEXP_{it}$)	4d	>1	<1	Insignificant	Rejected
Lagged Variables					
NPR Citation Lag Length (NCL_{it})	1d	<1	>1	Significant (10%)	Rejected
Patent Citation Lag Length (PCL_{it})	1e	<1	<1	Insignificant	Rejected
Variables Available After Patent Grant/Publication Date					
Forward Citations Received (five-year Window) (FC_{it})	3a	>1	>1	Significant (1%)	Accepted
Forward Citation Lag Length (FCL_{it})	3b	<1	>1	Significant (10%)	Rejected

Notes: The dependent variable in all regressions is FDA regulatory drug approval.

Chapter 6 Assessment of Model Predictive Ability

6.1 Introduction

Having modelled the relationship between the likelihood of drug approval and various explanatory variables in the prior chapter, this chapter presents an assessment of the potential predictive ability of the observed relationship between certain patent and firm-level publishing characteristics and the likelihood of a patent being linked to an approved drug. The objective here is to examine whether the results reported in Chapter 5 can be utilised to predict whether a patent is likely to be linked to an approved drug.

The remainder of the chapter is organised as follows. Section 6.2 provides a brief outline of the methods utilised to assess the predictive ability of the model. This is done by conducting in-sample analyses of the model, and by imitating the real-life application of the model using out-of-sample tests. Section 6.3 presents the main empirical results for an in-sample analysis and discusses the findings of these predictive assessments. Section 6.4 presents the main empirical results for an out-of-sample analysis and discusses the results of several robustness checks and Section 6.5 concludes the chapter.

6.2 Assessing the Predictive Ability of the Model

The models outlined previously in Chapter 5, Section 5.4.2 demonstrate that there is a statistically significant relationship between the likelihood of a patent being linked to an approved drug and certain firm patent and publishing characteristics. These findings suggest that the model can explain a significant proportion of the variation in patent drug approval-related outcomes. Whilst the model may possess explanatory power in describing the relationship between these characteristics and the likelihood of a drug approval link, it is also desirable that the model should also possess a level of predictive ability.

I employ two methods, in-sample and out-of-sample analysis, to assess the predictive ability of my model. Researchers have found that in-sample analysis can produce an overly optimistic assessment of a model's predictive ability. Mosteller and Tukey (1977) considered that assessing the model with the data used in its estimation process tends to overestimate its predictive performance. With in-sample analysis, prediction errors always tend to be lower than out-of-sample prediction errors, as the errors in model prediction for data withheld from the model's estimation are greater than the prediction errors made by a model containing the model estimation data (Fortmann-Roe, 2012). This leads many to consider out-of-sample model testing, using data that are not part of the estimation process,

as a crucial test for any statistical model. According to Ashley, Granger and Schmalensee (1980), empirical evidence based on an out-of-sample analysis of the model's predictive performance is considered more reliable than evidence based on in-sample predictive performance. Although Inoue and Kilian (2004) disputed this viewpoint and argued that in-sample testing is a more credible technique than out-of-sample analysis. I therefore also conduct an out-of-sample analysis on the model, with and without forward citations, to validate my in-sample analysis results.

The model being tested is formulated based on the data from the complete sample of 12,957 patents. I anticipate that the results of the in-sample analysis will demonstrate a superior predictive ability to that of the out-of-sample analysis. This difference in performance can be attributed to the fact that the data utilised to both estimate the model and test the model's predictive ability in the in-sample analysis are drawn from the same sample and produce a more optimistic assessment of the model's predictive ability.

6.3 In-Sample Analysis of the Model's Predictive Ability

I test two versions of my model; one version only includes the patent characteristics that are available at the time of patent publication, excluding citations received from subsequent patents (forward citations). See Chapter 5, Section 5.4.2.1.5 for Model 5 details. This allows a patent's predictive ability to be assessed as soon as it is published without the need to wait for the patent to receive citations from other patents, which may take several years. As the goals of this research are to provide a project decision tool for managers and a portfolio selection mechanism that provides a profitable investment strategy for investors, this early availability of predictive outcomes is beneficial. It allows managers to allocate scarce resources in a more informed manner and investors can select and hold potentially profitable stocks for a longer period, increasing the returns to their portfolio. I also test a version of the model that includes these forwards citations. See Chapter 5, Section 5.4.2.2.1 for Model 8 details. These models are also tested by excluding the controls for firm age, prior FDA experience, firm size and profitability. This allows me to assess whether it is the patent and firm-level publishing characteristics that are the main drivers of the predictive ability of the model and to determine that the control variables have no influence on this ability.

The results of this in-sample predictive ability assessment are presented in an accuracy matrix (Altman, 1968) format. In this analysis, I am testing whether any of my model formulations can correctly predict the existence of an approval-linked patent. When

assessing the predictive ability of these models I want to measure both the sensitivity and specificity measures for each model. As in Jones, Johnstone and Wilson (2017), I define sensitivity as the ratio of the number of correctly predicted approval-linked patents (labelled D in Table 6.1) to the total of actual approval-linked patents (C+D). Specificity is the ratio of correctly predicted non-approval-linked patents (labelled A in Table 6.1) to the total of actual non-approval-linked patents (A+B). The null hypothesis is that the patent is linked to an approved drug. The matrix displays the classification results produced by the in-sample analysis. The number of false positives (Type II errors, labelled B in Table 6.1), where patents that are not linked to an approved drug are misclassified as approval-linked and the number of false negatives (Type I errors, labelled C in Table 6.1), where approval-linked patents are misclassified as non-approval linked. The number of true negatives and true positives (A+ D) as a proportion of the total sample measures the overall accuracy of the model. The in-sample analysis permits the models to provide predicted likelihood measures. These predict the likelihood that a patent will be linked to an approved drug and the likelihood that no such link exists.

Table 6.1 A Patent Approval-Linked Outcome: Comparison of Actual and Predicted Values

	Prediction: No Approval Link (Coded 0)	Prediction: Approval Link (Coded 1)
Actual: No Approval Link (Coded 0)	A (True Negative)	B (False Positive-Type II Error)
Actual: Approval Link (Coded 1)	C (False Negative-Type I Error)	D (True Positive)

Notes: This table displays the patents that are classified correctly as not connected to a drug approval in Cell A, a true negative result. Cell B displays non-approval linked patents misclassified as approval-linked. These are false positives and are statistically classified as Type II errors. Cell C displays the approval-linked patents that are misclassified as having no approval link. These are false negatives and are statistically classified as Type I errors. Finally, cell D displays correctly classified drug approval-linked patents, a true positive result.

The predicted probabilities saved from the conditional logistic regressions are only probabilities. To use the model as a project decision tool or a portfolio selection mechanism the probability of an outcome occurring is not sufficient. To predict that a patent will be linked to an approved drug or not is based on selecting appropriate cut-off values for these probabilities with which such outcomes can be inferred. Any patent with an assigned probability equal to or higher than the chosen cut-off probability is classified as possessing a drug approval link and those patents with probabilities below the cut-off are classified as

not possessing a drug approval link. The choice of cut-off value can affect both the level of Type I and Type II errors.

Before discussing the optimal cut-off probability, it is important to define Type I and Type II errors in this context. The Type I error occurs when a patent is predicted to not be linked to an approved drug when such a link does exist, while a Type II error occurs when a patent is predicted to be linked to an approved drug when such a link does not exist. The assessment of the importance of the costs involved with both error types created some controversy in the literature. Palepu (1986) assumed that the costs of these error types are identical, whilst Barnes (1999) suggested that they should be weighted differently. According to Zhou, Obuchowski and McClish (2011) the overall accuracy of the model's predictive ability also depends on the cut-off value selected. There is no definitive rule for the selection of cut-off values. Assuming Type I and II errors are equal in terms of costs and benefits, Hosmer, Lemeshow and Sturdivant (2013) suggested that a cut-off value where Type I and Type II errors are equal should be selected whilst Ohlson (1980), Palepu (1986) and Espahbodi and Espahbodi (2003) recommended a cut-off value where both Type I and Type II errors are minimised. Barnes (1998, 1999) and Powell (2001) propose a differing view. They suggested that the assumption that that the costs of Type I and Type II errors are equal is unrealistic. If the goal of this model is to deliver a profitable investment strategy to investors, then the cut-off value selected should be that that maximises investor returns or minimises investor losses. This cut-off value would lie where the concentration of accurately predicted approval-linked patents (true positives) within the predicted approval sample would be at its highest level and where Type II errors (false positives) are at their lowest level. The choice of cut-off value depends primarily on the requirements of the analysis.

To assess model predictive ability, I firstly conduct an in-sample prediction test on the model with patent characteristic variables (excluding forward citations), firm-publishing behaviour variables and control variables. Utilising Stata's "predict" command, I produce and save predicted values for this model. In prior literature some studies have used a predicted probability values of 0.5 as a cut-off value (see Cramer (1999) for a discussion). However, this model is attempting to predict a relatively rare event, as approval-linked patents comprise 4.88% of the total patent sample. This means that both the in-sample and out-of-sample datasets have unequal numbers of approval-linked and non-approval-linked patents, making the use of a cut-off value of 0.5 inappropriate. I follow Powell (2001) and rank the predicted probability values into deciles; these deciles are used to select a range of cut-off

values at which the predictive ability of the model is then assessed (see Section 4.6.2.1 for a discussion of the methodology employed). The model's predictive ability is reported in tables at the highest and lowest 10% and 20% of these predicted values, and at the 50% level. This range of values provides an assessment of a model's predictive ability at both extremes of the likelihood measure, and at the mid-point. It can then be observed which cut-off value is optimal and provides the best combination of overall accuracy and low levels of Type I or Type II errors.

As can be observed from Table 6.2, as the cut-off value increases the number of correctly predicted approval-linked patents (true positives) decreases, but the approval concentration ratio (ACR) increases indicating that the upper 10% cut-off value is the optimal value. The number of patents correctly predicted to have no approval link (true negatives) increases, Type I errors (false negatives) increase, and Type II errors (false positives) decrease. Increasing the cut-off value increases the model's overall accuracy, increasing Type I errors and reducing Type II errors, where patents with no approval link are incorrectly classified as approval-linked.

It can be observed that at the upper 10% cut-off value the model accurately predicts 91.56% of patents without a link to an approved drug and 40.51% of patents linked to approved drugs, with overall predictive accuracy (true positives and negatives as a proportion of the whole sample) of 89.07%. The approval concentration (true positives as a proportion of all predicted approval-linked) is 19.75%. At the upper 20% cut-off value the model correctly predicts 81.47% of patents with no link to an approved drug and now predicts the approval-linked patents with an increased accuracy rate of 48.89%, however the overall accuracy and approval concentration drop to 79.89% and 11.92%, respectively. At the 50% level, the model accurately predicts 71.36% of approval-linked patents with an overall accuracy of 52.09% and approval concentration of 6.96%. Next, the lowest 10% and 20% of the predicted values are used as cut-off thresholds to assess the accuracy of the model's prediction ability. The model accurately predicts 79.44% of patents with no link to an approved drug at the lowest 20% cut-off value and 89.64% of these patents at the lowest 10% cut-off value. At the lowest 20% and 10% cut-off values, the model correctly predicts approval-linked patents with an accuracy rate of 9.97% and 3.32% respectively. The overall accuracy is 76.05% at the lowest 20% and 85.43% at the lowest 10% cut-off values. Approval concentrations are 2.43% at the lowest 20% and 1.62% at the lowest 10% cut-off values. These results are not unexpected as the portfolios created using the lowest cut-off

values contain patents that are predicted to be the least likely to be linked to an approved drug and therefore the model finds it difficult to accurately predict approval-linked patents.

Table 6.2 Model 5- In-Sample Analysis of Model Predictive Ability

Actual	**Prediction (Without Forward Citations Included)**		
	Non-Approval	Approval	Predicted Values: Cut-off
Non-Approval	11,285	1,040	0.010689 (Upper 10%)
	91.56%†	8.44%	
Approval	376	256	
	59.49%	40.51%+	
Approval Concentration Ratio 19.75%		Overall Accuracy 89.07%	
Non-Approval	10,041	2,284	0.0051503 (Upper 20%)
	81.47%†	18.53%	
Approval	323	309	
	51.11%	48.89%+	
Approval Concentration Ratio 11.92%		Overall Accuracy 79.89%	
Non-Approval	6,298	6,027	0.0015088 (50%)
	51.10%†	48.90%	
Approval	181	451	
	28.64%	71.36%+	
Approval Concentration Ratio 6.96%		Overall Accuracy 52.09%	
Non-Approval	9,791	2,534	0.0007635 (Lowest 20%)
	79.44%†	20.56%	
Approval	569	63	
	90.03%	9.97%+	
Approval Concentration Ratio 2.43%		Overall Accuracy 76.05%	
Non-Approval	11,048	1,277	0.0005618 (Lowest 10%)
	89.64%†	10.36%	
Approval	611	21	
	96.68%	3.32%+	
Approval Concentration Ratio 1.62%		Overall Accuracy 85.43%	
† Correctly Predicted True Negative (Specificity), + Correctly Predicted True Positive (Sensitivity)			
Whole Sample: 12,957 Patents (12,325 Non-Approvals and 632 Approvals)			

Notes: This table displays the patents that are classified as drug approval-linked patents or patents with no link to an approved drug. These classifications are based on cut-off thresholds selected from the predicted probabilities of the conditional logistic regression for Model 5. The classifications displayed are based on an in-sample analysis where Model 5 is estimated using all patent observations in the sample.

I next test the predictive ability of the model including forward citations (Model 8 in Chapter 5, Section 5.4.2.2.1) as a patent characteristic variable, firm-publishing behaviour variables and control variables are still present. The results in Table 6.3 demonstrate a very similar set of results to those observed for the model excluding forward citations. This model is marginally higher in its predictive ability at the upper 10% and 20% cut-off values in terms of overall accuracy, approval concentration, Type I and Type II errors. This model increases approval predictive ability by only 0.47% at the upper 10% cut-off value, by 3.17% at the upper 20% cut-off value and by 3.96% at the 50% cut-off value. Approval concentration is also marginally higher at 19.98 % at the upper 10% cut-off value, 12.69% at the upper 20% cut-off value and 7.35% at the 50% cut-off value. At the lowest 10% and 20% cut-off values, the results are virtually identical to those for the model that excludes forward citations. It can be observed that although the addition of forward citations to the model does improve predictive performance, the improvement is marginal. Increases in model complexity do not necessarily result in a more accurate forecasting ability. Forward citations alone do not appear to be the main predictor of a patent's likelihood of having a link to an approved drug. The other patent and firm-level publishing characteristics also appear to play an important role.

The prediction of approval-linked patents is a more difficult task than that of predicting patents with no link to an approved drug as approval-linked patents comprise just 4.88% of the total sample. However, these results confirm that both versions of the model can accurately predict these patents as well as those patents with no link to an approved drug, with a predictive ability that cannot be solely attributed to chance. The model without forward citations has an approval concentration of 19.75% that is significantly higher than the 4.88% that would have been achieved by chance alone. Although these results indicate that the models predict approval-linked patents better than chance, they also exhibit a strong ability to identify non-approval- linked patents. Both models predict non-approval-linked patents with an accuracy of almost 91.6% at the upper 10% cut-off value. This ability to predict both the existence and non-existence of an approval linkage, is necessary to generate excess returns from a portfolio selected based on these predicted probabilities. It can also be observed that the addition of forward citations does not markedly increase the model's predictive ability. This result means that patent characteristics other than forward citations are also important in predicting the likelihood of a patent being linked to an approved drug. The likelihood of a link to an approved drug is not predominantly driven by a patent possessing forward citations. It also means that the model can be used to predict this approval

linkage likelihood at an earlier point in the patent's lifetime, as it is not necessary to await the flow of citations from subsequent patents.

Table 6.3 Model 8-In-Sample Analysis of Model Predictive Ability

Actual	**Prediction (With Forward Citations Included)**		
	Non-Approval	Approval	Predicted Values: Cut-off
Non-Approval	11,288	1,037	0.0105772 (Upper 10%)
	91.59%†	8.41%	
Approval	373	259	
	59.02%	40.98%+	
Approval Concentration Ratio 19.98%		Overall Accuracy 89.12%	
Non-Approval	10,062	2,263	0.0051004 (Upper 20%)
	81.64%†	18.36%	
Approval	303	329	
	47.94%	52.06%+	
Approval Concentration Ratio 12.69%		Overall Accuracy 80.20%	
Non-Approval	6,321	6,004	0.0014459 (50%)
	51.29%†	48.71%	
Approval	156	476	
	21.68%	75.32%+	
Approval Concentration Ratio 7.35%		Overall Accuracy 52.46%	
Non-Approval	9,769	2,556	0.0006149 (Lowest 20%)
	79.26%†	20.74%	
Approval	595	37	
	94.15%	5.85%+	
Approval Concentration Ratio 1.43%		Overall Accuracy 75.68%	
Non-Approval	11,041	1,284	0.000436 (Lowest 10%)
	89.58%†	10.42%	
Approval	617	15	
	97.63%	2.37%+	
Approval Concentration Ratio 1.15%		Overall Accuracy 85.33%	
† Correctly Predicted True Negative (Specificity), + Correctly Predicted True Positive(Sensitivity)			
Whole Sample: 12,957 Patents (12,325 Non-Approvals and 632 Approvals)			

Notes: This table displays the patents that are classified as drug approval-linked patents or patents with no link to an approved drug. These classifications are based on cut-off thresholds selected from the predicted probabilities of the conditional logistic regression for Model 8. The classifications displayed are based on an in-sample analysis where Model 8 is estimated using all observations in the sample.

The ability to predict probable patent outcomes relatively early in a project's life cycle could prevent the misallocation of firm resources and improve investor returns. Managers would possess additional information that would allow research project funding to be more accurately targeted on projects with an increased probability of approval success, whilst funding would not be wasted on drug candidates being put through to the clinical trial stage of a project that is less likely to succeed. In addition, investors would be able to form a potentially profitable portfolio at an earlier point in time, allowing returns to be earned over a longer period.

6.3.1 Cost Implications of Model Misclassifications

The analysis, so far, has assumed that the error of misclassifying approval-linked patents is equal in costs/benefits to misclassifying those patents without a link to an approved drug. If this is not the case, then simply examining the misclassification rates is not sufficiently informative and may even be misleading.

The costs for incorrect predictions may be very different for Type I errors than for Type II errors. For example, those patents that are misclassified as approval-linked, 1,040 patents at the upper 10% cut-off value in Table 6.2 will proceed from pre-clinical trials into phase I trials if project decisions are informed by the model's predictions. This incurs additional costs as, if they were correctly classified; they would not have progressed to this project stage and would have been abandoned after the pre-clinical trials. There is also a potential opportunity cost of underfunding or delaying the development of other promising drug candidates in an environment where there are resource limitations. Also, the approval-linked patents that are misclassified as having no drug link, 376 patents, will be abandoned and will not be progressed to further drug trials and eventual approval. The firm holding this patent may avoid clinical trial costs, but it also sacrifices the expected future profits from the commercialisation of a drug candidate linked to this patent.

It is important to note that drugs are usually linked to multiple patents. Ouellette (2010) found that in 2005 drugs were linked, on average to 3.5 patents and that these numbers had been increasing over time. I therefore assume that a drug is linked to 4 patents. This would result in the drug candidates linked to the false positive patents progressing to 260 unnecessary phase I trials and 94 drug candidates would not be developed further. Both these actions incur costs, with estimates for the out-of-pocket expenses of a phase I trial ranging from $4m (Sertkaya, Birkenbach, Berlind and Eyraud, 2014) to $23m (Mestre-Ferrandiz,

Sussex and Towse, 2012), with an average cost of around $12m. Total phase I trial costs would therefore increase by $3,120m due to this misclassification. There are also cost implications for misclassified approval-linked patents. According to IMSHealth[73] the average annual sales of the top 100 drugs in 2011 was $1,900m (lowest annual sales were $680m). However, according to Grabowski and Vernon (1994), only 3 in 10 marketed drugs match or exceed their R&D costs and a report by the Pharmaceutical Research and Manufacturers of America (PhRMA) in 2013 put the figure as low as 2 in 10. Assuming that of the 94 drug candidates 28 would have covered their costs and made a profit, even at the lowest profit estimate of $680m, a conservative estimate of the total loss of potential revenue would be $19,040m. This implies that errors in classifying approval-linked patents are costlier than those for misclassifying non-approval-linked patents.

It may be more important to managers to select a lower cut-off value that produces a higher level of approval predictive accuracy but reduces overall predictive accuracy and the approval concentration ratio and in so doing increases the Type II error rate and reduces the Type I error rate. Whilst overall accuracy is reduced where a lower cut-off value is selected, the opportunity costs to the firm (not investing in a potentially successful drug candidate) are reduced. This decision would increase phase I trial costs but reduce the loss of potential future profits.

This finding, whilst informative should not mean that project decisions be solely based on predicted probabilities. Decisions concerning drug development should be based on these findings, traditional clinical success rate metrics and specific considerations linked to the drug candidate being investigated. Managers must make the final decision as whether to proceed with a research project but can do so in a more informed manner by using this model. Assessing the probability of success using this method may avoid the unnecessary testing costs associated with failed drug candidates, assisting in more logical resource allocation and potentially limiting risk-taking activity that consumes R&D resources.

However, if the main goal of the prediction phase of this analysis is to construct a portfolio of firm stocks whose patents are more likely to be linked to an approved drug an investor wants such a portfolio to hold as many correctly predicted successful stocks as possible, as this would produce the highest rate of portfolio returns. If the predictive ability of any of the

[73] https://www.drugs.com/stats/top100/2011/sales, accessed on 21st April 2017.

models is to be utilised to formulate a successful investment strategy then a cut-off value must be selected that increases the proportion of correctly classified approval-linked patents, so increasing the proportion of potentially profitable stocks in the portfolio that is constructed based on model predictions. A higher cut-off value is preferable in this instance and this is the focus of the portfolio construction approach utilised in this study.

6.3.2 Model Predictive Ability: In-Sample Analysis Excluding Control Variables

As outlined in Table 6.2 the model excluding forward citations possesses predictive ability. I next employ in-sample analysis on a version of this model that excludes the control variables for firm age, prior FDA experience, firm size and profitability. This determines whether the predictive ability of the model is driven by the patent and firm-level publishing characteristics or if the control variables exert an influence on the likelihood of a patent being linked to an approved drug. A procedure similar to that employed previously in this chapter is applied to this control-free model, with the upper (and lower) 10% and 20% levels and 50% level of predicted values again being employed as cut-off values to assess the model's predictive ability. The results are displayed in Table 6.4.

At the upper 10% (20%) cut-off value, the model accurately predicts 91.61% (81.54%) of patents with no link to an approved drug and 41.61% (50.16%) of patents linked to approved drugs. Overall accuracy is 89.17% (80.00%) and approval concentration is 20.28% (12.23%). When this in-sample analysis is applied to the model excluding controls but with the addition of forward citations, the model's predictive ability is almost identical to that for the model including controls (see Table 6.3). With the model accurately predicting 91.55% (81.74%) of patents with no link to an approved drug and 40.35% (52.53%) of patents linked to approved drugs at the upper 10% (20%) levels. Overall accuracy is 89.06% (80.32%) and approval concentration is 19.68% (12.86%) at these levels.

These results confirm that the predictive ability of the model is not influenced by the control variables, but that this ability is attributable to the patent and firm-level publishing characteristics. It can also be observed that the inclusion of forward citations only increases the accuracy in predicting approval-linked patents by less than 10% and only at the upper 20% and 50% cut-off values. At the upper 10% cut-off value, accuracy drops slightly by just over 1%. This again demonstrates that whilst the forward citation variable is a positive and

significant explanatory variable in Model 8, the variable is not a critical contributor to the predictive ability of the model. Whilst its inclusion does improve prediction accuracy, it only does so to a relatively small extent.

Table 6.4 In-Sample Analysis of Model Predictive Ability (Control Variable and Forward Citations Excluded)

Actual	**Prediction (Without Controls-Without Forward Citations)**		
	Non-Approval	Approval	Predicted Values: Cut-off
Non-Approval	11,291	1,034	0.0125979 (Upper 10%)
	91.61%†	8.39%	
Approval	369	263	
	58.39%	41.61%+	
Approval Concentration Ratio 20.28%		Overall Accuracy 89.17%	
Non-Approval	10,049	2,276	0.005735 (Upper 20%)
	81.54%†	18.46%	
Approval	315	317	
	49.84%	50.16%+	
Approval Concentration Ratio 12.23%		Overall Accuracy 80.00%	
Non-Approval	6,286	6,039	0.0015585 (50%)
	51%†	49%	
Approval	193	439	
	30.54%	69.46%+	
Approval Concentration Ratio 6.78%		Overall Accuracy 51.90%	
Non-Approval	9,780	2,545	0.0007714 (Lowest 20%)
	79.35%†	20.65%	
Approval	579	53	
	91.61%	8.39%+	
Approval Concentration Ratio 4.08%		Overall Accuracy 75.89%	
Non-Approval	11,049	1,276	0.0005893 (Lowest 10%)
	89.65%†	10.35%	
Approval	612	20	
	96.94%	3.16%+	
Approval Concentration Ratio 1.54%		Overall Accuracy 85.43%	
† Correctly Predicted True Negative (Specificity), + Correctly Predicted True Positive (Sensitivity)			
Whole Sample: 12,957 Patents (12,325 Non-Approvals and 632 Approvals)			

Notes: This table displays the patents that are classified as drug approval-linked patents or patents with no link to an approved drug. These classifications are based on cut-off thresholds selected from the predicted probabilities of the conditional logistic regression for Model 5 without control variables. The classifications displayed are based on an in-sample analysis where the model is estimated using all observations in the sample.

The in-sample analysis results support the premise that patent and firm-level publishing characteristics can assist in predicting the likelihood that a patent will be linked to an approved drug. They also accurately predict those patents with no link to an approved drug with an accuracy of over 80% and 90% at the upper 20% and 10% cut-off values respectively.

6.3.3 Model Predictive Ability: In-Sample Analysis Robustness Checks

This section presents the results of a series of robustness and sensitivity checks. The objective of these checks is to establish whether the results reported for the in-sample analysis are robust. As this is an in-sample analysis robustness test the sample cannot be partitioned but its composition can be altered. This is done by varying the firms in the sample.

The results presented in Tables 6.2 and 6.3 were based on a sample that included all the patents from the 81 firms in the sample. As a robustness check, I firstly examined firms with differing levels of patenting activity to observe if this affects the predictive ability of the model without forward citations. Alternative firm samples are utilised to assess if the in-sample analysis results are sensitive to a firm's patenting activity levels. Two firm samples were tested; firms with above median patenting levels of 42 patents and firms with below median patenting levels. There are 40 firms in the above median sample and 41 firms in the below median sample. The above median sample contains 444 approval-linked patents (3.62% of this 12,259-patent sample) and the below median sample contains 188 approval-linked patents (26.93% of this 698-patent sample). Both samples cover a similar time period. The results are displayed in Table 6.5.

The results for firms with above and below median patenting activity levels were very similar to the results displayed in Table 6.2. Results for the median-related samples were broadly similar for the prediction of patents with no link to an approved drug. The model is continuing to accurately predict 75-80% of these patents at the lowest 20% level and around 90% at the lowest 10% level. The overall accuracy for the above median sample is 88.16% at the 10% level and 78.84% at the 20% level for Model 5 (excludes forward citations). The prediction of patents linked to approved drugs exhibits reduced accuracy levels at both at the upper10% and 20% cut-off values with approval concentration ratios of 8.70% (10% level) and 6.16% (20% level), although this could be attributable to the reduced number of firms

Table 6.5 Robustness Checks for In-Sample Analysis

	Below Median (41 Firms)	**Above Median (40 Firms)**
Correct Prediction of Patents Linked to Approved Drugs		
Model without Forward Citations:		
Upper 10% Cut-off Value-Correctly Classified Approval-linked Patents	20.74%	23.87%
Approval Concentration Ratio	55.71%	8.70%
Overall Accuracy	74.21%	88.16%
Upper 20% Cut-off Value Correctly Classified Approval-linked Patents	29.79%	34.01%
Approval Concentration Ratio	40%	6.16%
Overall Accuracy	69.05%	78.84%
Model with Forward Citations		
Upper 10% Cut-off Value Correctly Classified Approval-linked Patents	21.28%	27.25%
Approval Concentration Ratio	57.14%	9.86%
Overall Accuracy	74.50%	88.34%
Upper 20% Cut-off Value Correctly Classified Approval-linked Patents	30.85%	39.41%
Approval Concentration Ratio	41.43%	7.14%
Overall Accuracy	69.63%	79.23%

Notes: This table displays the patents that are classified as drug approval-linked patents. These classifications are based on cut-off thresholds selected from the predicted probabilities of the conditional logistic regression for Model 5 (without forward citations) and Model 8 (with forward citations). The classifications displayed are based on an in-sample analysis where Models 5 and 8 are estimated using all observations in the below median patenting sample and again using all observations in the above median patenting sample.

being analysed or to the fact there is a lower proportion of approval-linked patents in this sample, 3.62% compared to 4.88% in the total sample. The model estimated and tested on the above median sample has greater overall predictive ability than the below median sample, with an overall accuracy rate of 88.16% compared to 74.21% at the upper 10% level. However, its approval concentration ratios are much higher, with an ACR of 55.71% and 40% at the upper 10% and 20% levels. The below median sample does have a much higher proportion of approval-linked patents in the sample, 26.93%.

The results for the model including forward citations are broadly similar with overall accuracy of 88.34% at the 10% level and 79.23% at the 20% level for Model 8 (including forward citations). Their addition improves predictive ability (approval concentration ratios) in the above median samples by 1.26% and 0.98% at the upper 10% and 20% levels respectively. Overall accuracy levels of the model with forward citations are lower compared to the above median sample, at 74.50% and 69.63% at the upper 10% and 20% levels. The approval concentration ratios are again much higher, with an ACR of 57.14% and 41.43% at the upper 10% and 20% levels, respectively.

These results demonstrate that a firm's level of patenting activity does not directly affect the overall accuracy of the predictive ability of the models. The above median sample contains many more patents than the below median sample but there is very little difference in the overall accuracy of the models based on these very dissimilar samples. However, the approval concentration ratio is greatly increased for the models when applied to the below median sample. This evidence supports the premise of this study, confirming that it is the characteristics of the patents the firm produces that have a predictive role and not necessarily the quantity of patents a firm produces.

6.4 Out-of-Sample Analysis of the Model's Predictive Ability

Good predictive ability, as assessed by in-sample analysis, is very useful in assessing the goodness of fit of the model but it does not exclude the possibility that the model may suffer from overfitting and it does not necessarily provide the most accurate predictions. To assist in excluding these possibilities, an out-of-sample analysis is also conducted to assess the model's predictive ability. It can be argued that the best performance indicator for a predictive model is how well the model is able to predict the occurrence of an approval-linked patent out-of-sample. Out-of-sample testing has grown to become the accepted method for validating the predictive ability of a model in finance research[74]. The focus of this section is to evaluate the model's ability to predict approval-linked patents using out-of-sample analyses. Performing an out-of-sample analysis is a good way to test the assumptions of the model and to test its predictive ability. To conduct this type of analysis, I withhold a portion of the sample data from the model estimation process, then using the resultant model formulation I make predictions for those patents that are excluded from the estimation sample. This process assesses how accurate these predictions are and determines whether they are similar to those made using the in-sample analysis approach.

The focus of the out-of-sample analysis is on Models 5 and 8 specified in Chapter 5, Sections 5.4.2.1.5 and Section 5.4.2.2.1. I estimate these models based on a subset of the total patent sample and using this to predict outcomes for observations that are not part of the estimation sub-sample. This prediction sub-sample is then used to evaluate the model's predictive performance. There are no definitive rules for how to select the estimation and test sub-samples. They can be based either on the number of observations or on a time measure. The

[74] See also, Zavgren (1985), Palepu (1986), Barnes (1998, 1999), Powell (2001), Shumway (2001) and Campbell, Hilscher and Szilagyi (2008).

estimation sample is usually larger than the test sample but there is nothing preventing the selection of a test sample that is larger than the estimation sample. According to Kuhn and Johnson (2013), in datasets with binary dependent variables where the frequency of positive outcomes is substantially different from 50%, such as this one where approval-linked patents comprise 4.88% of the sample, it can be advisable to choose an estimation sample that closely represents the characteristics of the whole original data sample. For my first out-of-sample test, I split my data based on time using only the observations for patents with grant years up to and including 1996. This produces an estimation sample containing a total of 8,473 patents (approximately 65% of the total sample) with 282 approval-linked patents (3.2%). This sub-sample has a similar dependent variable profile to the whole sample and I conduct robustness tests that vary this split later in this chapter. The "predict" command is again used to produce predicted probability values but it is now restricted to patents with grant years from 1997 onwards, which total 4,484 patents (approximately 35% of the total sample) with 350 approval-linked patents (7.8%). The model's predictive ability is again assessed using similar techniques to those employed for in-sample testing and reported at the upper and lower 10% and 20% cut-off values and also at the 50% cut-off. The results for Model 5 (excluding forward citations) are displayed in Table 6.6 and results for Model 8 (including forward citations) are displayed in Table 6.7.

The predictive ability of Model 5 was assessed first. It can be observed that the out-of-sample analysis produces similar results to the in-sample analysis when predicting the likelihood that a patent is not linked to an approved drug. However, the results for predicting patents linked to an approved drug have lower accuracy levels at the upper 10%, upper 20% and 50% cut-off levels. At the lowest 10% and 20% levels, the accuracy of the out-of-sample analysis is greater than that for the in-sample analysis. At the upper 10% level, the model accurately predicts 22.86% of patents linked to approved drugs compared to 40.51% for the in-sample analysis. At the upper 20% level, the model now predicts approval-linked patents with an accuracy rate of 36.29%; the in-sample predictive ability is 48.89%. The overall accuracy is 85.75% at the highest 10% and 77.83% at the highest 20% cut-off values. Approval concentrations are 17.82% at the highest 10% and 14.14% at the highest 20% cut-off values. This compares to 19.75% at the highest 10% and 11.92% at the highest 20% levels in the in-sample analysis. At the lowest 10% and 20% levels, the model's predictive ability for patents linked to an approved drug is superior to that of the in-sample analysis. The model accurately predicts 22.29% (9.97% for in-sample) at the 20% level and 11.71%

(3.32% for in-sample) at the 10% level. These higher levels of accuracy for the in-sample analysis at the upper cut-off values are not unexpected as the predictive ability of a model

Table 6.6 Model 5-Out-of-Sample Analysis of Model Predictive Ability

Actual	**Prediction (Without Forward Citations Included)**		
	Non-Approval	Approval	Predicted Values: Cut-off
Non-Approval	3,765	369	0.0333401 (Upper 10%)
	91.07%†	8.93%	
Approval	270	80	
	77.14%	22.86%+	
Approval Concentration Ratio 17.82%		Overall Accuracy 85.75%	
Non-Approval	3,363	771	0.0152897 (Upper 20%)
	81.35%†	18.65%	
Approval	223	127	
	63.71%	36.29%+	
Approval Concentration Ratio 14.14%		Overall Accuracy 77.83%	
Non-Approval	2,090	2,044	0.0045499 (50%)
	50.56%†	49.44%	
Approval	151	199	
	43.14%	56.86%+	
Approval Concentration Ratio 8.87%		Overall Accuracy 51.05%	
Non-Approval	3,315	819	0.0022048 (Lowest 20%)
	80.19%†	19.81%	
Approval	272	78	
	77.71%	22.29%+	
Approval Concentration Ratio 8.70%		Overall Accuracy 75.67%	
Non-Approval	3,725	409	0.0011415 (Lowest 10%)
	90.11%†	9.89%	
Approval	309	41	
	88.29%	11.71%+	
Approval Concentration Ratio 9.11%		Overall Accuracy 83.99%	
† Correctly Predicted True Negative (Specificity), + Correctly Predicted True Positive (Sensitivity)			
Patents Granted up to 1996 Used to Predict Outcomes for Patents Granted from 1997 Onwards. Sample 1997 Onwards: 4,484 Patents (4,134 Non-Approvals and 350 Approvals)			

Notes: This table displays the patents that are classified as drug approval-linked patents or patents with no link to an approved drug. These classifications are based on cut-off thresholds selected from the predicted probabilities of the conditional logistic regression for Model 5. The classifications displayed are based on an out-of-sample analysis where Model 5 is estimated using patents with grant years up to and including 1996. The test sub-sample includes patents with grant years of 1997 onwards.

often reduces when the model is applied to a new set of data (out-of-sample testing) rather than the dataset used in its estimation (in-sample testing). The model's out-of-sample overall predictive performance is very similar to that for the in-sample analysis. The largest variation in performance is at the upper 10% cut-off value where it correctly predicts 85.75% of all outcomes compared to 89.07% for the in-sample analysis. At the other reported cut-off values the accuracy of the in-sample results are all higher than for the out-of-sample results; however, the difference only ranges from 0.4% to 2.01%. The exception to this is the approval concentration ratio at the highest 20% cut-off value where the out-of-sample ratio of 14.14% exceeds the 11.92% of the in-sample analysis. In line with my hypothesis the model does indeed possess an ability to correctly predict the likelihood that a patent will be linked to an approved drug with a performance level that is better than chance alone.

When the same out-of-sample analysis is also applied to Model 8 (including forward citations), the model's predictive ability is almost identical to that for the model excluding these citations. As can be observed in Table 6.7, at the upper 10% (20%) cut-off values the model still accurately predicts 90.95% (81.23%) of patents with no link to an approved drug and 21.43% (34.57%) of patents linked to approved drugs. The overall accuracy is 85.53% at the highest 10% and 77.59% at the highest 20% cut-off values. Approval concentrations are 16.70% at the highest 10% and 13.49% at the highest 20% cut-off values. At the lowest 10% and 20% levels, the accuracy of the predictions for patents without a link to an approved drug are within 1% of those based on the model excluding forward citations. In comparison to the in-sample analysis of Model 8, the overall accuracy of the out-of-sample predictions are 2.6% lower at the highest 10% and 20% cut-off values. At the highest 10% cut-off value the approval concentration ratio is 3.3% lower, but at the highest 20% level it is a little higher by 0.8%.

These out-of-sample results confirm that both Models 5 and 8 can accurately predict both approval-linked patents and those patents with no link to an approved drug, with a predictive ability that is better than would be expected by chance. The models not only possess explanatory power but also predictive power.

Table 6.7 Model 8-Out-of-Sample Analysis of Model Predictive Ability

Actual	**Prediction (With Forward Citations Included)**		
	Non-Approval	Approval	Predicted Values: Cut-off
Non-Approval	3,760	374	0.032626 (Upper 10%)
	90.95%†	9.05%	
Approval	275	75	
	78.57%	21.43%+	
Approval Concentration Ratio 16.70%		Overall Accuracy 85.53%	
Non-Approval	3,358	776	0.0137609 (Upper 20%)
	81.23%†	18.77%	
Approval	229	121	
	65.43%	34.57%+	
Approval Concentration Ratio 13.49%		Overall Accuracy 77.59%	
Non-Approval	2,094	2,040	0.003264 (50%)
	50.65%†	49.35%	
Approval	147	203	
	42%	58%+	
Approval Concentration Ratio 9.05%		Overall Accuracy 51.23%	
Non-Approval	3,037	827	0.0007615 (Lowest 20%)
	80.00%†	20.00%	
Approval	280	70	
	80.00%	20.00%+	
Approval Concentration Ratio 7.80%		Overall Accuracy 69.31%	
Non-Approval	3,713	421	0.0003352 (Lowest 10%)
	89.82%†	10.18%	
Approval	322	28	
	92.00%	8.00%+	
Approval Concentration Ratio 6.24%		Overall Accuracy 83.43%	
† Correctly Predicted True Negative (Specificity), + Correctly Predicted True Positive (Sensitivity)			
Patents Granted up to 1996 Used to Predict Outcomes for Patents Granted from 1997 Onwards. Sample 1997 Onwards: 4,484 Patents (4,134 Non-Approvals and 350 Approvals)			

Notes: This table displays the patents that are classified as drug approval-linked patents or patents with no link to an approved drug. These classifications are based on cut-off thresholds selected from the predicted probabilities of the conditional logistic regression for Model 8. The classifications displayed are based on an out-of-sample analysis where Model 8 is estimated using patents with grant years up to and including 1996. The test sub-sample includes patents with grant years of 1997 onwards.

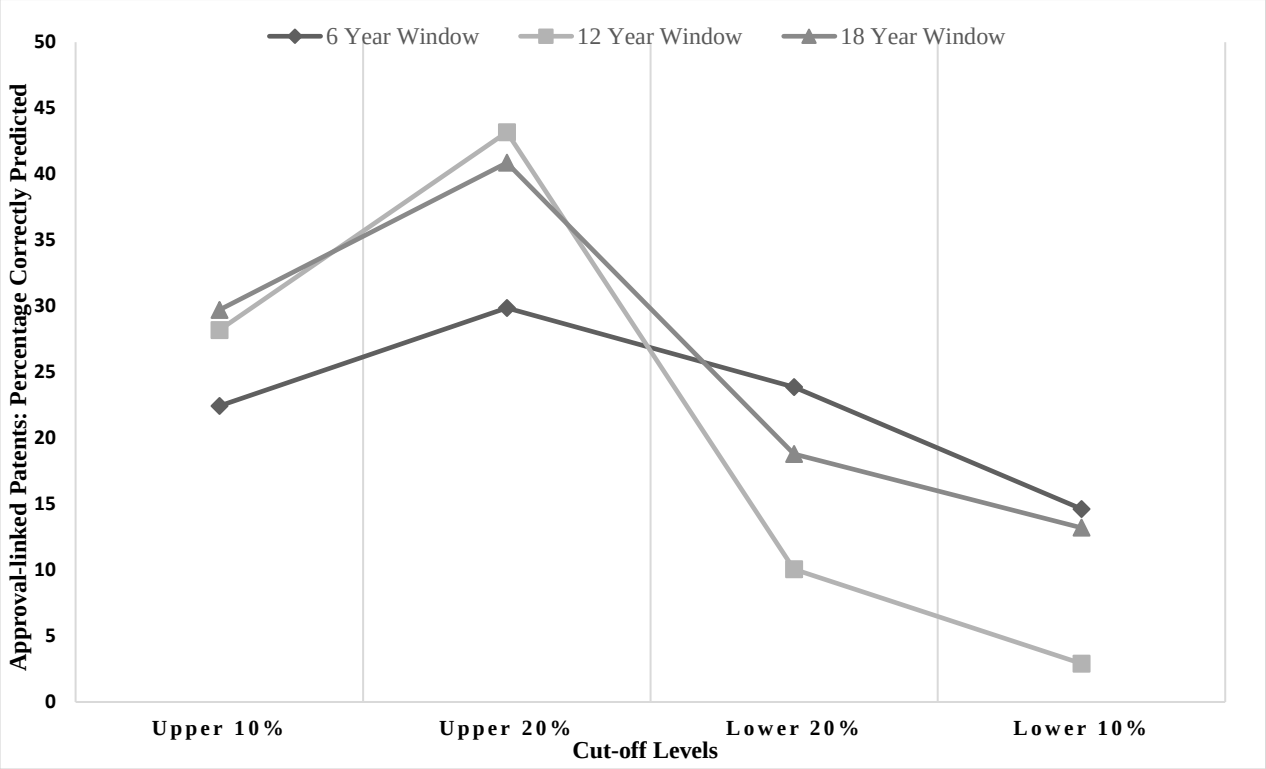
6.4.1 Model Predictive Ability: Testing the Out-of-Sample Analysis with Respect to Time

Next, I test the out-of-sample analysis to observe if it is robust to the selection of alternative time periods for the model estimation sub-sample and the model test sub-sample. In Section 6.4, I employed a thirty-four-year estimation sub-sample (patent grant years 1962 to 1996) and a seventeen-year test sub-sample (patent grant years 1997 onwards). To validate my results, I test a range of estimation and test samples. I introduce a shorter six-year estimation sample (patents granted during the period 1976-81) and then expand this estimation window, from the same 1976 start point, to twelve years (patents granted during the period 1976-87) and then to eighteen years (patents granted during the period 1976-93). The six-year sample contains 133 approval-linked patents (5.28% of the 2,521 patents in this sample), the twelve-year sample contains 185 approval-linked patents (4.05% of the 4,568 patents in this sample) and the eighteen-year sample contains 238 approval-linked patents (3.41% of the 6,985 patents in this sample). For each of these samples I estimate Models 5 and 8. I use the predicted probabilities to predict patent outcomes for the patents granted after the end of certain, selected time periods. That is, those patents granted from 1982 onwards (10,371 patents) for the six-year estimation sample, those patents granted from 1988 onwards (8,324 patents) for the twelve-year estimation sample and those patents granted from 1994 onwards (5,513 patents) for the eighteen-year estimation sample. As for Tables 6.5 and 6.6, the results indicate that patents linked to approved drugs are also predicted with similar or increasing accuracy rates.

As displayed in Figure 6-1, using the six-year estimation window the model, without forward citations, predicts 22.44% of approval-linked patents at the upper 10% level and 29.86% at the upper 20% level. The addition of forward citations to the model exhibited similar predictive ability, with slightly lower levels of accuracy. When the estimation sample size is increased from six years to twelve and then to eighteen years predictive accuracy at these upper cut-off levels continues to improve as the size of the estimation sample increases and size of the test sample decreases. As expected, increasing the volume of data used in the estimation of the model enhances its predictive ability. With the shorter six-year estimation sample, Model 5 can correctly predict 87.20% of all outcomes with an approval concentration ratio of 10.12% at the upper 10% cut-off level, but with the eighteen-year estimation sample 87.68% of all outcomes are correctly predicted with an approval concentration ratio of 21.69% at the upper 10% cut-off level. The predictive ability of the model is influenced by the size, in years, of the estimation samples with all three estimation

samples producing an overall accuracy level of between 87% and 88%. However, longer estimation sample periods produce increasing approval concentration ratios, with the twelve-year estimation sample having a ratio of 15.49%, increasing to 21.69% for eighteen-year estimation sample at the upper 10% cut-off levels.

Figure 6-1 Prediction of Patents Linked to Approved Drugs Using 6, 12 and 18 Year Estimation Samples (Model 5)



As another robustness check, I also move the starting point of the six-year estimation sample to test if the patent composition of the estimation sample influences the model's predictive ability. I estimate the model using a sample of patents with grant years of 1976 to 1981 (2,521 patents), patents with grant years of 1988 to 1993 (2,417 patents) and patents with grant years of 1994 to 1999 (3,563 patents). The number of patents in each of the estimation samples are broadly similar for the first two samples and over 40% larger for the sample starting in 1994, reflecting trends in patenting levels. The predicted probabilities from these models are then employed to predict patent outcomes for the remainder of the sample period; that is, patents with grant years of 1982 onwards, 1994 onwards and 2000 onwards. The results displayed in Figure 6-2 show that varying the sample start point influences the predictive ability of the model. The models estimated using samples with later time periods have an increasing ability to predict approval-linked patents at the 10% cut-off but even more so at the 20% cut-off values. In all three cases, the highest accuracy rates for predicting

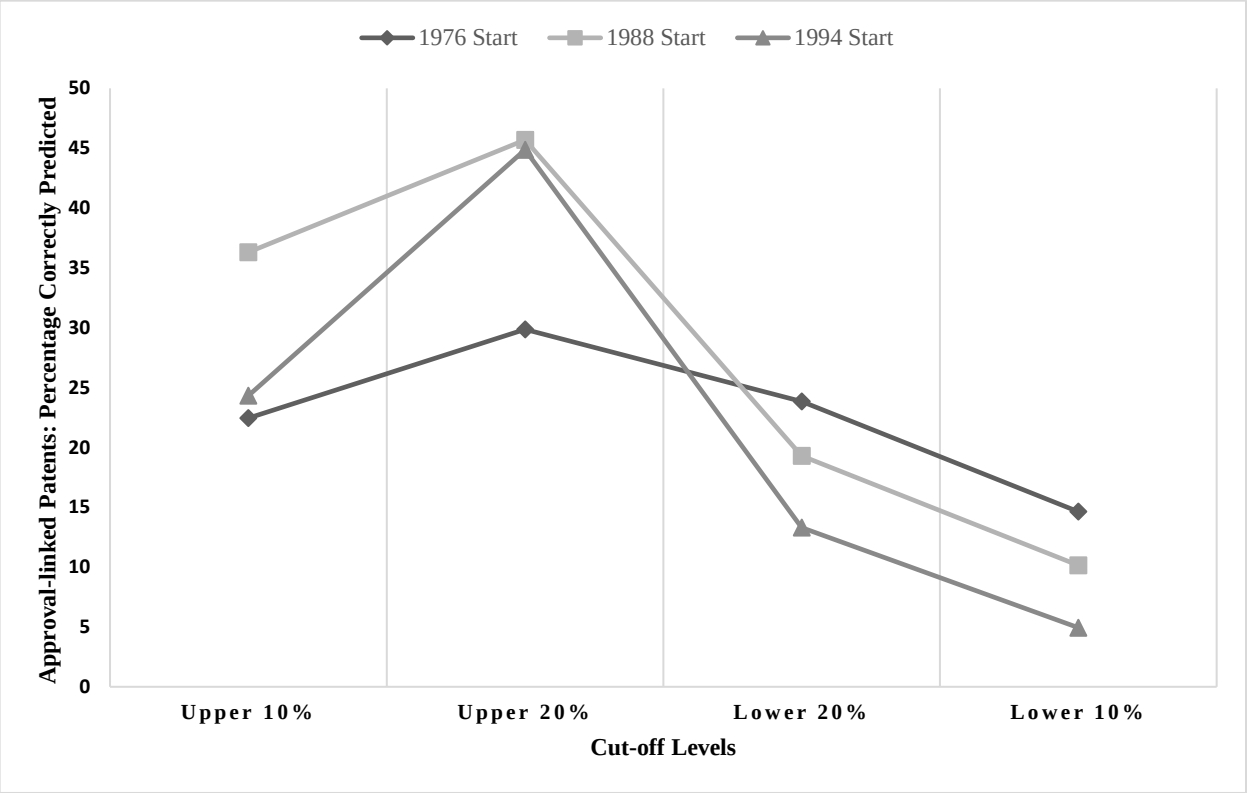
approval-linked patents are displayed at the upper 20% cut-off level, followed by a downward trend at the lower 20% and 10% cut-off levels. However, the overall accuracy of the model predictions is higher at the 10% cut-off level, with the predictions for 1982 onwards having an overall accuracy of 87.20%, 88.17% for 1994 onwards and 83.79% for 2000 onwards. Approval concentration ratios are also higher at this 10% level, with a ratio of 10.12% for predictions for 1982 onwards, 24.20% for 1994 onwards and 27.12% for 2000 onwards.

For the same length of estimation sample period, six years, the magnitude of the approval prediction outcomes measure does vary with the start point chosen but its trend follows a similar pattern. This might arise as varying the estimation sample start point means that the resulting test samples to be predicted vary in size with both the number of years and the number of patent outcomes to be predicted. The estimation samples with later start points of 1988 and 1994, are applied to test samples that are smaller in size, both in terms of years and patents, than the estimation sample starting in 1976. It is interesting to note that the estimation sample with a 1994 start point performs best in terms of its approval concentration ratio, this might be a result of its estimation sample being smaller than the test sample it is predicting (3,563 patents compared to 2,344 patents). These results show that whilst time can play a role in influencing the level of accuracy of the model's predictive ability, in terms of approval concentration ratios, the level of the overall accuracy of the model is reasonably stable. It does appear that time does not alter the ability of the model to predict patents linked to approved drugs and non-approval linked patent outcomes.

As a further robustness test, I also check if the patent or firm composition of the estimation and test samples influences the out-of-sample analysis results. To test this, I randomly sort the entire data sample of 12,957 patents in Stata, using two-thirds of these data (8,638 patents) to estimate the model. Predicted values are calculated for the remaining one-third of the data (4,319 patents) that comprise the test sample. The estimation sample contains 400 approval-linked patents 4.63% of the 8,638-patent sample). At the 10% level, the predictive ability of the model predicts nearly 40% of these patents; at the 20% level, the predictive performance of the model excluding forward citations exceeds 47% and exceeds 49% with the inclusion of these citations. The model excluding forward citations has an overall accuracy of 88.86% and an approval concentration ratio of 21.25%. This is the highest level of overall model predictive accuracy for any of the out-of-sample analyses conducted and is

very similar in magnitude to the in-sample results presented in Tables 6.2 and 6.4. The approval concentration ratio is also similar to those for the in-sample analyses.

Figure 6-2 Prediction of Patents Linked to Approved Drugs Using 6 Year Estimation Sample with Varying Sample Start Years



Overall, these checks on the robustness of the out-of-sample analysis results provide strong support for the predictive ability measures reported in Tables 6.5 and 6.6 using different firm sample compositions and alternative estimation and test samples. They provide evidence that Models 5 and 8, excluding and including forward citations, can consistently predict the likelihood that a patent will be linked to an approved drug with a predictive ability that cannot be attributed solely to chance. In addition, these tests also demonstrate that the models possess the ability to predict patents that are not linked to approved drugs.

6.5 Conclusion

The main motivation of this study is to examine whether certain firm patent and publishing characteristics can be utilised to assist in the prediction of the likelihood that a patent will be linked to the development of an approved drug. This chapter presented empirical results for the in-sample and out-of-sample analysis of the predictive ability of these firm patent and publishing characteristics. It also presented several robustness checks, including differing firm and patent sample compositions for the in-sample analysis. There were also tests of

results across differing size and time measures for the estimation and test samples utilised in the out-of-sample analysis.

Firstly, the results from the in-sample analysis section support the earlier conclusion that the models developed in Chapter 5 can indeed be employed to predict the likelihood that a patent will be involved in the development of an approved drug. These findings were robust to both changes in the firm and patent composition of the sample. Similarly, the results from the out-of-sample analysis of the model's predictive performance also support the view that the hypothesised relationship between the likelihood of approval and firm patent and publishing characteristics can predict the future approval status of a patent with an ability that cannot be solely attributed to chance. Again, the results were robust to changes in the size and timing of the selected estimation and test samples. Overall, these results suggest that for the U.S. pharmaceutical industry, a firm's patent and publishing characteristics can be employed to assist in the prediction of a patent's eventual drug approval status.

In the next chapter (Chapter 7), the main aim is to test whether abnormal returns can be earned by investing in the stocks of firms holding patents predicted to be approval-linked by the models in this chapter.

Chapter 7 Investing in Model Prediction Outcomes

7.1 Introduction

In Chapter 6, a drug approval prediction model is developed and tested for its ability to correctly predict actual drug approvals. The results from Chapter 6 indicate that an approval concentration ratio of up to 20% could be achieved by holding a portfolio of the 10th percentile of patents with the highest approval likelihood over the period 1962 to 2014. This is a substantial improvement over a random selection approach, given that approval-linked patents comprise 4.88% of the total patent sample.

As discussed in Chapter 2, Section 2.7, there is a consensus within the literature that receipt of an FDA drug approval significantly increases shareholder wealth. Hamill et al. (2013) found that significant abnormal returns accrued to U.S.-listed firms. However, there are few studies that evaluate the effect of the grant of an individual patent on stock returns using an event study methodology (see for example Austin, 1993; Liu, 2006 and Plumlee, Xie, Yan and Yu, 2015). Both Austin and Liu found that biotechnology patents earn excess returns, with Austin finding that these excess returns are larger for patents readily identifiable with end products. This chapter also investigates whether investing in firm stocks as they are granted patents can form the basis of a profitable investment strategy. However, my approach differs from this prior literature as it builds on the predictive results discussed in Chapter 6. I utilise the predictive models from this chapter to formulate an investment portfolio that is comprised of the stocks of firms that are predicted to be the most likely to hold patents that will subsequently be linked to an approved drug. I then evaluate the portfolio to assess whether it can earn long run abnormal returns. This evaluation is important for investors, such as fund managers, who are looking for superior investment strategies.

The methodology employed in this chapter (including the computation of returns and the formation of portfolios) is discussed in Chapter 4, Section 4.7. In Section 7.2, the market-adjusted and risk-adjusted returns generated by the model are presented and discussed. Section 7.3 presents robustness checks of the key results obtained in the chapter. Section 7.4 short run patent announcement returns are investigated. Section 7.5 concludes the chapter and discusses some of its implications to research and practice.

7.2 Excess Returns from Model Predictions

In this chapter, the issue of primary importance is whether abnormal stock returns can be earned by investing in stocks of the firms holding patents predicted to be approval-linked by the models in Chapter 6. Although the methodology discussed in Chapter 6 provides a statistical assessment of model performance, it has nothing to say about the economic usefulness of the model. It can be observed that Models 5 and 8 produce forecasts considerably better than chance. The model including the forward citation variable (Model 8) outperforms the model without their inclusion but only by 10%. If the model with forward citations were used to select portfolio stocks, investors would have to wait an additional 5 years to gather this data. This delay in portfolio formation is avoided by using the patents predicted to be approval-linked from prediction Model 5 to create an investment portfolio.

The investment strategy consists of adopting a buy-and-hold approach for the stocks within the portfolios. The portfolio stocks are selected based on the predictions included within the upper 10% cut-off grouping of both the in-sample and out-of-sample (1997 to 2014) analyses (see Tables 6.2 and 6.6). The results from the analyses indicate that an approval concentration ratio of up to 19.75% (in-sample) and 17.82% (out-of-sample) could be achieved by holding a portfolio comprised of firms predicted to hold approval-linked patents.

The methodology consists of simulating buying the stock on the publication day of the patent and, for correctly predicted approval-linked patents, selling it on the day the drug to which it is linked is approved by the FDA. For patents that the model incorrectly predicts as approval-linked (patents that do not actually possess a link to an approved drug), stocks are sold at a pseudo approval date, 1,078 calendar days later (approximately 750 trading days). This holding period is utilised as it is the median publication-to-approval date period (for patents that actually possess a link to an approved drug) as discussed in Chapter 4, Section 4.7. The patent publication date is treated as the event date rather than the application date as patent applications (delayed by 18 months) have only been published and made publicly available since November 2000.

The stock returns from the portfolio of firms predicted to hold approval-linked patents are calculated for a holding period that is determined by this publication to approval date period or pseudo approval date in the case of incorrect predictions. I assume that an investment is made in each firm that is a constituent of the approval-prediction portfolio. These investments are made in equal capital proportions and are held for the holding period

relevant to each firm's patent. For return calculation purposes, I utilise returns for the firms holding predicted approval-linked patents, which are adjusted for dividends and capitalisation changes.

Stocks of firms whose patents are predicted as potential approval-linked patents, based on their characteristics and firm-level publishing behaviour, are placed within a portfolio and the discrete daily returns on each stock within the portfolio are computed using an event study methodology. Equal-weighted portfolio returns are computed from the discrete daily returns of all the constituent firms within the portfolio. This equal-weighted portfolio approach assumes that an equal amount is invested in each firm in the portfolio. In this study, the equal-weighted portfolio daily returns are the arithmetic average of the returns to all the stocks within the portfolio for that day.

Non-risk and risk-adjusted portfolios (capital asset pricing models (CAPM)) are used to evaluate returns. These different methodologies are fully discussed in Section 4.7. I focus on the portfolios that were developed in Chapter 6, Tables 6.2 and 6.6. The returns generated by these portfolios are presented and discussed in the sections below.

7.2.1 Average Daily Returns

Fama and French (1998) favour value-weighted returns as they more accurately capture the total wealth effects experienced by investors. However, Loughran and Ritter (2000) showed that value-weighted returns tend to underestimate abnormal returns when the event being studied is a managerial choice variable. Choosing whether to patent the outcome of a firm's research project and the timing of such a patent application is the choice of its management team. To reflect these differing opinions, I firstly calculate returns for an equal-weighted portfolio and then confirm my findings by calculating value-weighted portfolio returns.

7.2.1.1 Market-Adjusted Returns

In this section, the simple market-adjusted daily returns generated by the model are calculated for each day of the holding period. The equation for abnormal return calculation (Equation 7.1) is shown below;

$$AR_{it} = R_{it} - R_{mt} \qquad \text{(Equation 7.1)}$$

In this equation, AR_{it} is the daily abnormal (or excess) return and is defined as a firm's daily return, $R_{it,}$ minus the value-weighted daily market return, R_{mt}[75]. The market return is the expected return for the same period. I calculate the abnormal return for each observation in the holding period (the event window) and the returns for all the portfolio firms treated as a group. In addition, I calculate the cumulative abnormal return (CAR) for all the portfolio firms treated as a group. The equation for CAR calculation (Equation 7.2) is shown below;

$$CAR_{it} = \sum_{t=1}^{\tau} AR_{it} \qquad \text{(Equation 7.2)}$$

Table 7.1 presents the results obtained when Model 5 is used to predict potential approval-linked patents during the period 1962 to 2014 (in-sample predictions). The portfolio of stocks is selected to include all firms that hold patents that are predicted approval-linked and fall within decile 10 of the prediction results displayed in Table 6.2. Decile 10 is chosen as all the patents in this portfolio are predicted to be approval-linked and it also has the highest proportion of correctly predicted approval-linked patents. Returns are calculated using a simple Market Adjusted Returns (MAR) model (see Equations 7.1 and 7.2). The abnormal return is the daily return on the drug approval (or pseudo approval) date, whilst the cumulative return is the sum of the daily returns for the holding period. The calculation of the approval concentration ratio is detailed in Section 4.6.2.1 of Chapter 4.

The results obtained using this portfolio construction strategy show that the portfolio (equal-weighted) generates an average daily abnormal return of 0.003 (0.3%) on the drug approval (or pseudo approval) date. Average cumulative abnormal returns of 0.6111 (61.1%) are earned over the holding period (patent publication date to approval/pseudo approval date) as a whole. The abnormal and cumulative abnormal returns are both significant at the 1% level. As a robustness test the Datastream Pharmaceutical and Biotechnology Industry index and Datastream Pharmaceutical Industry index were also used in place of Kenneth French's Drug Industry index as alternative measures of the market return. Approval date abnormal returns were also 0.3% for both Datastream indices. Cumulative abnormal returns were measured at 0.591 (59.1%) for the Pharmaceutical and Biotechnology Industry index and 0.643 (64.3%)

[75] This is the drug industry value-weighted return listed within the 48-industry portfolio in the Kenneth French Data Library.

for the Pharmaceutical Industry index (all returns are significant at the 1% level). These results indicate that the abnormal returns are not sensitive to the market return measure.

Table 7.1 Decile 10 Approval Date Portfolio Returns for the In-Sample Period 1962 to 2014

Equal-Weighted Decile 10 Portfolio: In Sample, 1962 to 2014		
	Approval Concentration Ratio (%)	**MAR Model**
AR_{it} at Drug Approval Date	19.75	0.003 (0.006)***
CAR_{it} at Drug Approval Date	19.75	0.611 (0.000)***

Notes: The table presents results for the univariate regressions of abnormal and cumulative abnormal returns on the drug approval or pseudo approval date. The equations for abnormal and cumulative abnormal return calculation (Equations 7.1 and 7.2) are shown below;

$AR_{it} = R_{it} - R_{mt}$ (Equation 7.1) and $CAR_{it} = \sum_{t=1}^{\tau} AR_{it}$ (Equation 7.2)

R_{it} is the day t simple return on a sample firm; R_{mt} is the day t simple market return and this is the drug industry value-weighted return listed within the 48-industry portfolio in the Kenneth French Data Library. The decile 10 portfolio contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. The approval concentration ratio indicates the percentage of correctly predicted approval-linked patents within the portfolio. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug, the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7.2 presents the results obtained when Model 5 is used to predict potential approval-linked patents from 1997 to 2014 using patents granted up to the end of 1996 (out-of-sample predictions). Again, a decile-based portfolio selection strategy is used to select the approval-linked patent portfolio.

The results obtained show that the portfolio generates average daily abnormal returns of 0.004 (0.4%) on the drug approval (or pseudo approval) date, although the returns are statistically insignificant. Average cumulative abnormal returns of 0.938 (93.8%) are earned over the holding period (patent publication date to approval/pseudo approval date) and are significant at the 1% level. Again, as a robustness test, the Datastream Pharmaceutical and Biotechnology Industry index and Datastream Pharmaceutical Industry index were again used in place of Kenneth French's Drug Industry index as alternative measures of the market return. Abnormal returns were also statistically insignificant and identical in value at 0.4% for both Datastream indices. Cumulative abnormal returns were measured at 0.901 (90.1%) for the Pharmaceutical and Biotechnology Industry index and 0.984 (98.4%) for the Pharmaceutical Industry index (both cumulative return measures are significant at the 1% level).

Table 7.2 Decile 10 Approval Date Portfolio Returns for the Out-of-Sample Period 1997 to 2014

Equal-Weighted Decile 10 Portfolio: Out-of-Sample, 1997 to 2014		
	Approval Concentration Ratio (%)	**MAR Model**
AR_{it} at Drug Approval Date	17.82	0.004 (0.137)
CAR_{it} at Drug Approval Date	17.82	0.938 (0.000)***

Notes: The table presents results for the univariate regressions of abnormal and cumulative abnormal returns on the drug approval or pseudo approval date. The equations for abnormal and cumulative abnormal return calculation (Equations 7.1 and 7.2) are shown below;

$AR_{it} = R_{it} - R_{mt}$ (Equation 7.1) and $CAR_{it} = \sum_{t=1}^{\tau} AR_{it}$ (Equation 7.2)

R_{it} is the day t simple return on a sample firm, R_{mt} is the day t simple market return and this is the drug industry value-weighted return listed within the 48-industry portfolio in the Kenneth French Data Library. The decile 10 contains 449 patents, 80 correctly predicted to be approval-linked and 369 patents incorrectly predicted to be approval-linked. The approval concentration ratio indicates the percentage of correctly predicted approval-linked patents within the portfolio. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are shown in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

7.2.1.2 Risk-Adjusted Returns

When computing portfolio abnormal returns, it can also be informative to observe the results when portfolio returns are adjusted for risk. This section discusses the model used in the computation of abnormal returns (or risk-adjusted returns) which are obtained by adjusting portfolio returns for risk factors such as portfolio risk. Several approaches (with different strengths and weaknesses) have been employed in previous literature in an attempt to adjust returns for the risks involved. To allow for robustness and comparison of the results with previous literature, portfolio daily returns are adjusted for risk by using Capital Asset Pricing Model (CAPM). The equation for CAPM (Equation 7.3) is shown below;

$$R_{it} - R_{ft} = \alpha_i + \beta_{mkt} (R_{mt} - R_{ft}) + \alpha_{it} \quad \text{(Equation 7.3)}$$

R_{it} is the discrete return on the portfolio i, on day t, R_{ft} is the risk-free rate on day t, α_i is the abnormal daily return, β_{mkt} is the risk measure and R_{mt} is the market return on day t. As per this equation, I fit daily excess portfolio returns ($R_{it} - R_{ft}$) to excess market return ($R_{mt} - R_{ft}$). The intercept or constant term from this regression provides an estimate of returns that cannot be explained by common risk factors.

Table 7.3 Decile 10 Portfolio Approval Date Returns for the In-Sample Period 1962 to 2014: Risk-Adjusted Returns (CAPM)

	CAPM
Alpha (α_i)	0.030 (0.000)***
Excess Market Return ($R_{mt} - R_{ft}$)	1.085 (0.000)***

Notes: The table presents results for the univariate regressions of abnormal and cumulative abnormal returns on the drug approval or pseudo approval date. The equations for abnormal and cumulative abnormal return calculation (Equations 7.3) is shown below;

$R_{it} - R_{ft} = \alpha_i + \beta_{mkt} (R_{mt} - R_{ft}) + \alpha_{it}$ (Equation 7.3)

R_{it} is the day t simple return on a sample firm; R_{ft} is the risk free rate on day t, α_i is the abnormal daily return, β_{mkt} is the risk measure and R_{mt} is the day t simple market return and this is the drug industry value-weighted return listed within the 48-industry portfolio in the Kenneth French Data Library. The decile 10 contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. The estimate of the intercept term, Alpha, tests the null hypothesis that the mean daily excess return on the portfolio is zero. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are shown in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

As shown in Table 7.3, I find that the portfolio generates average daily abnormal returns of 0.03% (equal-weighted portfolio) on the drug approval (or pseudo approval) date during the holding period. The alpha generated by the equal-weighted portfolio is significant at the 1% level. The excess market return is also positive and significant at the 1% level.

7.2.1.3 Buy-and-Hold Long Run Returns

The CAR results are encouraging and appear to support the premise that the investment portfolios created from the 1962 to 2014 (in-sample) and 1997 to 2014 (out-of-sample) prediction results of Model 5 can earn abnormal returns over the holding period. However, the CAR approach assumes that an investor buys a stock and sells it almost immediately. Using CARs to measure long run returns is therefore inappropriate, as I assume an investor buys a stock, not selling it immediately, but holding it for the period running from the patent publication date to the related approval (or pseudo approval) date. I therefore use the more appropriate measure of the buy-and-hold abnormal return (BHAR) in tests designed to detect long-run abnormal stock returns.

Initially, I use the simple market-adjusted returns model to estimate a portfolio firm's BHAR. The BHAR for an individual stock is the difference between the buy-and-hold return of the portfolio firm and that of the benchmark market return, assuming that an investor buys a stock and holds it until the end of the holding period.

The market-adjusted BHAR at the drug approval date is defined as:

$$BHAR_{it} = \prod_{t=1}^{\tau}[1 + R_{it}] - \prod_{t=1}^{\tau}[1 + R_{mt}] \quad \text{(Equation7.4)}$$

Where R_{it} denotes firm's daily stock return starting from the day a patent is published until the approval date of the drug to which it is related, or for patents without a link to an approved drug, the pseudo approval date. The patent-approved drug link is established using multiple sources including the FDA Orange Book and is discussed in detail in Chapter 4, Section 4.2. R_{mt} denotes the daily value-weighted market return over the same period[76].

Table 7.4 presents the BHAR results obtained when Model 5 is used to predict potential approval-linked patents from 1962 to 2014 (in-sample predictions) and two differing out-of-sample prediction periods, 1962 to 1996 and 1997 to 2014. Table 7.5 presents these returns winsorised at the top and bottom 1%, to exclude any potential influence from extreme return values. In both return calculations, a decile optimal portfolio selection strategy is used to select the approval-linked patent portfolio.

As can be observed from the results in Table 7.4, for all three sample periods BHARs are positive and increasing in the period approaching the approval date. There is no marked increase in returns on the approval date. This is not unexpected, as 80-90% of the patents in these three portfolios are not linked to an approved drug. There would therefore not be any approval day reaction by the market. In all but one instance, they are also all statistically significant. Although the in-sample prediction portfolio has a higher approval concentration ratio (19.75%) than the out-of-sample prediction portfolios, 1962 to 1996 (6.15%) and 1997 to 2014 (17.82%), it is the 1997 to 2014 out-of-sample prediction portfolio that generates higher levels of abnormal returns over the holding period. The 1962 to 1996 portfolio exhibits positive significant abnormal returns but they are at a lower level than those produced by both the in-sample and other out-of-sample portfolios. These lower returns may be attributable to the low proportion of correctly predicted approval-linked patents within this portfolio.

[76] This is the drug industry value-weighted return listed within the 48-industry portfolio in the Kenneth French Data Library.

Table 7.4 Decile 10 Portfolio Buy-and-Hold Returns for Patents Predicted to be Approval-Linked for the In-Sample Period 1962 to 2014 and Out-of-Sample Periods 1962 to 1996 and 1997 to 2014

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1962 to 2014) Buy-and-Hold Returns (BHAR)	MAR Model (1962 to 1996) Buy-and-Hold Returns (BHAR)	MAR Model (1997 to 2014) Buy-and-Hold Returns (BHAR)
-700	0.141 (0.092)*	0.016 (0.212)	0.077 (0.072)*
-650	0.124 (0.072)*	0.025 (0.076*	0.134 (0.026)**
-600	0.158 (0.039)**	0.047 (0.009)***	0.131 (0.032)**
-550	0.209 (0.012)**	0.062 (0.012)**	0.226 (0.07)***
-500	0.251 (0.001)***	0.073 (0.012)**	0.363 (0.001)***
-450	0.291 (0.000)***	0.111 (0.001)***	0.384 (0.000)***
-400	0.350 (0.000)***	0.115 (0.001)***	0.450 (0.000)***
-350	0.463 (0.000)***	0.141 (0.001)***	0.670 (0.000)***
-300	0.473 (0.000)***	0.116 (0.003)***	0.717 (0.000)***
-250	0.516 (0.000)***	0.132 (0.003)***	0.837 (0.000)***
-200	0.545 (0.000)***	0.146 (0.002)***	0.895 (0.000)***
-150	0.519 (0.000)***	0.164 (0.001)***	0.848 (0.000)***
-100	0.527 (0.000)***	0.169 (0.002)***	0.874 (0.000)***
-50	0.559 (0.000)***	0.205 (0.001)***	0.900 (0.000)***
0	0.604 (0.000)***	0.270 (0.000)***	0.904 (0.000)***

Table 7.5 Decile 10 Portfolio Buy-and-Hold Returns (Winsorised) for Patents Predicted to be Approval-Linked for the In-Sample Period 1962 to 2014 and Out-of-Sample Periods 1962 to 1996 and 1997 to 2014

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1962 to 2014) Winsorised Buy-and-Hold Returns (BHAR)	MAR Model (1962 to 1996) Winsorised Buy-and-Hold Returns (BHAR)	MAR Model (1997 to 2014) Winsorised Buy-and-Hold Returns (BHAR)
-700	0.050 (0.027)**	0.020 (0.082)*	0.077 (0.071)*
-650	0.056 (0.022)**	0.027 (0.039)**	0.134 (0.026)**
-600	0.083 (0.002)**	0.049 (0.004)***	0.129 (0.031)**
-550	0.126 (0.000)***	0.055 (0.004)***	0.213 (0.006)***
-500	0.180 (0.000)***	0.062 (0.006)***	0.329 (0.001)***
-450	0.219 (0.000)***	0.089 (0.000)***	0.369 (0.000)***
-400	0.257 (0.000)***	0.093 (0.001)***	0.420 (0.000)***
-350	0.349 (0.000)***	0.104 (0.000)***	0.595 (0.000)***
-300	0.355 (0.000)***	0.100 (0.003)***	0.630 (0.000)***
-250	0.371 (0.000)***	0.106 (0.004)***	0.716 (0.000)***
-200	0.412 (0.000)***	0.121 (0.003)***	0.768 (0.000)***
-150	0.427 (0.000)***	0.126 (0.004)***	0.759 (0.000)***
-100	0.410 (0.000)***	0.128 (0.005)***	0.744 (0.000)***
-50	0.429 (0.000)***	0.150 (0.002)***	0.779 (0.000)***
0	0.432 (0.000)***	0.174 (0.001)***	0.776 (0.000)***

Notes: Tables 7.4 and 7.5 present buy-and-hold abnormal returns for the period 700 trading days prior to the drug approval or pseudo approval date for equal-weighted decile 10 portfolios. The equations for buy-and-hold abnormal returns calculation (Equations 7.4) is shown below;

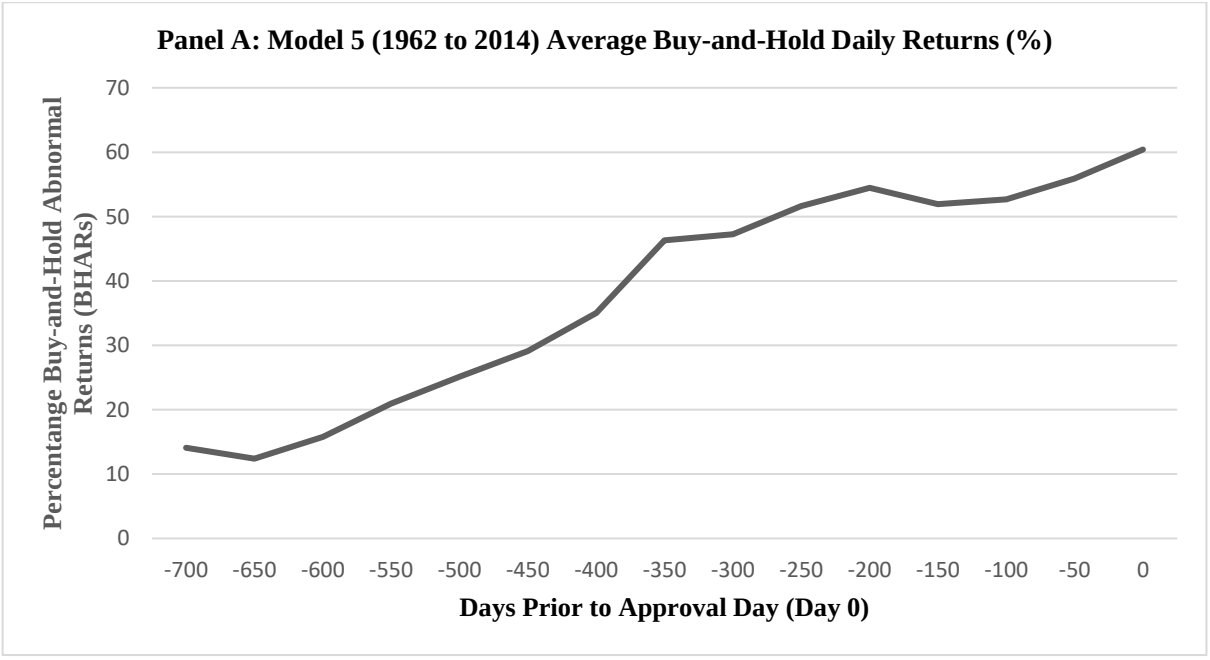
$BHAR_{it} = \prod_{t=1}^{\tau}[1 + R_{it}] - \prod_{t=1}^{\tau}[1 + R_{mt}]$ (Equation 7.4)

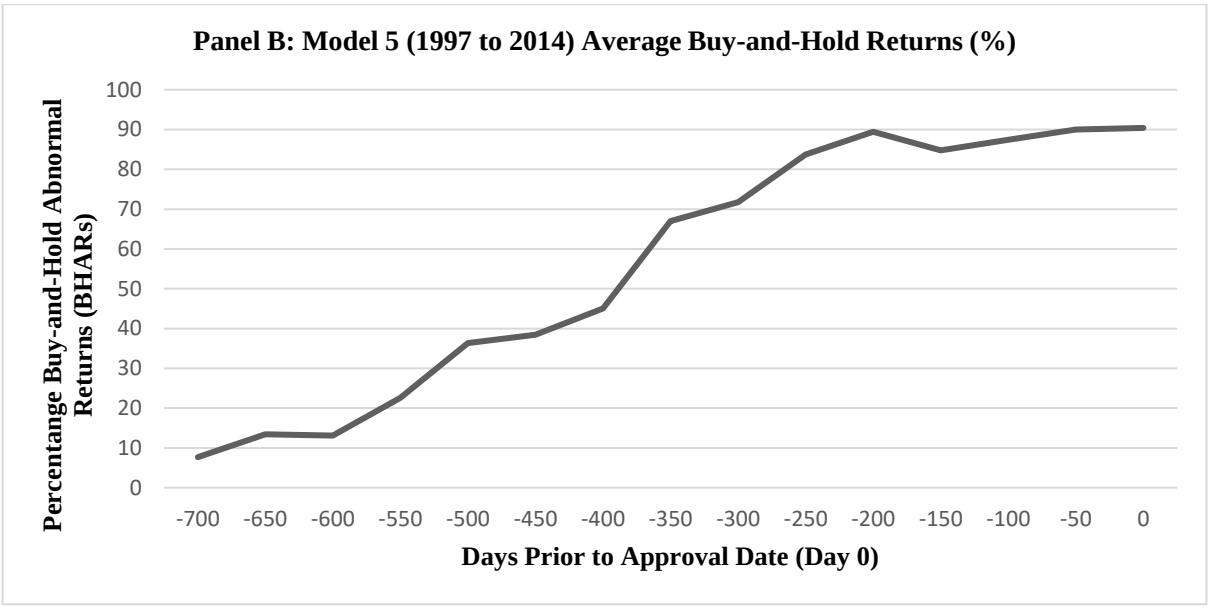
R_{it} is the day t simple return on a sample firm, R_{mt} is the day t simple market return and this is the drug industry value-weighted return listed within the 48-industry portfolio in the Kenneth French Data Library. The returns

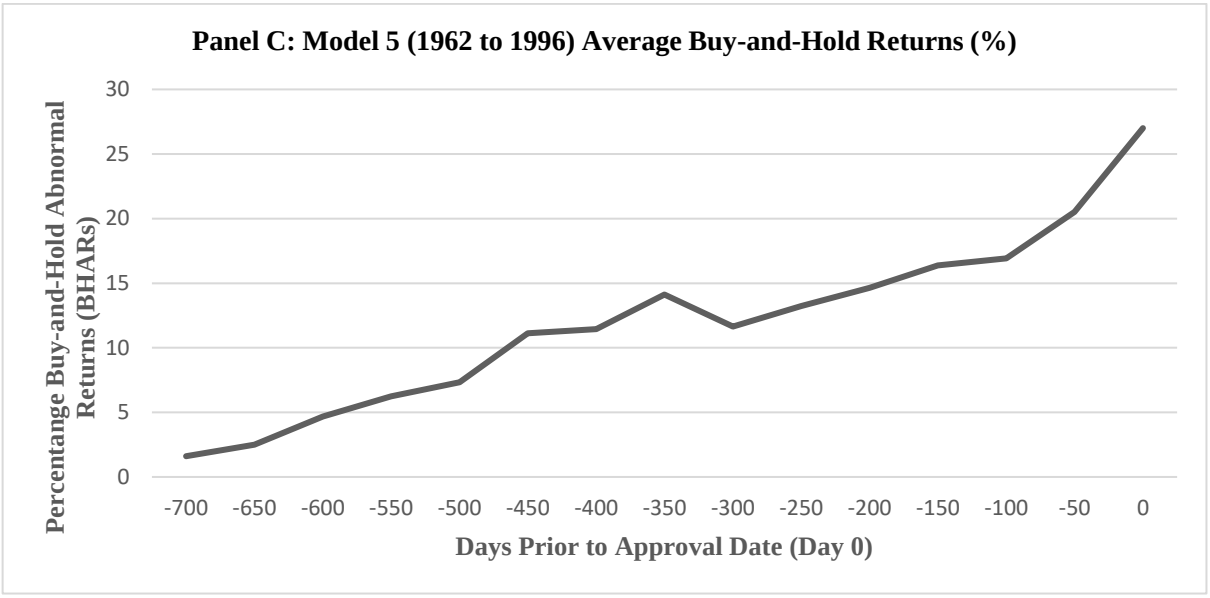
displayed are for the period from 700 trading days prior to the approval date. For the period 1962 to 2014, decile 10 contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. For the period 1962 to 1996, decile 10 contains 846 patents, 52 correctly predicted to be approval-linked and 794 patents incorrectly predicted to be approval-linked. For the period 1997 to 2014, decile 10 contains 449 patents, 80 correctly predicted to be approval-linked and 369 patents incorrectly predicted to be approval-linked. In Table 7.5, the BHARs have been winsorised at the top and bottom 1% to manage any potential influence from extreme return values. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are shown within parentheses, *** p<0.01, ** p<0.05, * p<0.1.

These patterns are clearly exhibited in Figure 7-1, Panels A, B and C. It appears that a higher approval concentration ratio is not a prerequisite for attaining higher BHARs, but a ratio markedly higher than that of chance (4.88%) does improve portfolio return performance. In Table 7.5, it can be observed that winsorising portfolio returns does reduce BHAR values, but portfolio BHARs are still positive, statistically significant and increasing during the holding period. Extreme return values appear to have the strongest influence on the 1962 to 1996 out-of-sample portfolio, with their exclusion reducing BHAR values by 28.5%. Extreme return values have the least influence on the 1997 to 2014 out-of-sample portfolio with a 14.2% reduction. Extreme return values do influence overall BHAR values, but they are not the sole drivers of the positive and statistically significant portfolio returns that are observed during the holding period for all three portfolios.

Figure 7-1 Average Decile 10 Portfolio Buy-and-Hold Daily Returns for the In-Sample Period 1962 to 2014 and Out-of-Sample Periods 1997 to 2014 and 1962 to 1996







Notes: Panels A, B and C display the growth in the average percentage buy-and-hold returns for the period from 700 trading days prior to the drug approval date for equal-weighted decile 10 portfolios. In Panel A, the portfolio consists of stocks of the firms that hold the 1,296 patents predicted to have an approval linkage by Model 5 for the full in-sample period 1962 to 2014. In Panel B, the portfolio consists of stocks of the firms that hold the 449 patents predicted to have an approval linkage by Model 5 for the predicted out-of-sample period 1997 to 2014. In Panel C, the portfolio consists of stocks of the firms that hold the 846 patents predicted to have an approval linkage by Model 5 for the predicted out-of-sample period 1962 to 1996.

As shown in Table 7.6 I find that the difference in BHARs between the predicted approval-linked and the predicted not approval-linked portfolios is present prior to the approval (or pseudo approval) date and that throughout the 700-day period, both portfolios exhibit positive and significant returns. However, the approval-linked portfolio exhibits higher levels of BHARs for the entire 700-day period prior to the drug approval date. With the

difference increasing over time as the returns to the predicted approval portfolio increase on the approach to the approval (or pseudo approval) date. This result confirms that if an investor constructs an investment portfolio that comprises the stocks of firms holding patents predicted to have the highest likelihood of having a link to a subsequent drug approval, then such a portfolio provides a more profitable outcome for an investor.

Table 7.6 Comparison of Decile 10 Predicted Approval and Non-Approval Portfolio Buy-and-Hold Returns for the 1962 to 2014 Out-of-Sample Period

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1962 to 2014) Buy-and-Hold Returns (BHAR) Predicted Approval	MAR Model (1962 to 2014) Buy-and-Hold Returns (BHAR) Predicted Non-Approval	Difference in BHAR for Predicted Approvals and Predicted Non-Approvals
-700	0.141 (0.092)*	0.019 (0.000)***	0.122 (0.147)
-650	0.124 (0.072)*	0.028 (0.000)***	0.096 (0.162)
-600	0.158 (0.039)**	0.031 (0.000)***	0.127 (0.099)*
-550	0.209 (0.012)**	0.038 (0.000)***	0.171 (0.042)**
-500	0.251 (0.001)***	0.052 (0.000)***	0.199 (0.009)***
-450	0.291 (0.000)***	0.058 (0.000)***	0.233 (0.005)***
-400	0.350 (0.000)***	0.070 (0.000)***	0.280 (0.002)***
-350	0.463 (0.000)***	0.084 (0.000)***	0.379 (0.000)***
-300	0.473 (0.000)***	0.095 (0.000)***	0.378 (0.000)***
-250	0.516 (0.000)***	0.102 (0.000)***	0.414 (0.000)***
-200	0.545 (0.000)***	0.112 (0.000)***	0.433 (0.000)***
-150	0.519 (0.000)***	0.119 (0.000)***	0.400 (0.000)***
-100	0.527 (0.000)***	0.130 (0.000)***	0.397 (0.000)***
-50	0.559 (0.000)***	0.147 (0.000)***	0.412 (0.000)***
0	0.604 (0.000)***	0.157 (0.000)***	0.447 (0.000)***

Notes: For the period 1962 to 2014, decile 10 approval-linked portfolio contains stocks for the firms that hold the 1,296 patents, of which 256 patents are correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. The non-approval-linked portfolio contains 11,661 patents, of which 11,285 patents are correctly predicted not to be approval-linked and 376 are incorrectly predicted not to be approval-linked. BHARs are calculated in accordance with Equation 7.4 and the returns displayed are for the period from 700 trading days prior to the approval (or pseudo approval) date for equal-weighted decile 10 portfolios. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are shown within parentheses, *** p<0.01, ** p<0.05, * p<0.1.

In Table 7.7, it can be observed that for the 1997 to 2014 out-of-sample period both the approval-linked and non-approval-linked portfolios also exhibit positive returns throughout the 700-day period prior to the approval (or pseudo approval) date. These returns are significant for the approval-linked portfolio for the whole of this period. Whilst the non-approvals-linked portfolio exhibits higher levels of BHARs for the first 100 days, but these returns are not statistically significant. Again, the difference between the two portfolios is increasing over time as the returns to the approval-linked portfolio increase on the approach

to the drug approval date, with the returns to both portfolios being higher than for the in-sample portfolios from 550 days prior to the approval (or pseudo approval) date.

Table 7.7 Comparison of Decile 10 Predicted Approval and Non-Approval Portfolio Buy-and-Hold Returns for the 1997 to 2014 Out-of-Sample Period

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1997 to 2014) Buy-and-Hold Returns (BHAR) Predicted Approval	MAR Model (1997 to 2014) Buy-and-Hold Returns (BHAR) Predicted Non-Approval	Difference in BHAR for Predicted Approvals and Predicted Non-Approvals
-700	0.077 (0.072)*	0.216 (0.185)	-0.139 (0.408)
-650	0.134 (0.026)**	0.174 (0.121)	-0.040 (0.753)
-600	0.131 (0.032)**	0.173 (0.069)*	-0.042 (0.710)
-550	0.226 (0.007)***	0.212 (0.084)*	0.014 (0.927)
-500	0.363 (0.000)***	0.256 (0.074)*	0.107 (0.555)
-450	0.384 (0.001)***	0.294 (0.087)*	0.090 (0.656)
-400	0.450 (0.000)***	0.352 (0.067)*	0.098 (0.667)
-350	0.670 (0.000)***	0.314 (0.015)**	0.356 (0.075)*
-300	0.717 (0.000)***	0.304 (0.003)***	0.413 (0.029)**
-250	0.837 (0.000)***	0.349 (0.012)**	0.488 (0.030)**
-200	0.895 (0.000)***	0.365 (0.009)***	0.530 (0.021)**
-150	0.848 (0.000)***	0.419 (0.024)**	0.429 (0.090)*
-100	0.874 (0.000)***	0.490 (0.046)**	0.384 (0.213)
-50	0.900 (0.000)***	0.497 (0.027)**	0.403 (0.170)
0	0.904 (0.000)***	0.471 (0.014)**	0.433 (0.109)

Notes: For the period 1997 to 2014, the decile 10 approval-linked portfolio contains stocks for the firms that hold the 449 patents, of which 80 patents are correctly predicted to be approval-linked and 369 patents incorrectly predicted to be approval-linked. The non-approval-linked portfolio contains 4,035 patents, of which 3,765 patents are correctly predicted not to be approval-linked and 270 are incorrectly predicted not to be approval-linked. BHARs are calculated in accordance with Equation 7.4 and the returns displayed are for the period from 700 trading days prior to the approval (or pseudo approval) date for equal-weighted decile 10 portfolios. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are shown within parentheses, *** p<0.01, ** p<0.05, * p<0.1.

I also calculate portfolio buy-and-hold returns for value-weighted portfolios for the three sample periods. The value-weighted portfolios assume that an investor allocates his investment to the stocks in the portfolio in proportion to their market value at the patent publication date. The results in Table 7.8 show that the value-weighted portfolios continue to generate positive buy-and hold approval date returns. In all three portfolios, buy-and-hold returns are reduced compared to those for the equal-weighted portfolios, but they remain positive. They also remain statistically significant for the entire period for the 1962 to 2014 in-sample and 1962 to 1996 out-of-sample portfolios. The 1962 to 2014 in-sample portfolio's approval date BHARs reduce from 0.604 (60.40%) to 0.234 (23.40%), a 61.92%

reduction, For the out-of-sample portfolios, the 1962 to 1996 portfolio approval date BHARs reduce to 0.150 (15.05%) from 0.270 (27%), a 44.44% reduction. The largest change can be observed for the 1997 to 2014 portfolio approval-date BHARs which reduce by 85.40%, from 0.904 (90.40%) to 0.132 (13.20%). These results suggest that the firms with higher market values are not the drivers of approval-date returns. Based on these results the evidence suggests that the value-weighted portfolios formed from the predictions of Model 5 can be employed by investors to generate consistently positive and significant returns in the long run but this is limited to the 1962 to 2014 in-sample and 1962 to 1996 out-of-sample portfolio periods. Equal-weighted portfolios generate higher and more consistently significant returns across all three portfolios. In the following sections, I confirm the results from the equal-weighted portfolios.

Table 7.8 Decile 10 Value-Weighted Portfolio Buy-and-Hold Returns for Patents Predicted to be Approval-Linked for the In-Sample Period 1962 to 2014 and Out-of-Sample Periods 1962 to 1996 and 1997 to 2014

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1962 to 2014) Buy-and-Hold Returns (BHAR)	MAR Model (1962 to 1996) Buy-and-Hold Returns (BHAR)	MAR Model (1997 to 2014) Buy-and-Hold Returns (BHAR)
-700	0.153 (0.002)***	0.016 (0.000)***	-0.010 (0.141)
-650	0.158 (0.001)***	0.021 (0.000)***	-0.011 (0.000)***
-600	0.174 (0.000)***	0.041 (0.000)***	-0.004 (0.345)
-550	0.165 (0.001)***	0.050 (0.000)***	-0.015 (0.000)***
-500	0.190 (0.000)***	0.061 (0.000)***	0.018 (0.000)***
-450	0.219 (0.000)***	0.081 (0.000)***	0.024 (0.000)***
-400	0.222 (0.000)***	0.089 (0.000)***	0.020 (0.000)***
-350	0.229 (0.000)***	0.105 (0.000)***	0.034 (0.000)***
-300	0.226 (0.000)***	0.110 (0.000)***	0.040 (0.000)***
-250	0.241 (0.000)***	0.115 (0.000)***	0.073 (0.085)*
-200	0.239 (0.000)***	0.129 (0.000)***	0.086 (0.113)
-150	0.212 (0.000)***	0.134 (0.000)***	0.066 (0.296)
-100	0.226 (0.000)***	0.137 (0.000)***	0.085 (0.257)
-50	0.219 (0.000)***	0.147 (0.000)***	0.096 (0.245)
0	0.234 (0.000)***	0.150 (0.000)***	0.132 (0.144)

Notes: The table presents buy-and-hold abnormal returns for the period 700 trading days prior to the drug approval or pseudo approval date for value-weighted decile 10 portfolios. BHARs are calculated in accordance with Equation 7.4 and the returns displayed are for the period from 700 trading days prior to the approval (or pseudo approval) date for equal-weighted decile 10 portfolios. BHARs are weighted by the firm's market value on the patent publication date. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. For the period 1962 to 2014, decile 10 contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. For the period 1962 to 1996, decile 10 contains 846 patents, 52 correctly predicted to be approval-linked and 794 patents incorrectly predicted to be approval-linked. For the period 1997 to 2014, decile 10 contains 449 patents, 80 correctly predicted to be approval-linked and 369 patents incorrectly predicted to be approval-linked. P-values are shown within parentheses, *** p<0.01, ** p<0.05, * p<0.1.

7.3 Robustness Checks

The evidence presented in Section 7.2 suggests that the approval-linkage predictions of Model 5 can be consistently employed by investors to generate significant positive returns in the long run. These results conflict with the Efficient Market Hypothesis (EMH), as they reject the main premise of the EMH, that significant positive and consistent abnormal returns cannot be generated in the long run from investment strategies that rely on publicly available information. In this section, I analyse these results in detail to confirm that the model's predictions are in fact the driver of this apparently successful investment strategy.

7.3.1 Non-Approval Linked Patents, Potential Performance Differences

Firstly, I expect that abnormal returns are earned if the model was able to predict perfectly with a 100% approval concentration ratio. I investigate whether the performance of non-approval-linked patents differs in the 1962 to 2014 in-sample portfolio and the 1997 to 2014 and 1962 to 1996 out-of-sample portfolios, and whether portfolios containing only stocks of firms holding approval-linked patents earn significantly higher abnormal returns than those that do not. To explore this argument, I compare the performance of all three portfolios with correctly predicted approval-linked patents removed and then again with incorrectly predicted approval-linked patents removed from the portfolio. The portfolios of in-sample and out-of-sample predicted approval-linked patents generate significant approval date BHARs of 0.604 (60.40%), 0.904 (90.40%) and 0.270 (27%) respectively. The approval date BHARs detailed in Table 7.9, show that when all correctly predicted approval-linked patents are removed from these portfolios, the portfolio BHARs are affected. BHARs earned by the in-sample portfolio reduces to from 0.604 (60.40%) to 0.426 (42.60%) and for the 1997 to 2014 out-of-sample portfolio they reduce from 0.904 (90.40%) to 0.774 (77.40%). The results are similar when the portfolios are winsorised at the top and bottom 1%, with both the in-sample and 1997 to 2014 out-of-sample portfolio BHARs reducing but remaining statistically significant at the 1% level.

When the incorrectly predicted approval-linked patents are removed from the portfolios leaving portfolios containing only correctly predicted approval-linked patents, I would expect the portfolio returns to improve as the portfolios now only contain approval-linked stocks. Indeed, BHARs increase to 1.452 (145.20%) for the in-sample portfolio and increase to 1.726 (172.6%) for the 1997 to 2014 out-of-sample portfolio. However, whilst the in-sample BHARs remain statistically significant at the lower 5% level, those for the 1997 to

2014 out-of-sample portfolio become statistically insignificant. Again, the results are similar when the portfolios are winsorised at the top and bottom 1%, with both the in-sample and 1997 to 2014 out-of-sample portfolio BHARs reducing. However, the BHARs for both portfolios are statistically significant at the 5% level. The superior performance of the incorrectly predicted approval-linked patents in the 1997 to 2014 out-of-sample portfolio might explain why this portfolio outperforms that for the in-sample period (in terms of generating BHARs).

When all correctly predicted approval-linked patents are removed from the 1962 to 1996 out-of-sample portfolio, the portfolio BHARs unexpectedly increase to 0.298 (29.80%). The results are similar when the portfolio is winsorised at the top and bottom 1%, with BHARs increasing but remaining statistically significant at the 1% level.

Table 7.9 Alternative Portfolio Compositions for Decile 10 Portfolio Approval Date Buy-And-Hold Returns for the In-Sample Period 1962 to 2014 and Out-of-Sample Periods 1997 to 2014 and 1962 to 1996

Portfolio Composition	Approval Date BHAR
MAR Model (In-Sample: 1962 to 2014) 256 Approval-linked Patents Excluded	0.426 (0.000)***
MAR Model (In-sample: 1962 to 2014) 1,040 Non-approval-linked Patents Excluded	1.452 (0.045)**
MAR Model (Out-of-Sample: 1997 to 2014) 80 Approval-linked Patents Excluded	0.774 (0.000)***
MAR Model (Out-of-Sample: 1997 to 2014) 369 Non-approval-linked Patents Excluded	1.726 (0.110)
MAR Model (Out-of-Sample: 1962 to 1996) 52 Approval-linked Patents Excluded	0.298 (0.000)***
MAR Model (Out-of-Sample: 1962 to 1996) 794 Non-approval-linked Patents Excluded	-0.151 (0.381)

Notes: The table presents approval date buy-and-hold returns for portfolios comprising either only firm stocks for correctly predicted approval-linked patents or stocks for patents incorrectly predicted to be approval-linked. BHARs are calculated in accordance with Equation 7.4 for equal-weighted decile 10 portfolios. For the in-sample period 1962 to 2014, the decile 10 portfolio contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. For the out-of-sample period 1997 to 2014, the decile 10 portfolio contains 448 patents, 80 correctly predicted to be approval-linked and 368 patents incorrectly predicted to be approval-linked. For the out-of-sample period 1962 to 1996, the decile 10 portfolio contains 846 patents, 52 correctly predicted to be approval-linked and 794 patents incorrectly predicted to be approval-linked. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are shown within parentheses, *** p<0.01, ** p<0.05, * p<0.1.

When the incorrectly-predicted approval-linked patents are removed from the 1962 to 1996 out-of-sample portfolio, leaving only correctly-predicted approval-linked patents, contrary to expectations the approval date BHARs decrease to -0.151 (-15.10%) and are statistically insignificant. Again, the results are similar when the portfolio is winsorised at the top and bottom 1%

The results from both robustness test of the 1962 to 1996 out-of-sample portfolio yield results that are contrary to expectations and call for further investigation. With more detailed examination of the firm-level BHARs, the correctly predicted portfolio's returns are influenced by the negative returns of a small number of firms, who experienced negative earnings. This accounts for the poorer performance of the correctly predicted approval-linked patent portfolio. In the incorrectly predicted portfolio, the positive returns are influenced by the large positive returns of a small number of firms.

I propose that while correctly predicted, approval-linked patents perform well. I also propose that those patents that are incorrectly predicted to be approval-linked, in the decile 10 portfolio, may perform better than those correctly predicted to be non-approval-linked. I investigate whether patents incorrectly predicted as approval-linked outperform other correctly predicted non-approval-linked patents. To investigate this proposition, I compare the performance of non-approval-linked patents in the predicted approval-linked decile 10 portfolio with the performance of non-approval-linked patents in decile 1 (the portfolio of firms with patents with the lowest approval likelihood). If the proposition is valid, I expect a significant difference in performance between non-approval-linked patents in decile 10 and non-approval-linked patents in decile 1, with the decile 1 non-approval-linked patents underperforming those in decile 10. I therefore investigate the performance of the decile 1 in-sample 1962 to 2014 prediction portfolio. This portfolio contains patents that are predicted to be the least likely to be linked to an approved drug. This portfolio also has the lowest approval concentration ratio of 1.62%. This portfolio generates a positive but statistically insignificant approval date BHAR of 0.333 (33.3%). As the patents in this portfolio are least likely to be linked to an approved drug based on their patent and holding-firm publishing characteristics**,** I would expect that the correctly predicted non-approval-linked patents in this portfolio would underperform the non-approval-linked patents, incorrectly predicted to be approval-linked, in the decile 10 portfolio. When all correctly-

predicted approval-linked patents are removed from the decile 1 portfolio, leaving only the patents not linked to an approval but incorrectly-predicted to be so, the portfolio BHARs reduce to 0.079 (7.9%) (significant at the 1% level). This indicates that the correctly predicted approval-linked patents are driving the portfolio returns in this decile.

7.3.2 Choice of Market Index

As a robustness check I test whether the BHARs are sensitive to the choice of market index employed in return calculation. The Datastream Pharmaceutical Industry index and Datastream Pharmaceutical and Biotechnology Industry index[77] were again used in place of the Kenneth French Drug Industry index as alternative measures of the market return. I also use the S&P 500 index as a market index alternative that is more general in nature. Sixteen sample firms are constituents within the Datastream Pharmaceutical and Biotechnology Industry and the S&P 500 indices, with six firms being constituents of the Datastream Pharmaceutical Industry index. This may affect the relevance of these indices when used as market index alternatives. I therefore also create a bespoke index containing the returns of sample firms weighted by their market capitalisation.

The BHARs, detailed in Table 7.10, are similar in magnitude for three of the indices, with the S&P 500 index producing lower approval date BHARs of 0.505 (50.5%) (significant at the 5% level) than the two Datastream indices. The approval date BHARs were measured at their highest value of 0.647 (64.7%) for the Pharmaceutical Industry index and at 0.572 (57.2%) for the Pharmaceutical and Biotechnology Industry index (both significant at the 1% level). These BHARs are at their lowest value of 0.154 (15.4%) (significant at the 10% level) for the sample firm weighted index. BHARs are positive and statistically significant for the entire 700 days prior to the approval date for both Datastream indices. This result is similar to that in Table 7.4 where the Kenneth French Drug Industry returns were used to calculate 1962 to 2014 BHARs. BHARs are positive but statistically significant for just under half of that period when the S&P 500 index is employed. This result is similar to the sample firm weighted index, where BHARs are insignificant in the earlier section of the pre-approval period and are positive at all but one measurement point.

[77] The Datastream Pharmaceutical Industry index is comprised of 15 constituent firms, with the Pharmaceutical and Biotechnology Industry index is comprised of 47 firms.

Table 7.10 Decile 10 Portfolio Buy-And-Hold Returns for the In-Sample Period 1962 to 2014: Alternative Market Index Measures

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1962 to 2014) BHAR (Market Index-S&P 500 Index)	MAR Model (1962 to 2014) BHAR (Market Index-Datastream Pharmaceutical Industry Index)	MAR Model (1962 to 2014) BHAR (Market Index-Datastream Pharmaceutical and Biotechnology Industry Index)	MAR Model (1962 to 2014) BHAR (Market Index-Weighted Index of Sample Firms)
-700	0.170 (0.199)	0.151 (0.074)*	0.138 (0.087)*	0.020 (0.686)
-650	0.176 (0.115)	0.133 (0.056)*	0.119 (0.071)*	-0.011 (0.786)
-600	0.210 (0.128)	0.169 (0.029)**	0.151 (0.041)**	0.004 (0.924)
-550	0.228 (0.145)	0.223 (0.009)***	0.202 (0.013)**	0.042 (0.415)
-500	0.250 (0.139)	0.270 (0.000)***	0.242 (0.001)***	0.065 (0.179)
-450	0.265 (0.105)	0.311 (0.000)***	0.281 (0.000)***	0.076 (0.134)
-400	0.297 (0.121)	0.372 (0.000)***	0.336 (0.000)***	0.112 (0.059)*
-350	0.328 (0.118)	0.487 (0.000)***	0.447 (0.000)***	0.210 (0.002)***
-300	0.313 (0.069)*	0.499 (0.000)***	0.455 (0.000)***	0.195 (0.006)***
-250	0.336 (0.40)**	0.545 (0.000)***	0.496 (0.000)***	0.207 (0.007)***
-200	0.352 (0.032)**	0.576 (0.000)***	0.522 (0.000)***	0.205 (0.007)***
-150	0.436 (0.027)**	0.552 (0.000)***	0.495 (0.000)***	0.146 (0.023)**
-100	0.475 (0.031)**	0.562 (0.000)***	0.501 (0.000)***	0.129 (0.062)*
-50	0.463 (0.024)**	0.598 (0.000)***	0.529 (0.000)***	0.138 (0.062)*
0	0.505 (0.022)**	0.647 (0.000)***	0.572 (0.000)***	0.154 (0.083*

Notes: The table presents approval date buy-and-hold returns for decile 10 portfolios comprising firm stocks for patents predicted to be approval-linked. BHARs are calculated in accordance with Equation 7.4 and the returns displayed are for the period from 700 trading days prior to the approval (or pseudo approval) date for equal-weighted decile 10 portfolios. For the in-sample period 1962 to 2014, the decile 10 portfolio contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. Alternative measures of the day t simple market returns are taken from the S&P 500 Index, Datastream Pharmaceutical Industry index, Datastream Pharmaceutical and Biotechnology Industry index and a market capitalisation-weighted index of the sample firms. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

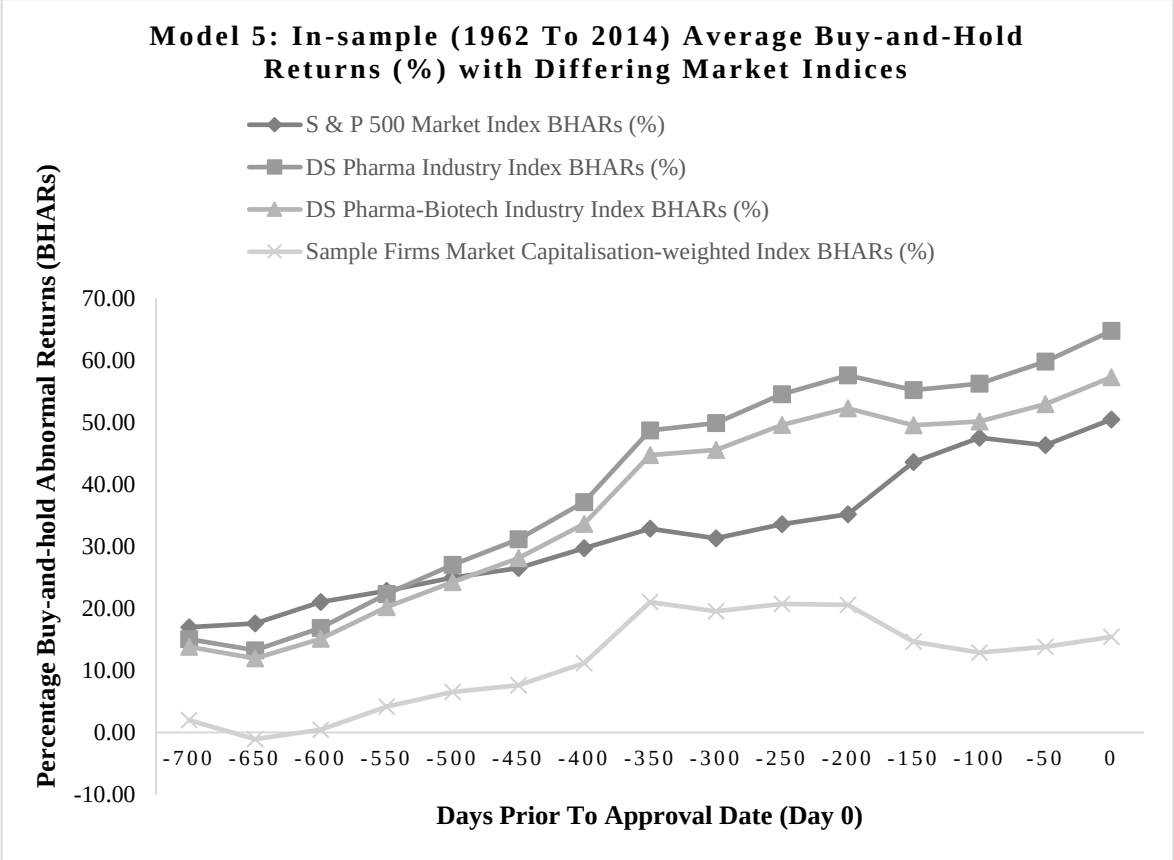
Results are similar for the 1997-2014 out-of-sample period BHARs. The approval date BHARs have their highest value of 0.965 (96.5%) for the Pharmaceutical Industry index and at 0.848 (84.8%) for the Pharmaceutical and Biotechnology Industry index (both significant at the 1% level). BHARs are at their lowest value of 0.450 (45%) (significant at the 5% level) for the weighted index. The BHARs for the Datastream indices are statistically significant for the whole period and from 500 days prior to the approval date for the weighted index.

As can be observed from Figure 7-2, buy-and-hold abnormal returns are generated by the portfolio irrespective of which proxy for the market index is used. Returns are at their highest levels when an industry-specific market index is employed.

The difference between general and industry-specific market index measures may be a result of events occurring in other industries affecting the general market index but not influencing the industry-specific indices to the same degree, for example oil price movements or the dot-com bubble. The industry-specific indices are a more reliable measure of performance within the industry as they reflect industry events unrelated to the market in general.

Although all firms compete in the prescription drug market it may be that the sample firms and the industry index firms are competing in different sections of this market. The prescription drug market is made up of many sections related to specific disease types. These market sections may be subject to differing market conditions, competition levels and revenue opportunities. This could produce differences in firm performance within the industry, especially if a firm concentrates its research efforts in only one or a very narrow range of disease types.

Figure 7-2 Average Buy-and-Hold Daily Returns for the In-Sample Period 1962 to 2014 with Returns Calculated Using Differing Market Indices



Notes: The graph displays the growth in the average percentage buy-and-hold portfolio returns for the period from 700 trading days prior to the drug approval date. The portfolio consists of stocks of the firms that hold the 1,296 patents predicted to have a drug approval linkage by Model 5 for the full in-sample period 1962 to 2014.

7.3.3 The Influence of Confounding Events on Portfolio Buy-and-Hold Returns

The presence of confounding events within the holding period (event window), especially those that are firm-specific, may have a significant effect on the portfolio returns. The longer the event window, the more likely it becomes that a confounding event might have had an influence on stock prices. Confounding events are therefore more likely to occur in this study due to the long holding periods involved. Controlling for all such events would not be without difficulty, especially as the majority of holding periods are over 700 trading days. I therefore control for a subset of some of the possible firm-specific events to estimate the influence of such events. I control for the following confounding events, other patent grants within the holding period, earnings announcements and merger and acquisition announcements.

7.3.3.1 Assessment of the Potential Effects on Portfolio Returns of the Publication of Other Patents within the Holding Period

Due to the important role of intellectual property within the pharmaceutical industry, many firms produce multiple patents each year. I firstly take account of the fact that some of the firms within the in-sample and out-of-sample portfolios publish multiple patents that have overlapping holding periods. It is therefore necessary to control for any confounding effects that may arise from patents that are published within the holding periods relating to other patents. If a firm has patents published with overlapping event windows, I calculate the holding period BHARs with and without the inclusion of these patents to assess the potential influence of the publication of these patents on portfolio returns.

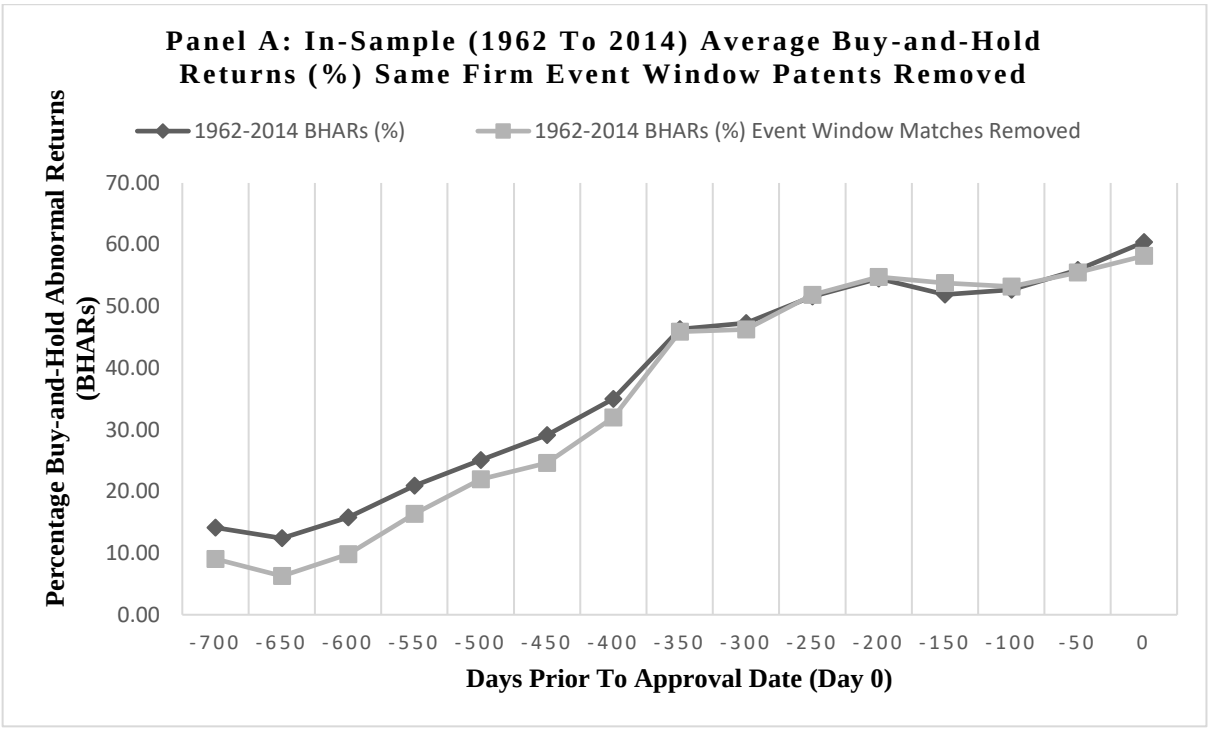
As can be observed from Table 7.11, the BHARs from the portfolios excluding firm patents with overlapping event windows are still positive and significant. The BHARs are not significant during the first year, only becoming significant a little nearer to the approval date and they are slightly lower in value. The publication of other patents by the same firm does influence BHAR values but the effect is not large in magnitude. Pharmaceutical firms produce many patents, with many of these expiring without ever being linked to any type of drug and others lapse due to non-payment of the fees required to keep them active. This result can clearly be observed in Figure 7-3, Panels A and B.

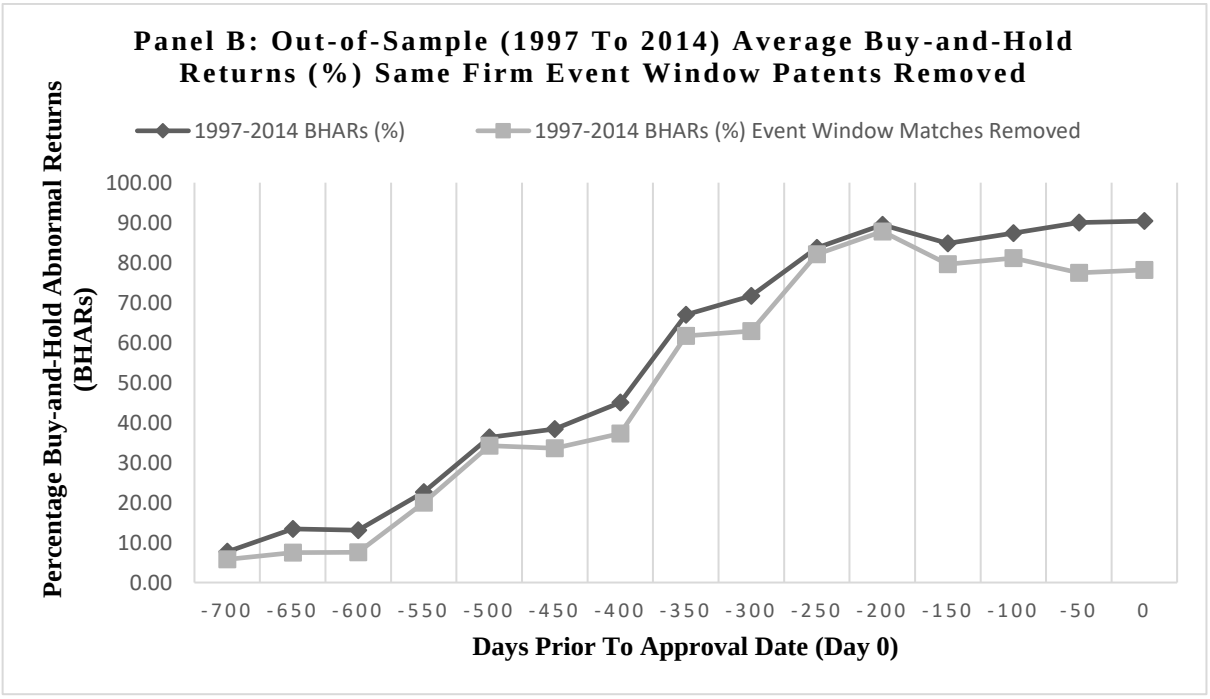
Table 7.11 Decile 10 Portfolio Buy-and-Hold Returns for the In-Sample Period 1962 to 2014 and the Out-of-Sample Period 1997 to 2014: Patents with Holding Period Overlaps Removed from Portfolios

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1962 to 2014) BHAR	MAR Model (1962 to 2014) BHAR-Multiple Patents within Holding Period Removed	MAR Model (1997 to 2014) BHAR	MAR Model (1997 to 2014) BHAR-Multiple Patents within Holding Period Removed
-700	0.141 (0.092)*	0.090 (0.310)	0.077 (0.072)*	0.058 (0.107)
-650	0.124 (0.072)*	0.063 (0.398)	0.134 (0.026)**	0.075 (0.190)
-600	0.158 (0.039)**	0.098 (0.233)	0.131 (0.032)**	0.076 (0.157)
-550	0.209 (0.012)**	0.163 (0.072)*	0.226 (0.007)***	0.200 (0.017)**
-500	0.251 (0.001)***	0.219 (0.011)**	0.363 (0.000)***	0.342 (0.003)***
-450	0.291 (0.000)***	0.246 (0.007)***	0.384 (0.001)***	0.336 (0.002)***
-400	0.350 (0.000)***	0.320 (0.002)***	0.450 (0.000)***	0.373 (0.000)***
-350	0.463 (0.000)***	0.459 (0.000)***	0.670 (0.000)***	0.616 (0.000)***
-300	0.473 (0.000)***	0.463 (0.000)***	0.717 (0.000)***	0.629 (0.000)***
-250	0.516 (0.000)***	0.519 (0.000)***	0.837 (0.000)***	0.821 (0.000)***
-200	0.545 (0.000)***	0.548 (0.000)***	0.895 (0.000)***	0.878 (0.000)***
-150	0.519 (0.000)***	0.538 (0.000)***	0.848 (0.000)***	0.796 (0.000)***
-100	0.527 (0.000)***	0.532 (0.000)***	0.874 (0.000)***	0.811 (0.000)***
-50	0.559 (0.000)***	0.555 (0.000)***	0.900 (0.000)***	0.775 (0.000)***
0	0.604 (0.000)***	0.582 (0.000)***	0.904 (0.000)***	0.782 (0.000)***

Notes: The table presents buy-and-hold returns for decile 10 portfolios comprising firm stocks for patents predicted to be approval-linked. BHARs are calculated in accordance with Equation 7.4 and the returns displayed are for the period from 700 trading days prior to the approval (or pseudo approval) date for equal-weighted decile 10 portfolios. Where a firm has other sample-patents published within the same event window these patents are removed from the portfolio and returns recalculated. For the in-sample period 1962 to 2014, decile 10 contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked, with 445 patents removed as multiple patents for same firm within the event window. For the out-of-sample period 1997 to 2014, the decile 10 portfolio contains 449 patents, 80 correctly predicted to be approval-linked and 369 patents incorrectly predicted to be approval-linked, with 139 patents removed as multiple patents for same firm within the event window. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Figure 7-3 Average Buy-and-Hold Daily Returns 1962 to 2014 and 1997 to 2014 with Same Firm Event Window Patents Removed





Notes: Panels A and B display the growth in the average percentage buy-and-hold returns for the period from 700 trading days prior to the drug approval date. In Panel A, the portfolio consists of stocks of the firms that hold the 1,296 patents predicted to have an approval linkage by Model 5 for the full in-sample period 1962 to 2014. In Panel B, the portfolio consists of stocks of the firms that hold the 449 patents predicted to have an approval linkage by Model 5 for the predicted out-of-sample period 1997 to 2014. Where a firm has other sample-patents published within the same event window these patents are removed from the portfolio and returns recalculated.

7.3.3.2 The Influence of Firm-Specific Announcements on Portfolio Returns

I also control for the impact of quarterly earnings announcements and mergers and acquisition (M&A) announcements on both in-sample and out-of-sample portfolio returns. This was done by identifying events that occurred within the portfolio stock holding period (patent publication date to approval/pseudo approval date). A final sample of 4,573 quarterly earnings announcement, 497 M&A acquirer and 176 M&A target related events were identified. When calculating buy-and-hold returns, return observations are removed from return calculations for several days around these events. For earnings announcements, return observations from 20 days prior to 20 days after the announcement date are excluded. For M&A target related announcements return observations from 30 days prior to 30 days after the announcement date are excluded and for acquirer related announcements, return observations from 15 days prior to 15 days after the announcement date are excluded.

Table 7.12 shows that the results obtained after removing the returns influenced by identified confounding events. The BHAR calculations for columns 3 and 5 used a market capitalisation-weighted index of sample firms in place of the Kenneth French Drug Industry index as a measure of the market return. In all portfolios, buy-and-hold returns are reduced but remain positive and statistically significant. The exception to this is the in-sample portfolio returns calculated using the weighted index. The 1962 to 2014 in-sample portfolio's approval date BHARs (column 2) reduce from 0.604 (60.40%) to 0.309 (30.90%), a 48.84% reduction. The approval date BHARs calculated using the weighted index increase from 0.154 (15.4%) to 0.224 (22.4%), a 45.45% increase when confounding events are excluded. For the 1997-2014 out-of-sample portfolio, the approval date BHARs reduce from 0.904 (90.40%) to 0.522 (52.2%), a 42.26% reduction, and the approval date BHARs calculated using the weighted index reduce by 13.11%, from 0.450 (45%) to 0.391 (39.10%). These results suggest that the presence of confounding events within the holding periods of the predicted approval portfolios positively influences their return performance, but they are not the sole drivers of positive and significant portfolio returns[78]. Based on these results the

[78] As a proxy for potentially missing an event that would have had a sudden influence on stock returns I exclude the top 1% of returns (after removing the defined confounding events) and estimated BHARs. BHARs remained positive and significant at 0.109 (10.9%) for 1962-2014 and 0.198 (19.8%) for 1997-2014.

evidence suggests that the portfolios formed from the predictions of Model 5 can be employed by investors to successfully generate significant positive returns in the long run.

Table 7.12 Decile 10 Portfolio Buy-and-Hold Returns for Patents Predicted to be Approval-Linked for the In-Sample Period 1962 to 2014 and Out-of-Sample Period 1997 to 2014: Controlling for Confounding Events

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1962 to 2014) Buy-and-Hold Returns (BHAR)	MAR Model (1962 to 2014) Buy-and-Hold Returns (BHAR) (Weighted Market Capitalisation Sample Index)	MAR Model (1997 to 2014) Buy-and-Hold Returns (BHAR)	MAR Model (1997 to 2014) Buy-and-Hold Returns (BHAR) (Weighted Market Capitalisation Sample Index)
-700	0.063 (0.011)**	0.049 (0.022)**	0.101 (0.045)**	0.085 (0.048)*
-650	0.001 (0.944)	-0.014 (0.016)	0.002 (0.937)	-0.026 (0.029)
-600	0.058 (0.131)	0.012(0.046)	0.175 (0.096)*	0.145 (0.106)
-550	0.105 (0.054)*	0.045 (0.059)	0.283 (0.047)**	0.242(0.143)*
-500	0.069 (0.035)**	0.034 (0.031)	0.122 (0.062)*	0.069 (0.064)
-450	0.195 (0.001)***	0.159 (0.058)***	0.504 (0.003)***	0.435 (0.171)**
-400	0.159 (0.006)***	0.089 (0.064)	0.400 (0.019)**	0.311 (0.170)*
-350	0.227 (0.001)***	0.146 (0.073)**	0.482 (0.015)	0.387 (0.197)*
-300	0.190 (0.000)***	0.099 (0.058)*	0.380 (0.004)***	0.271 (0.130)**
-250	0.213 (0.000)***	0.143 (0.049)***	0.347 (0.003)***	0.228 (0.117)*
-200	0.285 (0.000)***	0.213 (0.064)***	0.653 (0.000)***	0.520 (0.176)***
-150	0.341 (0.000)***	0.253 (0.081)***	0.661 (0.004)***	0.499 (0.223)**
-100	0.324 (0.000)***	0.206 (0.087)**	0.710 (0.003)***	0.571 (0.231)**
-50	0.332 (0.000)***	0.209 (0.079)***	0.675 (0.000)***	0.534 (0.182)***
0	0.309 (0.000)***	0.224 (0.065)***	0.522 (0.002)***	0.391 (0.165)**

Notes: The table presents buy-and-hold returns for decile 10 portfolios comprising firm stocks for patents predicted to be approval-linked. Observations linked to confounding events, such as earnings announcements and merger and acquisition announcements, are excluded from return calculations. For earnings announcements, return observations from 20 days prior to 20 days after the announcement date are excluded from return calculation. For M&A target related announcements return observations from 30 days prior to 30 days after the announcement date are excluded from return calculation. For M&A acquirer related announcements, return observations from 15 days prior to 15 days after the announcement date are excluded from return calculation. BHARs are calculated in accordance with Equation 7.4 and the returns displayed are for the period from 700 trading days prior to the approval (or pseudo approval) date for equal-weighted decile 10 portfolios. For the in-sample period 1962 to 2014, decile 10 contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. For the out-of-sample period 1997 to 2014, decile 10 contains 449 patents, 80 correctly predicted to be approval-linked and 369 patents incorrectly predicted to be approval-linked. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are shown within parentheses, *** p<0.01, ** p<0.05, * p<0.1.

7.3.4 Other Decile Portfolios' Return Performance

As a further robustness test, I also observe how other investment portfolios constructed based on the predictions of Model 5 perform, to ascertain if these alternate portfolio formulations can also earn abnormal returns for investors. The results are displayed in Table 7.13.

Table 7.13 Approval Date Buy-and-Hold Portfolio Returns for Deciles 1 to 10 for In-Sample and Out-of-Sample Periods

Equal-Weighted Decile Portfolios-Market-Adjusted Model				
Portfolio	**ACR (%) (1962 to 2014)**	**MAR Model (1962 to 2014) BHAR**	**ACR (%) (1997 to 2014)**	**MAR Model (1997 to 2014) BHAR**
Decile 10	19.75	0.604 (0.000)***	17.82	0.904 (0.000)***
Decile 9	4.09	0.354 (0.000)***	10.49	1.028 (0.000)***
Decile 8	3.55	0.300 (0.000)***	5.61	0.798 (0.001)***
Decile 7	3.32	0.973 (0.184)	4.43	0.228 (0.000)***
Decile 6	4.09	0.899 (0.114)	5.80	0.130 (0.021)**
Decile 5	3.78	0.640 (0.216)	5.17	0.073 (0.171)
Decile 4	2.55	0.107 (0.000)***	5.53	0.022 (0.577)
Decile 3	2.78	0.167 (0.005)***	5.80	-0.066 (0.020)**
Decile 2	3.24	0.163 (0.000)***	8.26	0.110 (0.219)
Decile 1	1.62	0.343 (0.166)	9.13	2.206 (0.220)

Notes: The table presents approval date buy-and-hold returns for decile 1 to 10 portfolios comprising firm stocks for patents predicted to be approval-linked. BHARs are calculated in accordance with Equation 7.4 for equal-weighted decile 10 portfolios.

For the period 1962 to 2014, decile 10 contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. Decile 9 contains 1,296 patents, 53 correctly predicted to be approval-linked and 1,243 patents incorrectly predicted to be approval-linked. Decile 8 contains 1,295 patents, 46 correctly predicted to be approval-linked and 1,249 patents incorrectly predicted to be approval-linked. Decile 7 contains 1,296 patents, 43 correctly predicted to be approval-linked and 1,253 patents incorrectly predicted to be approval-linked. Decile 6 contains 1,295 patents, 53 correctly predicted to be approval-linked and 1,242 patents incorrectly predicted to be approval-linked. Decile 5 contains 1,296 patents, 49 correctly predicted to be approval-linked and 1,247 patents incorrectly predicted to be approval-linked. Decile 4 contains 1,296 patents, 33 correctly predicted to be approval-linked and 1,263 patents incorrectly predicted to be approval-linked. Decile 3 contains 1,295 patents, 36 correctly predicted to be approval-linked and 1,254 patents incorrectly predicted to be approval-linked. Decile 2 contains 1,296 patents, 42 correctly predicted to be approval-linked and 1,254 patents incorrectly predicted to be approval-linked Decile 1 contains 1,296 patents, 21 correctly predicted to be approval-linked and 1,275 patents incorrectly predicted to be approval-linked.

For the period 1997 to 2014, decile 10 contains 449 patents, 80 correctly predicted to be approval-linked and 369 patents incorrectly predicted to be approval-linked. Decile 9 contains 448 patents, 47 correctly predicted to be approval-linked and 401 patents incorrectly predicted to be approval-linked. Decile 8 contains 446 patents, 25 correctly predicted to be approval-linked and 421 patents incorrectly predicted to be approval-linked. Decile 7 contains 451 patents, 20 correctly predicted to be approval-linked and 431 patents incorrectly predicted to be approval-linked. Decile 6 contains 448 patents, 26 correctly predicted to be approval-linked and 422 patents incorrectly predicted to be approval-linked. Decile 5 contains 445 patents, 23 correctly predicted to be approval-linked and 422 patents incorrectly predicted to be approval-linked. Decile 4 contains 452 patents, 25 correctly predicted to be approval-linked and 427 patents incorrectly predicted to be approval-linked. Decile 3 contains 448 patents, 26 correctly predicted to be approval-linked and 422 patents incorrectly predicted to be approval-linked. Decile 2 contains 448 patents, 37 correctly predicted to be approval-linked and 411 patents incorrectly predicted to be approval-linked Decile 1 contains 449 patents, 41 correctly predicted to be approval-linked and 408 patents incorrectly predicted to be approval-linked.

The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

For the 1962 to 2014 in-sample period, five of the other decile portfolio formulations do earn positive abnormal returns but at levels lower than those for the decile 10 portfolio. The decile 5, 6 and 7 portfolios approval date returns are higher that decile 10's returns, but they are statistically insignificant. Whilst the decile 1 portfolio also generates positive abnormal returns, these lower than those for decile 10 and are not statistically different from zero. The decile portfolios for the out-of-sample period, 1997 to 2014, produce only two portfolios, 1 and 9, that produce higher approval date returns that the decile 10 portfolio. Decile 1's returns, although strongly positive, are statistically insignificant. Decile 9's approval date BHARs are 13.7% higher than those for decile 10 are and are statistically significant. This result may be driven by the influence of extreme return values in this decile. To confirm this, I observed the returns and noted several extreme values, I therefore winsorised both deciles 9 and 10 approval date BHARs firstly at the top and bottom 1%. This reduces approval date returns for both deciles, with decile 9's returns reducing to 0.852 (85.2%) and the return for decile 10 to 0.776 (77.6%). However, the returns for decile 10 outperform those for decile 9 when returns are winsorised at the top and bottom 5%. Decile 9's approval date BHAR is 0.327 (32.7%) compared to decile 10's return of 0.567 (56.7%).

The large positive (2.206), but insignificant approval date return for decile 1 was also found to be driven by extreme values. When the approval date returns were winsorised at the top and bottom 1% the returns reduced in value to -0.061 (-6.1%), but were still statistically insignificant, with a p-value of 0.203. The returns reduced further to -0.131 (-13.1%) and became statistically significant with a p-value of 0.000, when they were winsorised at the top and bottom 5%.

These results and robustness checks indicate that the predictions of Model 5 can be employed to construct an investment portfolio that can generate market-adjusted abnormal returns consistently across several time periods and employing differing market index measures.

7.3.5 Size-Adjusted Portfolio Returns

Barber and Lyon (1997) and Lyon, Barber and Tsai (1999) suggested that the market-adjusted BHAR has many biases, such as rebalancing bias, new-listing bias and skewness

bias. Moreover, Lyon et al. (1999) and Bouwman, Fuller and Nain (2007) believed that a size-adjusted BHAR is a more reliable indicator for long-term firm performance. To address this issue, I also estimate the size-adjusted BHAR.

The size-adjusted BHAR at the drug approval date is defined as:

$$BHAR_{it} = \prod_{t=1}^{\tau}[1 + R_{it}] - \prod_{t=1}^{\tau}[1 + (R_{pt})] \quad \text{(Equation 7.5)}$$

Where, R_{it} denotes firm's daily stock return starting from the day a patent is published. R_{pt} denotes the daily stock for the firms matched to the decile 10 firms using market capitalisation and market-to-book ratios. To build the size-adjusted portfolio, all CRSP pharmaceutical firms (SIC codes 2833, 2834 2835, 2836 and 8731) are sorted in descending order by market capitalisation. I then match the firms in the decile 10 portfolio to non-sample firms whose market capitalisation is within plus or minus 20% of that of the sample firms. The firm that has the closest market capitalisation match is selected and then checked to ensure that it matches using a second criterion that is that their market-to-book ratio is within plus or minus 20% of that of the sample firm. If market-to-book ratio is outwith the +/- 20% range, then the firm with the second nearest market capitalisation match is selected and its market-to-book ratio checked. This process is carried out for all the firms within the portfolio. As this study is analysing a single industry rather than the market as a whole there are a limited number of firms with which portfolio firms can be matched and so it is not possible to create a size-matched portfolio of pharmaceutical firms. I do, however; create a size-matched portfolio based on firms drawn from other industries.

As shown in Table 7.14, BHARs are positive and significant for the entire 700-day period and are significant at the 1% level at the approval date. These returns indicate that, for the in-sample period, the decile 10 portfolio firms earn higher levels of BHARs when compared to a portfolio of sized-matched firms. The returns are also higher than any of the returns calculated using various alternate market return index measures. This result may be influenced by the fact that the approval concentration ratio, the proportion of correctly predicted approval-linked patents, for the size-matched decile 10 portfolio is 27.29% compared to the 19.75% for the non-matched decile 10 portfolio.

Table 7.14 Decile 10 Portfolio Buy-and-Hold Returns for the In-Sample Period 1962 to 2014: Size-Matched Portfolio

Days Prior to Drug Approval or Pseudo Approval Date	MAR Model (1962 to 2014) Buy-and-Hold Returns (BHAR)
-700	0.458 (0.071)*
-650	0.393 (0.060)*
-600	0.459 (0.041)**
-550	0.513(0.030)**
-500	0.542 (0.009)***
-450	0.605 (0.008)***
-400	0.639 (0.009)***
-350	0.745 (0.003)***
-300	0.789 (0.002)***
-250	0.869 (0.003)***
-200	0.900 (0.004)***
-150	0.722 (0.004)***
-100	0.749 (0.008)***
-50	0.853 (0.004)***
0	0.972 (0.009)***

Notes: The table presents results for the univariate regression of buy-and-hold abnormal returns on the drug approval or pseudo approval date for an equal-weighted portfolio. The equations for buy-and-hold abnormal returns calculation (Equations 7.5) is shown below;

$BHAR_{it} = \prod_{t=1}^{\tau}[1 + R_{it}] - \prod_{t=1}^{\tau}[1 + (R_{pt})]$ (Equation 7.5)

R_{it} is the day t simple return on a sample firm, R_{pt} is the day t simple return for the size-matched portfolio firm, matched on market capitalisation and market-to-book ratio. Returns are displayed for the period from 700 trading days prior to the approval date. For the in-sample period 1962 to 2014, decile 10 contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. For the size-matched portfolio 601 decile 10 firms are matched on market capitalisation and market-to-book ratio, 164 patents correctly predicted to be approval-linked and 437 patents incorrectly predicted to be approval-linked are matched. The portfolio-holding period for approval-linked patents is the time from the publication date of the patent to the approval date of the drug to which it is linked. For patents not linked to an approved drug the portfolio holding period is the time from the publication date of the patent to the pseudo approval date, 1,078 calendar days later. P-values are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

7.4 Short Run Excess Returns

Sections 7.2 and 7.3 provide evidence that investors earn significant positive long run returns if they invest in Model 5's predictions. To do so they must often hold stocks for an extended period of time, with the patent publication to approval period often being measured in years.

As a further test of Model 5's predictions I test whether investors can earn significant positive short run returns by investing in and holding a stock for a short time after patent publication. I again calculate the returns for the decile 10 portfolio of firms, calculating the CAR for all the portfolio firms treated as a group. Returns are calculated using a simple Market Adjusted Returns (MAR) model with the abnormal return and CAR calculation being those detailed in equations 7.1 and 7.2 in Section 7.2.1.1. Each stock in the portfolio is purchased on the day of patent publication (day 0). The CAR is calculated for a two-day

event window (from day 0 to day l). I also calculate the CAR for a three-day event window (from day -1 to day +1) and a five-day event window (from day -2 to day +2) to measure any movements in the market prior to patent publication. As a market index I utilise the weighted index introduced in Section 7.3.2.

In addition, I control for earnings announcements, merger and acquisition announcements and multiple patent grants by a firm. I therefore exclude patents where firms announce quarterly earnings, merger-related news or whose holder has multiple patents published within the event windows.

The CARs, detailed in column 2 of Table 7.15, show that average CARs of -0.3% are earned over the 2-day event window period as a whole. This reduces to -0.336% for the 3-day window, increasing slightly to -0.269% for the 5-day window. These returns are significant at the 10% level for the 2 and 3-day windows but insignificant for the 5-day window. When patents with confounding events are excluded from calculations, all three event windows have CARs that are negative and insignificant.

Table 7.15 Decile 10 Short Window Portfolio Returns for the In-Sample Period 1962 to 2014

	CAR (%)	CAR (%) Excluding Confounding Events
2-Day Event Window (day 0 to day 1)	-0.300 (0.057)*	-0.240 (0.136)
3-Day Event Window (day -1 to day 1)	-0.336 (0.094)*	-0.276 (0.177)
5-Day Event Window (day -2 to day 2)	-0.269 (0.285)	-0.264 (0.264)

Notes: The table presents results for the univariate regression of cumulative abnormal returns on day 1 for the 2-day event window and day 2 for the 3 and 5-day event windows, where day 0 is the patent publication date. Cumulative abnormal returns are calculated using Equations 7.2 from Section 7.2.1.1. R_{it} is the day t simple return on a sample firm; R_{mt} is the day t simple market return for the market capitalisation-weighted index of the sample firms. The decile 10 portfolio contains 1,296 patents, 256 correctly predicted to be approval-linked and 1,040 patents incorrectly predicted to be approval-linked. P-values are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

These results contradict those of Austin (1993) and Plumlee et al. (2015), who find positive and significant returns to patent grants. However, Dornelles and Ali (2014) find evidence of negative but insignificant abnormal returns in their multi-industry study. Whilst Koreamaki and Takalo (2013, p. 205), in their iPhone case study, find that "patent grants do not appear to bring new information to the market in a significant manner".

My results may be explained by the nature of pharmaceutical patent documents. They contain complex and technical information that investors may find difficult to interpret and assign value to (Liu, 2006). Also, given the highly uncertain nature of innovative events, investors may struggle to assess the subsequent development potential of a particular invention. This is made more difficult by the high patenting volumes of the pharmaceutical industry.

7.5 Conclusion

In the previous chapter, I demonstrated that the models developed based on certain patent and firm publishing characteristics exhibit a level of drug approval predictive ability that is higher than would be expected by chance alone. Results indicate that Model 5 can achieve an approval concentration ratio of 19.75% for in-sample predictions and 17.82% for out-of-sample predictions. Based on this encouraging predictive performance the objective of this chapter is to test whether investors can utilise this model to earn abnormal stock returns, a challenge to the Efficient Market Hypothesis. These results demonstrate that an investment in the firms predicted to hold approval-linked patents results in significant abnormal returns being earned by an investor. The analysis shows that this model can generate substantial abnormal returns in certain periods. For example, the model generates significant positive BHARs for patents (published during the period 1997 to 2014) predicted to be linked to an approved drug. For these patents, the model can generate BHARs of 90.40% for the publication to approval date holding period.

The chapter also explores potential determinants of these returns to approval portfolios (including the presence of incorrect predictions, Type II errors), the effects of confounding firm-specific events and attempts to explain how these factors may affect the model's return performance.

I firstly explore whether the presence of non-approval-linked patents in predicted approval-linked portfolios (in-sample and out-of-sample) affect the performance of these portfolios. I find that the presence of non-approval-linked patents still generate positive and significant holding period returns but at lower levels than actual approval-linked patents. When all non-approval-linked patents are excluded the BHARs increase by 240% for the in-sample portfolio and 190% for the 1997 to 2014 out-of-sample portfolio, although in this instance they are not statistically significant. When all approval-linked patents are excluded, BHARs are reduced but are still positive and significant. With the 1997 to 2014 out-of-sample

portfolio, BHARs are 77.4%, which is 86% of their whole portfolio level. For the in-sample period BHARs are 42.6% which is 71% of their whole portfolio level. This finding partly explains why the in-sample portfolio underperforms this out-of-sample portfolio, even though it achieves a higher approval concentration ratio.

Secondly, I explore whether the presence of confounding firm-specific events within the portfolio-holding period influences returns. I find that the exclusion of certain confounding events (earnings announcements and mergers and acquisitions announcements) from the portfolios leads to reductions in long run portfolio returns. For example, the model generates significant positive approval date BHARs of 0.604 (60.40%) for the in-sample portfolio, which reduce to 0.309 (30.90%) when these confounding events are excluded from return calculations. For the 1997 to 2014 out-of-sample portfolio approval date BHARs reduce from 0.904 (90.40%) to 0.522 (52.2%) when confounding events are excluded from return calculations. These results suggest that the presence of highly confounding events within the predicted approval portfolios positively influences their return performance, but they are not the sole drivers of positive and significant portfolio returns.

I also test whether BHAR calculations are sensitive to the use of differing market return measures for the 1962 to 2014 in-sample period. I find that using differing market return measures leads to changes in the long run portfolio returns, but the portfolios still earn positive and significant long run portfolio returns. The model generates significant positive approval date BHARs of 0.154 (15.40%) when using the sample firm weighted index, BHARs of 0.505 (50.50%) when using the S&P 500 index and 0.572 (57.20%) with the Datastream Pharmaceutical and Biotechnology Industry index, whilst using the Datastream Pharmaceutical Industry increased BHARs to 0.647 (64.70%).

Overall, I find that approval prediction portfolios generated using the Model 5 outperform the market. Misclassification of non-approval-linked patents as approval-linked negatively impacts on the returns to predicted approval-linked portfolios, but not to the extent that positive and significant BHARs cannot be generated. The initial publication of a patent has no significant effect on returns (when confounding events are excluded), with significant positive returns being earned during the extended holding period.

Chapter 8 Conclusion

8.1 Introduction

The main objectives of this thesis were to firstly investigate the characteristics of USPTO pharmaceutical patents and to evaluate whether the characteristics of the subset of these patents that are developed into approved drugs differs from those that do not receive regulatory approval. Secondly, if a difference exists, can certain patent and firm-level characteristics be employed as indicators of the likelihood of receiving regulatory approval and be utilised to predict such an event? Finally, I assess whether drug approval prediction modelling can form the basis of a profitable investment strategy and explain why investing in predicted drug approval-linked firms is likely to be an optimal investment strategy. I do not examine what factors make a drug commercially successful as this depends on factors outwith the firm environment, for example, competition conditions and the disease type that the drug treats. I chose the U.S. pharmaceutical industry as a sample for the study given its size, the level of patenting activity within this industry and its unique institutional features (see Chapter 1, Section 1.2). Several interesting findings have emerged from the study. These findings contribute to the understanding of why certain patents are more likely to be linked to an approved drug.

This chapter summarises the important findings of the empirical work carried out in the thesis and presents its implications and limitations, as well as suggestions for improvements and extensions of the study. The remainder of the chapter is organised as follows. Section 8.2 presents the key findings and contributions of the study and provides a summary of the main empirical findings from Chapters 5, 6 and 7, which investigated the factors that influence drug approval outcomes, predictive modelling and investment strategy. Section 8.3 discusses the implications of these findings for future studies, firms and investors. Section 8.4 discusses some of the potential limitations of the study, offering suggestions for improvement, and Section 8.5 highlights potential opportunities for future research.

8.2 Summary and Discussion of the Main Findings of the Thesis

8.2.1 Introduction

This section presents a short overview of the main research findings of empirical Chapters 5, 6 and 7 of this thesis. These can be summarised into the following areas. Firstly, I conducted an analysis of the characteristics of approval-linked patents to ascertain if they differ from patents that do not result in an approved drug. In addition, I determine the nature

of the publishing behaviour of firms that hold approval-linked patents. Secondly, can these patent characteristics and firm publishing behaviour be utilised to construct an approval prediction modelling methodology. Finally, can using the predictions of such a model and investing in firms that hold predicted approval-linked patents be the foundation of a profitable investment strategy.

8.2.2 Characteristics of Approval-Linked Patents

The first empirical chapter, Chapter 5, examined the relationship between the patent characteristics and firm-level publishing characteristics and the likelihood of a patent possessing a link to a subsequent FDA drug approval. I empirically test several hypotheses that attempt to explain how approval-linked patents exhibit certain patent characteristics.

Five of the hypotheses tested relate to information that is detailed within the patent document. These include the number of inventors, the number of references to non-patent literature, the number of patent claims, the age of the references to non-patent literature and the age of the references to previous patents. Two others relate to the number and age of subsequent citations a patent receives. I also introduce two new patent-related hypotheses that test the role of patent independent and dependent claims.

Prior researchers investigated the patent characteristics within these seven hypotheses in different combinations in relation to measures of patent quality, patent value, firm value or firm performance. This prior research postulated that patents that had larger inventor teams, cited more NPR, made more patent claims and cited newer patent and non-patent sources were of higher economic value. The characteristics of approval-linked patents are not fully known or understood. To my knowledge, this study is the first to develop and test any of these hypotheses with a goal to predict the likelihood of a link to a drug approval.

Cassiman et al. (2008) and Liu (2014) postulated that the number of inventors is an indicator of the level of resources a firm commits to a research project. Whilst, Singh and Fleming (2010) suggested that teams of inventors rather than sole inventors are more likely to generate breakthrough inventions. It follows from their argument that firms whose research teams are larger will produce more breakthrough inventions and breakthrough inventions often lead to the creation of new drugs in the pharmaceutical industry. Results from prior empirical tests of this hypothesis are inconsistent with Guler and Nerkar (2012) who found that inventor team size had a negative but insignificant relationship with the launch of new

drug products. Consistent with the findings of Singh and Fleming (2010), I find that the patents linked to an approved drug have inventor teams that are on average 30% larger than for patents not linked to an approved drug. I therefore conclude that approval-linked patents have higher numbers of inventors listed in their patent documents when compared to patents with no approval linkage.

The empirical literature on the role of citations to non-patent literature and their relationship with firm value exhibits more consistent evidence with Deng et al. (1999) and Hirschey and Richardson (2004), in their studies of U.S. high technology firms, finding a positive relationship between the level of NPR cited by a firm's patents and firm market value. I find that, on average, approval-linked patents contain nearly three times as many references, citing 27 references compared to 10 such references in patents with no approval linkage. I therefore conclude that approval-linked patents reference significantly higher numbers of NPR that those with no approval linkage.

Evidence from the patent claims literature tends to focus on the relationship between patent claims and the likelihood of litigation and views litigated patents as being of higher value. Tong and Frame (1994) and Lanjouw and Schankerman (1999, 2001) found that increased numbers of patent claims were linked to higher expected patent value. Whilst Reitzig (2004) found a positive correlation between the number of independent claims and patent value. I find that, on average, approval-linked patents make 19 claims within the patent document compared to 13 claims for those patents without such a link. When these total claims figures are broken down into their independent and dependent claim components, the approval-linked patents make 3 independent claims and 16 dependent claims compared to 2 independent and 11 dependent claims for those patents with no drug approval link.

The influence of the age of the patent and non-patent knowledge upon which a patent is based has been tested empirically by prior literature, again with mixed results. Schoenmakers and Duysters (2010) found that patents covering radical inventions cite older knowledge, whilst Fabrizio (2009), in her study of the U.S. pharmaceutical and biotechnology firms, found that patents with shorter average backward citation lags are associated with possessing greater importance in their area of research. Arts and Veugelers (2012) also found that breakthrough inventions in pharmaceutical and biotechnology have a shorter average backward citation lag compared to patents not associated with a breakthrough invention. I find that the youngest patent cited by approval-linked patents is less than half the age of that

referenced by patents with no drug approval link, the youngest patent reference cited is 331 versus 694 days old. For NPR there is also a marked difference in the age of the references cited between approval and non-approval-linked patents. Approval-linked patents cite NPR that are younger than those references cited in non-approval-linked patents, the youngest reference cited is 1,004 versus 1,555 days old.

Prior literature is more in agreement concerning the role of patent citations received, referred to as forward citations, as an indicator of a patent's technological or economic value. This indicator is the most widely used proxy for the value of the underlying invention with Hall et al. (2000) finding that firms with patents receiving higher levels of forward citations have higher stock market values. Gambardella et al. (2008) found that, in surveys of patent holders, their estimates of the actual value earned from specific inventions are positively correlated with forward citations. I find that citations from subsequent patents also differ markedly between approval and non-approval linked patents. Approval-linked patents receiving an average of 14 citations within a five-year post-publication period compared to 7 citations for non-approval-linked patents. Approval-linked patents also receive their first citation more quickly that patents with no approval link, receiving their first citation an average of 230 days post publication compared to 839 days for patents without an approval link.

8.2.3 Publishing Behaviour of Firms Holding Approval-linked Patents

There are three aspects of firm publication behaviour analysed in Chapter 5 total firm publications, the percentage of publications co-authored between the firm and academia and whether a firm co-authored with other firms. In the area of firm publishing behaviour there have been mixed results from prior literature with Kapoor and Klueter (2015), in their global pharmaceutical study, found that a firm's scientific publications have a negative but statistically insignificant influence on the likelihood of a drug entering clinical trials. Cockburn and Henderson (1998) found that publication volume has a generally negative and insignificant effect on a patent's importance but that co-authoring with the public sector has a positive and significant effect. However, Gittelman and Kogut (2003) found, for U.S. biotechnology firms, that publishing volumes significantly raise the level of patent citations received but that co-authoring has a positive but insignificant effect. I use the co-authoring percentage as a proxy for the degree of connectedness with university scientists. This follows a similar approach to that employed by Cockburn and Henderson (1998) and Fabrizio

(2009). I find that firms that hold approval-linked patents produce higher median levels of publications of 138 per year compared to 60 for firms holding patents without a link to an approved drug, although their mean levels are similar at 227 publications per year. The difference between the two firm groups is not as substantive when investigating the proportion of these publications that are co-authored with academia. Approval-linked patent holders have a mean co-authoring share of 32.8% compared to 32.3% for holders of non-approval-linked patents. The results for co-authoring with other firms is also similar between the two groups with 96.8% of approval-linked patent holders co-authoring with other firms and 97.5% of the holders of non-approval linked patents doing so.

8.2.4 Other Firm Characteristics

Studies that have considered firm size in the context of patenting and firm performance have reached inconsistent conclusions about the role that firm size plays. DiMasi (2014) found that larger firms were less successful in obtaining U.S. regulatory approval, but smaller firms had lower returns in terms of sales. However, in 2015 he and his co-authors found that the number of approved applications sponsored by larger firms were more than double those for smaller firms. This result was contradicted by Wang, Avorn and Kesselheim (2013), who found that smaller firms sponsored 59% of drug applications immediately approved by the FDA. These inconclusive results from prior literature make it essential that this study controlled for firm size. I find that firms holding approval-linked patents are smaller than those holding non-approval-linked patents with average total assets of $2,079 million compared to $2,482 million.

Prior literature has found a variety of findings around the role firm age plays in the innovation process. Barnett (1990) found that firm age has a negative effect on innovation levels within a firm. However, Sorensen and Stuart (2000) found a positive effect, as increased experience appears to enhance innovation ability within a firm. I find that approval-linked patents belong to firms that are younger, having an average age of 70 years compared to 82 years for non-approval-linked patents.

Profitability has been found to be an important firm-specific variable that affects patenting (Deng et al., 1999). It is one of the key characteristics that are associated with the extent of patenting and the performance of the firm. When profitability is high firms have more available funds to invest and tend to spend more on R&D, increasing patenting levels. I find

that approval-linked patents belong to less profitable firms that have a return on assets of -1.5% compared to 4% for non-approval-linked patents.

Prior experience in the regulatory drug development process also appears to be a factor in the likelihood of approval success. Heemstra, Leufkens, Rodgers et al. (2011) found that 63% of the U.S. approved drugs they investigated were sponsored by firms that had prior experience of developing that type of drug. However, Putzeist, Heemstra, Garcia et al. (2012) found that 63% of drug approval applications were sponsored by firms with no prior experience. They also found that these firms had 54.2% of their applications approved, whilst firms with prior development experience had 80.9% of their applications approved. I find that 77.4% of the firms that hold approval-linked patents have prior experience of the FDA regulatory process compared to 87.4% of the firms that hold non-approval linked patents.

From these results it can observed that these patent and firm-level publishing characteristics can be utilised as a way of identifying those patents more likely to possess a link to a successful regulatory drug approval.

8.2.5 Predictive Modelling

Chapter 6 develops new binary logistic models for the prediction of regulatory approval success based on my chosen patent and firm-level characteristics. To my knowledge, this is the first study to attempt to predict the likelihood of regulatory approval using publicly available patent and firm publication data. This thesis proposes that our knowledge of the factors that drive the likelihood of drug approval can be substantially improved by modelling the relationship between patent and firm-level publishing characteristics and regulatory drug approval success. I demonstrate this by identifying several new hypotheses for predicting regulatory drug approval success, and test these for the first time in this study.

I measure the importance of my chosen predictors to the likelihood of regulatory success. The univariate regression analysis results for the patent-related variables show that citations received from subsequent patents (forward citations) was ranked the most important predictor with an odds ratio of 1.585. The number of independent claims listed in the patent document ranked second with an odds ratio of 1.512. All but two of these variables, the citation lag-related measures, had highly statistically significant associations with regulatory success. The age of cited patent and non-patent literature were inversely related to regulatory

success but were not statistically significant. The results also suggest that the publications variables, whilst statistically significant, are inversely related to regulatory success, with firm publishing having an odds ratio of 0.925 and co-authoring with academia an odds ratio of 0.973. My findings are consistent with those of Kapoor and Klueter (2015), as I find that firm publications, whether they be solely authored by firm scientists or co-authored by academia, have a negative influence on the likelihood of the firm's patents being linked to approved drugs. However, in contrast, the relationship I find is statistically significant. Whilst firms may publish and co-author, which may assist in the early stages of the drug discovery process, this may have less impact in the later developmental stages of the process.

I also found that larger, profitable firms with prior experience of the regulatory process were more likely to hold patents that produced drugs with regulatory approval than smaller, less profitable firms with no prior experience. Whilst the age of a firm holding the patent appeared to have almost no effect on the likelihood of regulatory success, having an odds ratio of 1.034. These results provide strong support for their inclusion in a prediction model. To ensure important interactions between predictors were not missed, I used multivariate regression techniques to help identify models with the optimal predictor variable combination.

For my research, I selected a binary logistic regression technique that links a selection of predictors to the probability of regulatory success. This method is extensively used in the literature on bankruptcy prediction (see, for example, Lin and Piesse (2004) and Campbell et al. (2008)), and takeover target prediction (see for example Palepu (1986), Espahbodi and Espahbodi (2003) and Danbolt, Signaos and Tunyi (2016)). From the logistic models estimated, I developed two new prediction models, whose parameters are shown in Chapter 5, Tables 5.11 and 5.12. These models performed particularly well in terms of their log likelihood measures and have a superior classification and predictive ability when compared with pure chance. The results on model performance are robust to several model specifications and modelling choices (discussed in Chapter 5, Section 5.5).

The first of these models contained only the patent characteristics available at the patent publication date, whilst the second also included patent forward citations. I found that four of the patent-related variables and two of the publication-related variables, taken together, provided an effective foundation for predictive purposes. In the publication date model, the patent-related variable that has the greatest influence on the likelihood of regulatory success,

with a statistically significant odds ratio of 1.445, is the independent patent claims variable. In the other chosen model, forward citations remain the patent-related variable that appears to have the greatest influence on the likelihood of regulatory success, with a statistically significant odds ratio of 1.548. In both models the results suggest that there is an inverse relationship between firm age, prior regulatory experience and both publishing-related variables and the probability of regulatory success, although prior experience is statistically insignificant. That is, regulatory success is less likely for older, more experienced firms and firms whose researchers publish more, both on their own or with academic researchers. In both models, firm size and profitability positively influence the likelihood of regulatory success, but only size is statistically significant with odds ratios of 2.209 for the publication date model and 2.429 for the model including forward citations.

The objective of this thesis is to construct a model that as well as exhibiting explanatory power also possesses predictive ability. To use either of my selected models as project decision-making tools or an investment portfolio selection mechanism knowing the probability of an outcome occurring is not enough, they also need to predict whether a patent will be linked to an approved drug or not. My models are attempting to predict a relatively rare event, as approval-linked patents comprise 4.88% of the total patent sample.

The results in Chapter 6 show that, while some of the new variables are found to be statistically insignificant in the regression model, their inclusion significantly improves its performance. The predictors identified using these regressions resulted in two well-performing models. The two selected models are tested for classification and predictive ability using in-sample and out-of-sample analysis in Chapter 6. For the in-sample prediction period, 1962 to 2014, the publication-date model correctly predicted 256 out of 632 approval-linked patents, with correctly predicted approval-linked patents comprising 19.75% of the approval-linked portfolio. The model including the forward citation-related variables predicted 259 out of 632 approval-linked patents, with correctly predicted approval-linked patents comprising 19.98% of the portfolio. Through a comparative analysis of the forward citation model against the publication-date model, it can be concluded that their performances are almost identical, as the differences in predictive accuracy are very small, with the forward citations model outperforming the publication-date model for the prediction of approval-linked patents, by less than 1%. Nevertheless, the publication-date model has the advantage of being based on data that are available as soon as a patent is published. Both models exhibit a predictive performance that is superior to that of a pure

chance-based patent selection approach with both models out-performing chance by a factor of four.

The relevance of the selected models is based on their forecasting ability, that is, their ability to generate correct out-of-sample predictions of approval and non-approval linked patents. This is illustrated by testing the models' predictive ability on an out-of-sample period using data from 1997 to 2014. The publication-date model predicts 80 out of 350 approval-linked patents, with correctly predicted approval-linked patents comprising 17.82% of the portfolio. The model including the forward citation related variables predicted 75 out of 350 approval-linked patents, with correctly predicted approval-linked patents comprising 16.70% of the portfolio.

8.2.6 Investing in Firms Holding Patents Predicted to be Approval-linked

This thesis develops an investment strategy to predict regulatory success within the pharmaceutical industry and to manage a portfolio of potential approval-linked firms to maximise potential returns. It concentrates on the efficient use of publicly available information to forecast future regulatory success and uses this as a foundation to inform the construction of an investment portfolio. The thesis explores the possible economic gains accruing to a portfolio of firms holding predicted approval-linked patents. By utilising the predictions from individual models, a portfolio of firms holding predicted approval-linked patents is created that achieves abnormal returns and exhibits lower misclassification rates than a selection approach based on pure chance.

As one potential application of these new models, I evaluate whether they can be used as an investment tool to generate consistently positive abnormal returns for investors. The motivation for this test is the consensus that firms and investors gain substantially from drug development and approval announcements (See, for example, McNamara and Baden-Fuller (2007), Dedman et al. (2008), Hamill et al. (2013) and Hamill et al. (2018)).

I find that the predicted approval-linked portfolio generates positive abnormal returns and that these returns are driven by the correctly predicted approval-linked patents. That is, the returns to the portfolio decrease in magnitude when approval-linked patents are excluded from the portfolio. This finding suggests that model predictive ability does explain the returns to predicted approval-linked portfolios. The evidence in this study also shows that

false positives earn positive abnormal returns and, on average, perform better than firms holding non-approval linked patents in the lowest approval probability decile. The long run returns to approval-only portfolios are positive and statistically significant.

Prior studies, generally, attribute the ability to generate positive abnormal returns to market inefficiency. Nonetheless, to date, there has been no explanation of how market inefficiency behaves in this case.

Overall, the results in this thesis provide evidence in favour of the hypothesis that an investment strategy can achieve abnormal returns using publicly available information. These findings provide investors with important asset allocation information especially in periods of uncertainty.

In summary, this study demonstrates that approval prediction models can be constructed through the introduction of relevant prediction hypotheses and the application of empirical methods for prediction. I develop and adopt a robust modelling and testing framework, which allows for the development of a model that can predict patents that will result in a successful regulatory drug approval. This model is, perhaps, useful for key stakeholders such as investors and management who may want to gain a clearer understanding of the characteristics underlying successful patents or the likelihood that some patents will be linked to drug approvals in the future. Nonetheless, I find that there is evidence that an approval prediction model can help investors to consistently outperform the market in the long run. Overall, this thesis contributes to the literature in four main dimensions. Firstly, I extend the literature on the role of characteristics of successful patents. I also introduce a novel prediction modelling and testing framework. I provide a simple model, which can be used to ascribe approval probabilities to future U.S. patents, in research and practice. I provide some suggestions as to why using approval prediction as an investment strategy is likely to generate high abnormal returns for investors.

8.3 Implications and Recommendations

The empirical findings of the study have particular relevance for the theoretical framework of patent research and future empirical studies. They may assist firm-level R&D managers, who may have an interest in understanding the influence of the firm's research environment and the characteristics of the patents it produces. The study also provides guidance to policy makers who might benefit from an insight into how patenting operates in the U.S.

pharmaceutical industry. Finally, this study also provides investors with a potentially profit-making investment strategy.

The model (excluding forward citations) may be useful for key stakeholders such as regulators and management. It allows them to more fully understand the quality of the outputs of the drug discovery process, provides information on the motivations underlying project selection or assists in understanding why some firms will be more successful in obtaining an FDA drug approval in the future.

8.3.1 Implications for Future Empirical Research

This study extends the literature on the identification of quality/valuable patents and their relationship to successful commercialisation. It provides a simple model which can be used to assign approval probabilities to FDA drug candidates in future research and practice.

The study finds that the characteristics of approval-linked patents are different from patents with no approval linkage. For example, using a combination of patent-level characteristics, including the number of inventors, the number of independent claims, citations to NPR and citations received have greater power in explaining commercialisation success in comparison to simple counts of citations received. Studies conducted on firm-level innovation, should therefore recognise these differences and treat potential approval-linked patents as a distinctive patent group, with identifiable success-related patent-level characteristics.

Whilst prior literature has found patent claims to be a significant indicator of patent value. This study finds that it is a patent's independent patent claims that are the most important determinant of approval success, and so potential value, at the patent publication date. Future studies would therefore benefit by including measures of patent independent claims in their models to capture the determinants of patent value.

8.3.2 Implications/Recommendations for Firms

Drug developers are continuously involved in a decision-making process regarding the formulation of their portfolios of drugs candidates and whether during the drug development process to abandon or proceed with further development. I have developed a relatively simple, easy-to-apply predictive modelling tool to assist those drug development decisions. Firms can use the influential patent characteristics identified in this study to draw more

accurate inferences regarding their own and also their competitors' patents and future drug candidates. In so doing, R&D managers can gain a better understanding of how to achieve a successful drug approval outcome.

This model whilst informative should not mean that decisions be solely based on predicted probabilities. Decisions concerning drug development should be based on these findings, traditional clinical success rate metrics and specific considerations linked to the drug candidate being investigated. Managers must make the final decision as whether to proceed with a research project but can do so in a more informed manner by using this model. Assessing the probability of success using this method may avoid the unnecessary testing costs associated with failed drug candidates, assisting in more efficient resource allocation and potentially limiting risk-taking activity that consumes potentially scarce R&D resources. The economic impact from avoiding unnecessary drug trial costs associated with drug candidate failures may be substantial (See, for example, DiMasi, Kim and Getz (2014)).

An advantage of a model-based decision is that it can assist drug developers in assigning probabilities that are more appropriate to drug candidates that are based on stronger patents. Firms should focus time and resources on these candidates.

8.3.3 Implications/Recommendations for Policymakers

U.S. regulators and policymakers should encourage patent examiners operating in the USPTO to identify pharmaceutical patents that exhibit the characteristics that this study identifies as having a higher likelihood of drug approval success. Once these patents have been identified, a mechanism should be put in place to process these patents more quickly than those patents that do not possess these characteristics. This could fast track patent approval, potentially shortening product development times and the initial market availability for the drug candidates based on these patents.

When firms submit an Investigational New Drug (IND)[79] application to the FDA, in order to commence clinical human trial testing, the FDA should consider prioritising INDs that are based on patents that possess the approval success-related characteristics identified in

[79] After pre-clinical testing is completed a company then files an IND with the FDA to commence testing the drug candidate on human subjects (See glossary of terms in Appendix 2 for further detail).

this study. Again, this could potentially shorten product development times and bring forward the initial market availability for these drug candidates.

8.3.4 Implications/Recommendations for Investors

This study provides suggestions as to why drug approval prediction, as an investment strategy, is likely to generate high abnormal returns. Investors are continually searching for profitable investment opportunities. The findings of this study provide investors with an investment tool that allows the construction of potentially profitable stock portfolios. These portfolios do not require expert knowledge and neither do they involve large costs in terms of time or data gathering.

8.4 Limitations of the Study

This thesis contains several important findings, but it may also contain several limitations, which need to be acknowledged. This section briefly covers some of the potential limitations of my research and their implications for this thesis.

In the pharmaceutical industry, a small number of patents cover a product. In other industries, a large number of patents cover a product. In these industries patent portfolios, rather than individual patents, play a key role. This might reflect on the extent to which my conclusions can be generalised to other industries.

This study investigates a range of patent-level characteristics. The selection of these characteristics was determined by theory, along with time and data-processing limitations. This selection may have excluded certain patent characteristics that may influence the approval likelihood. For example, in the initial stage of the drug discovery process pharmaceutical firms select the disease their research project will target and any future patents will be linked to. When making this disease target choice, selecting a non-communicable, lifestyle related condition such as Type 2 Diabetes, may increase the likelihood of approval and higher future revenues. For approval-linked patents this information is readily available in the FDA approval letter. However, for those patents with no drug approval-linkage this information is not so easy to collect. This disease target information may not be discussed in the patent and even if it is, it is discussed in complex, scientific terminology in the descriptive section of the patent document (see patent 5346915 in Appendix 5). As this information is contained in a section of the patent document that I do not analyse it could potentially be a factor that influences the likelihood that a drug is

approved. However, as this data requires a large amount of processing time, this proposition could not be confirmed within the time constraints of this study.

This research has included publicly listed firms in the U.S.. Whilst these firms vary in size and experience differing levels of trading volume, they are all regularly traded stocks. Trading volume and share price volatility have normally been found to be positively correlated in empirical studies. Barber and Odean (2007) found a strong positive correlation between trading volume and the attention paid to a stock. There are several views as to why this occurs but most of the research in this area examines the flow of news to the market. Oh (2017) found that information linked to smaller firms, with lower levels of analyst coverage, is transmitted more slowly than firm-specific information for larger firms. In addition, according to Cohen and Lou (2012), larger firms tend to attract more investor attention, so reducing the potential for stock mis-valuation. For this study, low trading volumes were not examined as the focus is on NYSE/NASDAQ firms, which have the largest market capitalisation and therefore reduces concerns over low trading volumes and share price volatility.

This research has excluded patents that are not created by the firm that submits the NDA. This may have introduced some bias into the sample selection process. This may affect the extent to which my conclusions can be applied to firms whose research strategy is to license or acquire patents from other organisations rather than creating drug candidates themselves.

Citation-received variables may be influenced by other non-patent related factors. The results of clinical trial outcomes may positively influence citation numbers. However, clinical trials often occur several years after patent applications are filed so the application of a five-year time window measure for citations received may limit any influence from trial results.

Some patent-level characteristics may be influenced by factors endogenous to the firm. The patent-holder may possess confidential information related to the drug candidate covered by the patent or information obtained from the USPTO during the patent application process. The drafting of the patent document may be influenced by these unknown factors.

The analysis of portfolio returns in this study (as well as prior studies) ignores the transaction costs involved in investing in portfolios of firms holding predicted approval-linked patents. Given that the portfolio-holding period is rather long, nearly three years for at least 95% of

the sample, the inclusion of transaction costs is unlikely to significantly alter the conclusions. Given that the model generated small positive returns, future studies may wish to investigate how these returns are affected when transaction costs are considered.

When calculating portfolio returns I controlled for the presence of confounding events within the holding period, as firm-specific events may have a significant effect on the portfolio returns. Controlling for all such events is not without difficulty, as the majority of holding periods are over 700 trading days. Whilst it was not possible to control for all confounding events, I believe that the events I excluded are those of most importance.

Finally, this study further used past data to investigate the profitability of the prediction-based portfolio as an investment strategy. A strategy that provides profits using past data does not imply that such an approach can generate abnormal profits in the future. Changes in pharmaceutical industry patenting behaviour or changes in market conditions may reduce or eliminate the profitability of this proposed investment strategy. However, I have used out-of-sample analysis and robustness tests employing different sub-periods and results remain strong, increasing the likelihood that this investment strategy would be successful in the future.

I investigated the age of patents linked to drugs approved up to 31st December 2014 listed in the FDA Orange Book, Merck Index or Harvard FDA Project dataset. I noted that for FDA drug approvals approved in December 2014 (the latest in the sample) the oldest patents linked to these approvals were granted in 2003. In my methodology, I therefore assumed that patents published prior to 1st January 2003 were not linked to an FDA approved drug if they were not listed in any of these three sources. There may the possibility that a small number of patents that I assumed were not linked to an approved drug, do if fact possess such a linkage. However, this could not be confirmed as the necessary data were not available.

My research findings need to be interpreted considering the limitations outlined above. However, the questions they raise can provide potential avenues for further research.

8.5 Suggestions for Further Research

This study is one of the few studies that examine the potential predictive properties of certain patent characteristics and firm publishing behaviour. In so doing it can suggest areas of future opportunities.

The FDA controls drug approval in the U.S. The drug approval process studied in this research is linked to discoveries created by scientific research, driven by the profit and product pipeline concerns of pharmaceutical firms. The influence of consumer demand is not considered in this study. Investigating the potential influences of market demand factors may provide additional insight that may improve predictive performance.

Future research could investigate whether the findings of this study are applicable to pharmaceutical patents issued by the European Patent Office, which has differing patent requirements, and drug approvals granted by both the U.S. FDA and the European Medicines Agency.

These research findings could also be applied to test the relationship between USPTO patent characteristics and product commercialisation outcomes in other high-technology industries.

Although my study found that firm publishing behaviour had a negative influence on the likelihood of drug approval, the firm's publications were not examined in detailed, only recording the nature of co-authoring. With more time available, it may prove informative to examine composition of a firm's publication output, to refine the publication variables utilised for predictive purposes. It may prove informative to determine the proportion of publications published in highly ranked journals.

I exclude firms whose research strategy is to license or acquire patents rather than creating their own inventions. The small but negative influence of publishing behaviour on the likelihood of drug approval may encourage the extension of a future study sample to include these firms. This would assist in confirming that it is patent characteristics and not the firm-publishing environment that are the largest influence on approval likelihood.

My study uses a small amount of the data that are potentially available within the patent document. Prediction performance may be improved by utilising some of the other information held within patent documents. As discussed in section 8.4 identifying the disease target the patent is linked to may be a potential influence on FDA approval and the future revenue potential of the drug. With the increasing use of machine learning techniques for pattern recognition purposes, it may prove informative to apply these techniques to identify other useful predictor variables within these documents. For example, the Background of the Invention section of a patent document specifies the disease area the patent is linked to. Patent number 5346915 in Appendix 5, specifies that the invention the patent covers is linked

to “agents for treating dermatoses, such as acne”. This could then be used to classify a patent’s disease target into communicable or non-communicable, an additional data source that could be employed in the approval prediction process.

Appendices

Appendix 1 U.S. Drug Development Process

Basic Research

Long before a new drug can even be imagined, scientists are working to gain a basic understanding of a disease or of specific normal chemical pathways that are subverted in an abnormal cell. This research might be conducted in academic laboratories and research institutes around the world, and some of it is paid for by industry. Industry also plays a large role in the development of novel technologies, such as new approaches to sequencing of the human genome.

It is through better understanding of disease processes and pathways that targets for new treatments are identified. This might be a gene or protein instrumental to the disease process that a new treatment could interfere with, for example, by blocking an essential receptor. Biologists seek out and identify such targets that might be attacked with a new drug.

Once researchers have identified a target, they then validate them by determining whether the target is relevant to the disease that they are studying. Researchers will then search for a molecule or compound that acts on this target. Historically, researchers have looked to natural compounds from plants, fungi or marine animals to provide the basis for these candidate drugs but with scientific advances, scientists are using knowledge gained from the study of genetics and proteins to create new molecules using computers. As many as 10,000 compounds may be considered and whittled down to just 10 to 20 that could theoretically interfere with the disease process.

Based on an assessment of their properties researchers must then determine if a drug could affect the target enough to alter the course of the disease. Compounds are then assigned probabilities of success. For a particular biological mechanism and indication, one compound is selected as the lead compound. To do this, they use biochemical, cellular, or animal models to validate the biological mechanism of the target gene or protein. Other compounds that target the same mechanism—usually with lower probabilities of success—are designated back-up compounds. Back-up compounds are selected precisely because of their correlation with the lead compound with respect to a

particular biological mechanism. If the lead compound exhibits severe side effects, for example, the back-up compound may not, while still being effective against the target.

Searching for Compounds

When a potentially relevant target for an identified disease is identified, researchers then search for a molecule or compound that might modify the target or targets. To optimise the molecules being investigated, scientists use computers to model the structure of the lead compounds and how they link to the target protein. They screen vast compound libraries to develop a list of potential chemicals that might someday become a new medicine.

The molecules/compounds identified through this kind of screening can provide powerful research tools that help provide a better understanding of biological processes. This, in turn, may lead to new targets for potential drug discoveries.

Lead optimisation produces a drug candidate that has promising biological and chemical properties for the treatment of a disease.

A good target needs to be efficacious, safe, meet clinical and commercial needs and, above all, be “druggable”. A “druggable” target is accessible to the known drug molecule, be that a small molecule or larger biologicals and upon binding, elicit a biological response which may be measured both in vitro and in vivo.

Researchers select a target, such as a gene or protein, then select a molecule or compound, that may act on the target to alter the disease.

4 years: Average number of years taken to develop successful medicine

5,000 - 10,000 candidates: Number of medicinal candidates tested to achieve one approved medicine

Purpose: Assess safety, biological activity and formulations

Preclinical Studies: The Medicine

The drug candidate is then tested for its pharmacokinetic behaviour in animals, including its gastrointestinal absorption, body distribution, metabolism, and excretion. It is also tested for its pharmacodynamics, which refers to the relative effectiveness of the molecule.

Once a single compound is selected safety and efficacy tests are conducted using computerised models, cells and animals. These preclinical studies are performed to evaluate a drug's safety, efficacy, and potential toxicity in animal models. These studies are also designed to prove that a drug is not carcinogenic (i.e., it does not cause cancer when it is used at therapeutic doses, even over long treatment intervals), mutagenic (i.e., it does not cause genetic alterations), or teratogenic (i.e., it does not cause foetal malformations).

Preclinical studies with animals help determine the initial dose to be evaluated in the clinical trials and help identify safety evaluation criteria.

Around half of candidates make it through this pre-clinical testing stage. The five to 10 remaining compounds are now ready to be tested in humans for the first time. In the US, approval by the FDA is required before any testing in humans can occur. The company will put in a IND, which will be reviewed by medical and scientific experts, who will decide whether or not sufficient preliminary research has been conducted to allow testing in humans to go ahead.

Preclinical Trial Average Duration: 1 year

5 years: Average number of years (in total) taken to develop successful medicine

10-20 candidates: Number of medicinal candidates tested to achieve one approved medicine

Investigational New Drug Application (IND)

After preclinical testing is completed the company then files an IND with the FDA to begin testing the drug in people. The IND will become effective if the FDA does not

disprove it within 30 days. The IND will show results of previous experiments, how, where and by whom the new studies will be conducted. The IND also looks at the chemical structure of the compound, how it works in the body, and any toxic effects found in the animal studies. The IND will also look at how the compound is manufactured. The IND must be reviewed and approved by the Institutional Review Board where the study will be conducted, and progress reports on clinical trials must be submitted at least annually to the FDA.

Phase I Clinical Trials: Safety

Phase I trials are the first time that a drug is tested in humans. These trials may involve small numbers (20 to 100) of healthy volunteers, or they may include patients with specific conditions for which targeted pathways have been identified as potentially relevant to the disease under study. A phase I study may last for several months. The focus of a phase I study is the evaluation of a new drug's safety, the determination of a safe dosage range, the identification of side effects, and the detection of early evidence of effectiveness if the drug is studied in patients with disease, for example in patients with cancer. Wong and Siah (2018) calculated that 66.4% of compounds entering phase I trials successfully proceed to phase II testing.

Phase I Clinical Trial Average Duration: 1.5 years

6.5 years: Average number of years (in total) taken to develop successful medicine

5-10 candidates: Number of medicinal candidates tested to achieve one approved medicine

Purpose: Determine safety and dosage

Phase II Clinical Trials: Proof of Concept

In phase II clinical trials, the study drug is tested for the first time for its efficacy in 100 to 500 patients with the disease or the condition targeted by the medication. These studies may have up to several hundred patients and may last from several months to a few years. They help determine the correct dosage, common short-term side effects and the best regimen to be used in larger clinical trials. This usually begins with phase IIa clinical

trials, in which the goal is to obtain an initial proof of concept (POC). The phase IIb trials are larger and may use comparator agents and broader dosages to obtain a much more robust POC. Wong and Siah (2018) calculated that 58.3% of compounds entering phase II trials successfully proceed to phase III testing.

Phase II Clinical Trial Average Duration: 1.5 years

8 years: Average number of years (in total) taken to develop successful medicine

2-5 candidates: Number of medicinal candidates tested to achieve one approved medicine

Purpose: Evaluate effectiveness, look for side-effects

Phase III Clinical Trials: Regulatory Proof

Phase III clinical trials are designed to prove the candidate drug's benefit in a large targeted patient population with the disease. These trials confirm efficacy, monitor side effects, and sometimes compare the drug candidate to commonly used treatments. Researchers also use these clinical trials to collect additional information on the overall risk-benefit relationship of the drug and to provide an adequate basis for labelling after successful approval of the drug.

Phase III studies are conducted with large populations consisting of several hundred to several thousand patients (1,000 to 5,000) with the disease or the condition of interest. They typically take place over several years and at multiple clinical centres around the world. These studies provide the proof needed to satisfy regulators that the medicine meets the legal requirements needed to be approved for marketing. The study drug may be compared with existing treatments or a placebo. Phase III trials are, ideally, double blinded; that is, neither the patient nor the researcher knows which patients are receiving the drug and which patients are receiving placebos during the course of the trial. Phase III trials are usually required for FDA approval of the drug. If the trials are successful, then a New Drug Application is submitted to the FDA.

Wong and Siah (2018) calculated that 59% of compounds entering phase III trials successfully proceed to an NDA submission.

Phase III Clinical Trial Average Duration: 2.5 years

10.5 years: Average number of years (in total) taken to develop successful medicine

1-2 candidates: Number of medicinal candidates tested to achieve one approved medicine

Purpose: Confirm effectiveness, monitor adverse reactions from long-term use.

Licensing Approval

Information and results from all the studies is compiled and submitted to the regulatory agencies.

After all of the clinical trials are completed, then the company analyses the data and files an new drug application (NDA) with the FDA if the data successfully demonstrates safety and effectiveness. The NDA is usually about 100,000 pages or more and contains all of the scientific information that the company has gathered. By law, the FDA is allowed six months to review an NDA. In most cases the time from first submission of an NDA and final FDA approval usually exceeds six months. The process of review usually takes 10 to 12 months and may include an advisory committee review, but such a review is at the discretion of the FDA. The U.S. Government Accountability Office found that the FDA approved 85.3% NDA submissions.

Approval Review Average Duration: 1.5 years

12 years: Average number of years (in total) taken to develop successful medicine

1 medicine

Phase IV Clinical Trials: Marketing and Safety Monitoring

Phase IV trials are studies conducted after a drug receives regulatory approval from the FDA. They may be used primarily for medical marketing. In some cases, the FDA may require or companies may voluntarily undertake post-approval studies to generate additional information about a drug’s long-term safety and efficacy, including its risks, benefits, and optimal use. These studies may take a variety of forms, including studies

that use data from the administrative databases of health plans as well as observational studies and additional clinical trials.

Post-approval trials may also be designed to test the drug with additional patient populations (e.g., with children), in new delivery modes (e.g., as a timed-release capsule), or for new uses or indications (i.e., for the treatment of a different medical condition). Because these post-approval trials are intended to provide the basis for FDA approval of further uses or delivery modes, they must meet the same standards as the phase III trials conducted for initial approval

Patenting

Pharmaceutical companies will patent any molecule that shows promise early in the development process (Torjesen, 2015). Pharmaceutical companies typically file patents early on in the drug development process, often before the commencement of clinical trials. Patenting prevents other companies copying it for 20 years and covers many aspects of the intellectual property of a drug, including its manufacture, formulation and, in some cases, its use.

The purpose of a patent is to enable the pharmaceutical company that developed it to recoup their development costs and to make a profit to cover the development costs of drugs that failed during the testing process, as well as to invest in the development of future innovative drugs. By the time a drug has undergone the required testing and been licensed, half the patent period will usually have expired.

Appendix 2 Glossary of Terms

Assignee	The entity that is the recipient of a transfer of a patent application, patent, trademark application or trademark registration from its owner of record (assignor).[1] When the inventors are employed by a company they assign their invention to the company they are employed by at the time the invention was created.
Backward Citation	A backward citation is a citation of a previously published document (prior art) that was already available publicly before the new patent application was filed. These citations can relate to granted patents, patent applications or non-patent related material. The patent examiner ultimately decides which and how many citations are included in a patent.
Backward Non-patent Citation	These are references that are not related to patent documents. These include references to a various non-patent documents, such as scientific articles, technical papers, conference proceedings, etc. The knowledge contained within a non-patent reference is not usually as formalised or codified as the knowledge contained within patents. These references are listed in the Other Publications section at the beginning of the patent document, in alphabetical title order.
Backward Non-patent Citation Lag	This is the time elapsed (in days) between the filing year of patent application and the publication year of newest (youngest) non-patent reference this patent application cites.
Backward Patent Citation	All prior art that is patent-related must be referenced in a patent application. These patent references are listed in the References Cited section at the beginning of the patent document. U.S. patent documents are listed first, in patent number/publication date order, followed by foreign patent documents (again in patent number/publication date order).
Backward Patent Citation Lag	This is the time elapsed (in days) between the filing date of a patent application and the publication date of newest (youngest) patent reference this patent application cites.
Blockbuster Drug	Is defined to be an extremely popular drug, that generates annual sales of at least $1 billion.
Clinical Trial	A controlled experiment involving a defined set of human subjects, having a clinical event as an outcome measure, and intended to yield scientifically valid information about the efficacy or safety of a drug, vaccine, diagnostic test, surgical procedure, or other form of medical intervention.

Dependent Patent Claim	A claim that refers back to (depends on) and further limits a preceding dependent or independent claim. A dependent claim shall include every limitation of the claim from which it depends.[1] The description of each dependent claim is listed at the end of the patent document. These claims always contains a reference to another claim and includes the phrase "of claim" followed by the relevant claim number, e.g. "The Method of claim 10" (start of claim 11 of U.S. patent number 7741269, see Appendix 5).
Drug	A medicine or other substance which has a physiological effect when ingested or otherwise introduced into the body.
Forward Citation	For a particular patent, a forward citation is another (subsequent) patent document's citation back to that particular earlier patent (from the perspective of the earlier patent). Forward citations refer to the number of times a patent has been cited by subsequent patents, indicating that these newer patents are technologically built upon the cited (previously filed) patent.
Forward Citation Lag	This is the time elapsed (in days) between the publication date of the patent application and the filing date of the first forward citation it receives from a subsequent (later) patent application.
Independent Patent Claim	A claim that does not refer back to or depend on another claim.[1] The description of each independent claim is listed at the end of the patent document.
Invention	Any art or process (way of doing or making things), machine, manufacture, design, or composition of matter, or any new and useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States.[1]
Inventor	One who contributes to the conception of an invention. The patent law of the United States of America requires that the applicant in a patent application must be the inventor.[1] The names of the inventors are listed on the front page of the patent document.
Investigational New Drug Application (IND)	After preclinical testing is completed a company then files an IND with the U.S. Food and Drug Administration (FDA) to commence testing the drug candidate on human subjects. The IND includes results of previous experiments, how, where and by whom any new studies will be conducted. The IND also looks at the chemical structure of the compound being tested, how it works in the body and any toxic effects found in the animal studies. Unless the FDA places a hold on this application, a company may begin clinical testing 30 days after the filing of the IND.

New Drug Application (NDA)	For those drugs that make it to through phase III trials, a submission for marketing authorisations is made to the U.S. Food and Drug Administration (FDA). The submission contains preclinical and clinical information obtained during testing, including information about the chemical makeup and manufacturing process, pharmacology and toxicity of the compound, human pharmacokinetics, results of the clinical trials, and proposed labelling.
New Molecular Entity (NME)	A new molecular entity is an active ingredient that has never before been marketed in the United States in any form. It contains no active moiety that has been approved by the FDA in any other application submitted under section 505(b) of the Federal Food, Drug, and Cosmetic Act. Certain drugs are classified as new molecular entities (NME) for the purposes of the FDA review process. This review is carried out by FDA's Center for Drug Evaluation and Research (CDER). (Source: FDA)
Patent	A property right granted by the Government of the United States of America to an inventor "to exclude others from making, using, offering for sale, or selling the invention throughout the United States or importing the invention into the United States" for a limited time in exchange for public disclosure of the invention when the patent is granted.[1]
Patent Claims	They define the invention and are what aspects are legally enforceable. The specification must conclude with a claim particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention or discovery. The claim or claims must conform to the invention as set forth in the remainder of the specification and the terms and phrases used in the claims must find clear support or antecedent basis in the description so that the meaning of the terms in the claims may be ascertainable (clearly understood) by reference to the description. (See 37 CFR § 1.58(a)).[1] The total number of claims is listed at the end of the first page of the patent document. With the description of each claim listed at the end of the patent document.
Patent Classification	Patents are classified (organised) in the U.S. by a system using a 3 digit class and a 3 digit subclass to describe every similar grouping of patent art. A single invention may be described by multiple classification codes.[1] The patent classification codes are listed on the front page of the patent document.
Patent Examiner	In the United States Patent and Trademark Office (USPTO) examiners review patent applications to determine whether the invention(s) claimed in each of them should be granted a patent or whether the application should be refused. An examiner reviews patent applications to determine if they comply with

	basic format, rules and legal requirements. A claimed invention must meet patentability requirements of novelty, inventive step or non-obviousness, industrial application (or utility) and sufficiency of disclosure. One of the tasks of a patent examiner is to review the disclosure in the application and to compare it to the prior art. This involves reading and understanding a patent application, searching the prior art (including prior patent applications and patents, scientific literature databases, etc.) to determine what contribution the invention makes over the prior art. They also determine the scope of the invention claimed by the inventor.
Patent Scope	What is included - as in the scope of a claim.[1] The scope of a patent defines the extent of the legal coverage that a patent provides.
Patentable	Suitable to be patented; entitled by law to be protected by the issuance of a patent[1]. This means that the process, machine, composition of matter, or manufacture to be patented must be the first of its kind and cannot be of content that is already patented, published, or available on the market (See also 35 U.S. Code § 101 - Inventions patentable). An inventor's idea must be **novel** (i.e. at least some aspect of it must be new) in order to be eligible for a patent (See 35 U.S. Code § 102-Conditions for patentability; novelty). It must also be **non-obvious** and **useful**. If the invention covered by the patent application would be obvious to experts or the general public, then that invention cannot be patented (See 35 U.S. Code § 103-Conditions for patentability; non-obvious subject matter). An invention is "useful" if it provides some identifiable benefit and is capable of use.
Phase I Clinical Trial	In these trials the drug candidate is tested in people for the first time. Studies are conducted with about 20 to 100 healthy volunteers. The tests study a drug's safety profile, including the safe dosage range. The studies also determine how a drug is absorbed, distributed, metabolized and excreted, and the duration of its action.
Phase II Clinical Trial	In these trials researchers evaluate the drug candidate's efficacy in about 100 to 500 patients who have the condition the drug is intended to treat. To avoid unnecessarily exposing a volunteer to a potentially harmful substance, these studies use the fewest number of patients possible to provide sufficient statistical power to determine efficacy. The aim of phase II trials is to determine the most effective dose and method of delivery (for example, oral or intravenous), the appropriate dosing interval and to reconfirm product safety.
Phase III Clinical Trial	The aim of these phase II trials is to reconfirm the phase 2 findings in a larger population and to identify the best dosage

	regimen. Researchers study the drug candidate's in around 1,000 to 5,000 patients. Physicians monitor patients closely to determine the efficacy and identify any adverse reactions. In doing this the drug company needs to generate sufficient safety and efficacy data to demonstrate an overall risk-benefit for the drug to allow a submission to be made for a licensing application to the regulatory authority (FDA).
Pre-clinical Trial	These are early safety and efficacy tests are undertaken in computational models, cells and animals. No human subjects are involved.
Prescription Drug	A drug that can be obtained only by means of a physician's or dentist’s prescription.
Prior Art	Prior art, in most systems of patent law, is constituted by all information that has been made available to the public in any form before a given date that might be relevant to a patent's claims of originality. If an invention has been described in the prior art or would have been obvious over what has been described in the prior art, a patent on that invention is not valid.
Regulatory Drug Approval	A drug becomes approved for market when it receives regulatory approval from the U. S. Food and Drug Administration (FDA) An approval is issued at the end of a formalised review process. This review is carried out by FDA’s Center for Drug Evaluation and Research (CDER). (Source: FDA)
Utility Patent	May be granted to anyone who invents or discovers any new, useful, and nonobvious process, machine, article of manufacture, or composition of matter, or any new and useful improvement thereof.[1]

1 Source: https://www.uspto.gov/learning-and-resources/glossary

Appendix 3 Sample Firms

A. H. Robins Company Incorporated	Du Pont	Ortho Pharmaceutical Corporation
Abbott Laboratories	E. R. Squibb & Sons	OSI Pharmaceuticals
Adolor Corporation	Eli Lilly And Company	Parke Davis
Agouron Pharmaceuticals Inc.	Endo Pharmaceuticals Inc.	Pfizer
Air Reduction	Exelixis, Inc.	Pharmacia
Allergan, Inc.	G. D. Searle & Co.	Pharmacia & Upjohn
American Cyanamid	Geltex Pharmaceuticals Inc.	Pharmacyclics Inc.
American Home Products	Genentech, Inc.	Pharmos Corporation
American Hospital Supply Corporation	Genzyme Corporation	Praecis Pharmaceuticals, Inc.
Amylin Pharmaceuticals Inc.	Gilead Sciences	Procter & Gamble
Arena Pharmaceuticals Inc.	ICN Pharmaceuticals	QLT Phototherapeutics Inc
Ariad Pharmaceuticals Inc.	Icos Corporation	Richardson Merrell Inc
Avanir Pharmaceuticals	Incyte Corporation	Riker Laboratories
Biocryst Pharmaceuticals, Inc.	Isis Pharmaceuticals Inc.	Rowell Laboratories
Biogen Idec Inc.	Janssen	Schering Corporation
Biomarin Pharmaceutical Inc.	Ligand Pharmaceuticals Inc.	Schering-Plough Corporation
Bone Care International, Inc.	McNeil Lab	Scios Inc.
Bristol-Myers Squibb	Mead Johnson & Company	Sepracor Inc.
Carter-Wallace, Inc.	Medi-Physics, Inc.	Sterling Drug Inc.
Celgene Corporation	Merck & Co Inc	Sugen Inc.
Celtrix Pharmaceuticals, Inc.	Merrell Dow Pharmaceuticals Inc.	Syntex Corporation
Cephalon Inc.	Millennium Pharmaceuticals, Inc.	The Upjohn Company
Cor Therapeutics	Morton Norwich Products Inc	Theratechnologies, Inc.
Cubist Pharmaceuticals Inc.	Neurex Corporation	U.S. Bioscience, Inc.
CV Therapeutics Inc.	NPS Pharmaceuticals Inc.	Vertex Pharmaceuticals Incorporated
Cygnus Inc.	Onyx Pharmaceuticals	Warner-Lambert Company
Dow Pharmaceutical Sciences, Inc.	Orphan Medical Inc.	Wyeth

Appendix 4 Sample Patent Classes

Patent Class	Class Title
424	Drug, bio-affecting and body treating compositions
435	Chemistry: molecular biology and microbiology
436	Chemistry: analytical and immunological testing
514	Drug, bio-affecting and body treating compositions
530	Chemistry: natural resins or derivatives; peptides or proteins; lignins or reaction products thereof
534	Organic compounds -- part of the class 532-570 series
536	Organic compounds -- part of the class 532-570 series
540	Organic compounds -- part of the class 532-570 series
544	Organic compounds -- part of the class 532-570 series
546	Organic compounds -- part of the class 532-570 series
548	Organic compounds -- part of the class 532-570 series
549	Organic compounds -- part of the class 532-570 series
552	Organic compounds -- part of the class 532-570 series
558	Organic compounds -- part of the class 532-570 series
560	Organic compounds -- part of the class 532-570 series
562	Organic compounds -- part of the class 532-570 series
564	Organic compounds -- part of the class 532-570 series

Appendix 5 Sample Patent Documentation

U.S. Patent 5346915, Full Patent Document

Section 75 lists all of the researchers linked to the invention.

Section 45 specifies the grant date of the patent application.

US005346915A

United States Patent [19]

Chandraratna

[11] **Patent Number: 5,346,915**

[45] **Date of Patent: * Sep. 13, 1994**

[54] **CHROMANS AND THIOCHROMANS WITH PHENYLETHYNYL SUBSTITUENTS AT THE 7-POSITION HAVING RETINOLD-LIKE BIOLOGICAL ACTIVITY**

[75] Inventor: **Roshantha A. S. Chandraratna,** El Toro, Calif.

[73] Assignee: **Allergan, Inc.,** Irvine, Calif.

[*] Notice: The portion of the term of this patent subsequent to Dec. 25, 2007 has been disclaimed.

[21] Appl. No.: **1,009**

[22] Filed: **Jan. 6, 1993**

Related U.S. Application Data

[63] Continuation of Ser. No. 655,524, Feb. 13, 1991, abandoned.

[51] **Int. Cl.**[5] **A61K 31/38;** A61K 31/35; C07D 335/06; C07D 311/04

[52] **U.S. Cl.** **514/432;** 514/456; 549/23; 549/398

[58] **Field of Search** 549/23, 407, 398; 514/432, 456

Section 73 lists the company that initially owned the patent.

Section 22 lists the filing date of the patent application.

Section 52 lists all the 3 digit U.S. patent classes along with any related subclasses that the patent covers. In this example the first patent class listed is 514, subclass 432. I use the first 3 digit patent class listed, for this patent that would be 514.

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,096,341	6/1978	Frazer	560/85
4,326,055	4/1982	Loeliger	562/473
4,391,731	7/1983	Boller et al.	252/299.62
4,695,649	9/1987	Magami et al.	560/80
4,723,028	2/1988	Shudo	560/8
4,739,098	4/1988	Chandraratna	560/8
4,740,519	4/1988	Shroot et al.	514/443
4,810,804	3/1989	Chandraratna	549/23
4,826,969	5/1989	Maignan et al.	536/552
4,855,320	8/1989	Chatterjee et al.	514/473
4,895,868	1/1990	Chandraratna	549/23
4,980,369	12/1990	Chandraratna	514/432
4,992,468	2/1991	Chandraratna	514/532
5,006,550	4/1991	Chandraratna	549/405
5,013,744	5/1991	Chandraratna	514/435

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

0130795	1/1985	European Pat. Off.	C07D 311/58
176034A	4/1986	European Pat. Off.	C07C 63/66
0284288	9/1988	European Pat. Off.	C07D 401/04
0350846	7/1989	European Pat. Off.	C07D 311/58
3708060	9/1987	Fed. Rep. of Germany	C07D 311/04

Section 56 splits patent citations by geographical source. Firstly listing U.S. patents or patent applications with the next section listing all the foreign (non-U.S.) patents or patent applications a patent cites. In each section the patents/applications are listed by age with the oldest first.

For the U.S. documents the patent/application number, grant date, inventor(s) and leading U.S. patent class are detailed. The foreign documents the patent/application number, grant date, issuing patent office and leading International patent class are detailed.

The age of the newest patent/application (U.S. and Foreign) in this section is compared to the filing date of this patent. This gives the value for the variable patent citation lag length (PCL).

OTHER PUBLICATIONS

A. King et al A General Synthesis of Terminal and Internal Arylalkynes by the Palladium–Catalyzed Reaction of Alkynylzinc Reagents with Aryl Halides, *J. Org. Chem.* 43 No. 2, 1978 p. 358.

E. Negishi et al Conversion of Methyl Ketones into Terminal Acetylenes and (E)–Tri–substituted Olefins of Terpenoid Origin, *J. Org. Chem.* 45 No. 12, 1980 p. 2526.

(List continued on next page.)

This section lists all non-patent publications cited by the patent. Publications are listed by age, oldest first. Author(s), title, publication source and year are referenced.

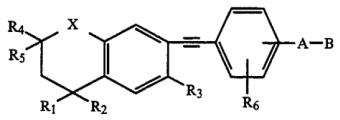
The total number of publications in this section gives the value for the variable non-patent references (NPR).

The age of the newest NPR in this section is compared to the filing date of this patent. This gives the value for the variable non-patent citation lag length (NCL).

Primary Examiner—Glennon H. Hollrah
Assistant Examiner—Mary C. Cebulak
Attorney, Agent, or Firm—Gabor L. Szekeres; Robert J. Baran; Martin A. Voet

[57] **ABSTRACT**

Novel compounds of the formula



where X is S, O; R_1–R_5 are hydrogen or lower alkyl; R_6 is lower alkyl, lower alkenyl, lower cycloalkyl having 1 to 6 carbons, or halogen; A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ where n is 0–5; and B is hydrogen, COOH or a pharmaceutically acceptable salt thereof, $COOR_8$, $COONR_9R_{10}$, $—CH_2OH$, CH_2OR_{11}, CH_2O-COR_{11}, CHO, $CH(OR_{12})_2$, $CHOR_{13}O$, —COR″, $CR''(OR_{12})_2$, or $CR''OR_{13}O$, where R″ is an alkyl, cycloalkyl or alkenyl group containing 1 to 5 carbons, R_8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R_8 is phenyl or lower alkylphenyl, R_9 and R_{10} independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl, R_{11} is lower alkyl, phenyl or lower alkylphenyl, R_{12} is lower alkyl, R_{13} is divalent alkyl radical of 2–5 carbons, have retinoic acid like activity.

30 Claims, No Drawings

This specifies the total number of claims the patent is making. It does not list claim details, these are listed at the end of the patent document. For this patent they are detailed from the foot of page columns 26 to 28 on pages 295 and 296.

U.S. PATENT DOCUMENTS

5,015,658	5/1991	Chandraratna	514/432
5,023,341	6/1991	Chandraratna	549/23
5,037,825	8/1991	Klaus et al.	514/233.8
5,045,551	9/1991	Chandraratna	514/337
5,053,523	10/1991	Chandraratna	549/398
5,068,252	11/1991	Chandraratna	514/543
5,089,509	2/1992	Chandraratna	514/332
5,130,335	7/1992	Chandraratna	560/100
5,134,159	7/1992	Chandraratna	514/456
5,162,546	11/1992	Chandraratna	549/23
5,175,185	12/1992	Chandraratna	514/445
5,183,827	2/1993	Chandraratna	514/444

OTHER PUBLICATIONS

Sporn et al. in *J. Amer. Acad. Derm.* 15:756–764 (1986).

S. Takahashi et al A Convenient Synthesis of Ethynylarenes and Diethynylarenes *Synthesis* 1980 p. 627–630.

Shudo et al. in *Chem. Phar. Bull.* 33:404–407 (1985).

Kagechika et al. in *J. Med. Chem.* 31:2182–2192 (1988).

M. Dawson et al., Chemistry and Biology of Synthetic Retinoids, published by CRC Press Inc., 1990, pp. 334–335, 354.

Meinc et al., Synthesis of 2,2′–Diacyl–1,1′–biaryls. Regiocontrolled Protection of . . . , *J. Org. Chem.,* No. 45, pp. 4720–4725, 1980.

L. Olson et al. A Dopamine Receptor Model and Its Application in the Design of a New Class of Rigid Pyrrolo[2,3–g]isoquinoline Antipsychotics, *American Chemical Society,* 1981, vol. 24, No. 9, pp. 1026–1031.

Williams et al, 6.2.3 Conformational Restriction, *Drug Discovery and Development,* 1987 The Humana Press, pp. 54–55.

Davis et al., V. Retinoid Structure–Biological Activity Relationships, Chemistry and Biology of Synthetic Retinoids, pp. 324–56 *J. Organometallic Chem* 387 (1990) 381–390.

CHROMANS AND THIOCHROMANS WITH PHENYLETHYNYL SUBSTITUENTS AT THE 7-POSITION HAVING RETINOLD-LIKE BIOLOGICAL ACTIVITY

CROSS-REFERENCE TO RELATED APPLICATION

This application is a continuation of application Ser. No. 07/655,524 filed on Feb. 13, 1991, now abandoned.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention is directed to novel compounds which have retinoic acid-like biological activity. More specifically, the present invention relates to compounds having an ethynyl benzoic acid portion and a second portion which is a 2-and/or 4-substituted thiochromanyl, or chromanyl group. The acid function may also be converted to an alcohol, aldehyde or ketone, or derivatives thereof, or may be reduced to —CH_3.

2. Related Art

European Patent Application 176034A (published Apr. 2, 1986) discloses tetrahydronaphtalene compounds having an ethynylbenzoic group. U.S. Pat. No. 4,739,098 discloses compounds wherein three olefinic units from the acid-containing moiety of retinoic acid are replaced by an ethynylphenyl functionality. These compound have retinoic acid-like biological activity.

U.S. Pat. No. 4,810,804 (issued on Mar. 7, 1989) based on an application of the same inventor and assigned to the same assignee as the present application, discloses such disubstituted acetylene compounds wherein one of the substituents of the acetylene (ethyne) group is a substituted phenyl group, and the second substituent is a substituted or unsubstituted 6-chromanyl, 6-thiochromanyl or 6-tetrahydroquinolinyl group. The compounds disclosed and claimed in U.S. Pat. No. 4,810,804 have retinoic acid-like biological activity.

Several co-pending applications of the present inventor, which applications are assigned to the assignee of the present application, are directed to further types of disubstituted acetylene compounds wherein one substituent of the acetylene (ethyne) moiety is a substituted phenyl or a substituted heteroaryl group, and the other substituent is a substituted or unsubstituted 6-chromanyl, 6-thiochromanyl or 6-tetrahydroquinolinyl group. The disubstituted acetylene compounds described and claimed in the aforesaid co-pending applications have significant retinoic acid-like activity.

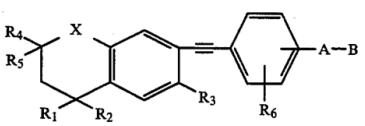
A published European patent application of the present applicant (Publication No. 0284288, published on Sep. 28, 1988) describes compounds having retinoic acid like activity which are 4,4 disubstituted chroman-6-yl, 4,4 disubstituted—thiochroman-6-yl acetylenes also substituted by a substituted heteroaryl group.

Retinoic acid-like activity has been generally recognized in the art to be associated with useful biological activity. Specifically, compounds having retinoic acid-like activity are useful as regulators of cell proliferation and differentiation, and particularly as agents for treating dermatoses, such as acne, Darier's disease, psoriasis, icthyosis, eczema, atopic dermatitis and epithelial cancers, for treating arthritic diseases and other immunological disorders (e.g. lupus erythematosus) for promoting wound healing, for treating dry eye syndrome and for reversing and preventing the effects of sun damage to skin.

With respect to the synthetic processes of the present invention which involve either the formation of an acetylenic (ethynyl) function in the compounds of the invention, or the coupling of the compounds of the invention which already have the ethynyl function with a halogen substituted phenyl group, the following articles comprise background information: A General Synthesis of Terminal and Internal Arylalkynes by the Palladium-Catalyzed Reaction of Alkynylzinc Reagents with Aryl Halides by Anthony O. King and Ei-ichi Negishi, *J. Org. Chem.* 43 1978 p 358; Conversion of Methyl Ketones into Terminal Acetylenes and (E)-Trisubstituted Olefins of Terpenoid Origin by Ei-ichi, Anthony O. King, and William L. Klima, *J. Org. Chem.* 45 1980 p.2526, and A Convenient Synthesis of Ethynylarenes and Diethynylarenes by S. Takahashi, Y. Kuroyama, K. Sonogashira, N. Hagihara, *Synthesis* 1980 p 627–630.

SUMMARY OF THE INVENTION

This invention covers compounds of Formula 1



Formula 1

wherein X is S, O; R_1–R_5 are hydrogen or lower alkyl; R_6 is lower alkyl, lower alkenyl, lower cycloalkyl having 1 to 6 carbons, or halogen; A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ where n is 0–5; B is hydrogen, COOH or a pharmaceutically acceptable salt thereof, $COOR_8$, $COONR_9R_{10}$, —CH_2OH, CH_2OR_{11}, CH_2O-COR_{11}, CHO, $CH(OR_{12})_2$, $CHOR_{13}O$, —COR″, CR″$(OR_{12})_2$, or CR″$OR_{13}O$, where R″ is an alkyl, cycloalkyl or alkenyl group containing 1 to 5 carbons, R_8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R_8 is phenyl or lower alkylphenyl, R_9 and R_{10} independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl, R_{11} is lower alkyl, phenyl or lower alkylphenyl, R_{12} is lower alkyl, R_{13} is divalent alkyl radical of 2–5 carbons.

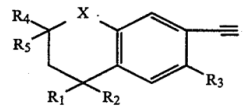
In a second aspect, this invention relates to the use of the compounds of Formula 1 for treating dermatoses, such as ache, Darier's disease, psoriasis, icthyosis, eczema, atopic dermatitis and epithelial cancers. These compounds are also useful in the treatment of arthritic diseases and other immunological disorders (e.g. lupus erythematosus), in promoting wound healing, in treating dry eye syndrome and in reversing the effects of sun damage to skin.

This invention also relates to a pharmaceutical formulation comprising a compound of Formula 1 in admixture with a pharmaceutically acceptable excipient.

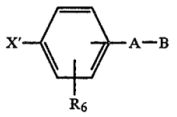
In another aspect, this invention relates to the process for making a compound of Formula 1 which process comprises reacting a compound of Formula 2 with a compound of Formula 3 in the presence of cuprous

iodide and $Pd(PQ_3)_2Cl_2$ (Q is phenyl) or a similar complex

Formula 2

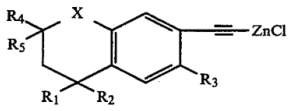


Formula 3



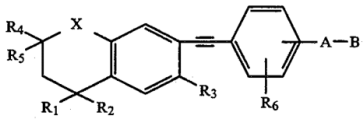
where R_1–R_6 are the same as described above, X' is a halogen, preferably I; and A is the same as defined above; and B is H, or a protected acid, alcohol, aldehyde or ketone, giving the corresponding compound of Formula 1; or to the process of making a compound of Formula 1 which consists of reacting a zinc salt of Formula 4 with a compound of Formula 3 in the presence of $Pd(PQ_3)_4$ (Q is phenyl) or a similar complex.

Formula 4



where R_1–R_6, and X, are the same as defined above, giving the corresponding compound of Formula 1; or homologating a compound of the Formula 5

Formula 5



where n is 0–4 to give an acid of Formula 1; or

converting an acid of Formula 1 to a salt; or

forming an acid addition salt;

converting an acid of Formula 1 to an ester; or

converting an acid of Formula 1 to an amide; or

reducing an acid of Formula 1 to an alcohol or aldehyde; or

converting an alcohol of Formula 1 to an ether or ester; or

oxidizing an alcohol of Formula 1 to an aldehyde; or

converting an aldehyde of Formula 1 to an acetal; or

converting a ketone of Formula 1 to a ketal.

GENERAL EMBODIMENTS

Definitions

The term "ester" as used here refers to and covers any compound falling within the definition of that term as classically used in organic chemistry. Where B (of Formula 1) is —COOH, this term covers the products derived from treatment of this function with alcohols, preferably with aliphatic alcohols having 1–6 carbons. Where the ester is derived from compounds where B is $—CH_2OH$, this term covers compounds of the formula $—CH_2OOCR$ where R is any substituted or unsubstituted aliphatic, aromatic or aliphatic-aromatic group, preferably with 1–6 carbons in the aliphatic portions.

Preferred esters are derived from the saturated aliphatic alcohols or acids of ten or fewer carbon atoms or the cyclic or saturated aliphatic cyclic alcohols and acids of 5 to 10 carbon atoms. Particularly preferred aliphatic esters are those derived from lower alkyl acids or alcohols. Here, and where ever else used, lower alkyl means having 1–6 carbon atoms and includes straight as well as branched chain alkyl groups. Also preferred are the phenyl or lower alkylphenyl esters.

Amide has the meaning classically accorded that term in organic chemistry. In this instance it includes the unsubstituted amides and all aliphatic and aromatic mono-and di-substituted amides. Preferred amides are the mono- and di-substituted amides derived from the saturated aliphatic radicals of ten or fewer carbon atoms or the cyclic or saturated aliphatic-cyclic radicals of 5 to 10 carbon atoms. Particularly preferred amides are those derived from lower alkyl amines. Also preferred are mono- and di-substituted amides derived from the phenyl or lower alkylphenyl amines. Unsubstituted amides are also preferred.

Acetals and ketals include the radicals of the formula —CK where K is $(—OR)_2$. Here, R is lower alkyl. Also, K may be $—OR_1O—$ where R_1 is lower alkyl of 2–5 carbon atoms, straight chain or branched.

A pharmaceutically acceptable salt may be prepared for any compound of this invention having a functionality capable of forming such salt, for example an acid or an amine functionality. A pharmaceutically acceptable salt may be any salt which retains the activity of the parent compound and does not impart any deleterious or untoward effect on the subject to which it is administered and in the context in which it is administered.

Such a salt may be derived from any organic or inorganic acid or base. The salt may be a mono or polyvalent ion. Of particular interest where the acid function is concerned are the inorganic ions, sodium, potassium, calcium, and magnesium. Organic amine salts may be made with amines, particularly ammonium salts such as mono-, di- and trialkyl amines or ethanol amines. Salts may also be formed with caffeine, tromethamine and similar molecules. Where there is a nitrogen sufficiently basic as to be capable of forming acid addition salts, such may be formed with any inorganic or organic acids or alkylating agent such as methyl iodide. Preferred salts are those formed with inorganic acids such as hydrochloric acid, sulfuric acid or phosphoric acid. Any of a number of simple organic acids such as mono-, di- or tri-acid may also be used.

The preferred compounds of this invention are those where the ethynyl group and the B group are attached to the 1 and 4 positions respectively of a benzene ring (i.e. where the phenyl moiety of the compound is para substituted) A is $(CH_2)_n$ and n is 0; and B is —COOH, an alkali metal salt or organic amine salt, or a lower alkyl ester thereof, or $—CH_2OH$ and the lower alkyl esters and ethers thereof, (formed with a lower alkanol) or —CHO and acetal derivatives thereof. The more preferred compounds shown in Formula 6 are:

ethyl 4-(4,4-dimethyl-7-thiochromanyl)-ethynyl-benzoate (Compound 1, X=S, R_3=H, R_4=H, R_5=H, R''=C_2H_5);

4-[(4-4-dimethyl-7-thiochromanyl)-ethynyl]-benzoic acid (Compound 2, X=S, R_3=H, R_4=H, R_5=H, R''=H);

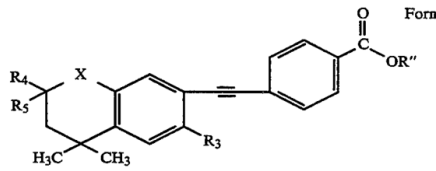
ethyl 4-(2,2,4,4-tetramethyl-7-thiochromanyl)-ethynylbenzoate (Compound 3, X=S, R_3=H, R_4=CH_3, R_5=CH_3, R"=C_2H_5);

ethyl-4-(2,2,4,4-tetramethyl-7-chromanyl)-ethynyl-benzoate (Compound 4, X=O, R_3=H, R_4=CH_3, R_5=CH_3, R"=C_2H_5).

4-(2,2,4,4-tetramethyl-7-thiochromanyl)-ethynyl benzoic acid (Compound 49, X=S, R_3=H, R_4=CH_3, R_5=CH_3, R"=H).

4-(2,2,4,4-tetramethyl-7-chromanyl)-ethynyl benzoic acid (Compound 50, X=O, R_3=H, R_4=CH_3, R_5=CH_3, R"=H).

Formula 6



The compounds of this invention may be administered systemically or topically, depending on such considerations as the condition to be treated, need for site-specific treatment, quantity of drug to be administered, and similar considerations.

In the treatment of dermatoses, it will generally be preferred to administer the drug topically, though in certain cases such as treatment of severe cystic acne, oral administration may also be used. Any common topical formulation such as a solution, suspension, gel, ointment, or salve and the like may be used. Preparation of such topical formulations are well described in the art of pharmaceutical formulations as exemplified, for example, *Remington's Pharmaceutical Science*, Edition 17, Mack Publishing Company, Easton, Pa. For topical application, these compounds could also be administered as a powder or spray, particularly in aerosol form.

If the drug is to be administered systemically, it may be confected as a powder, pill, tablet or the like, or as a syrup or elixir for oral administration. For intravenous or intraperitoneal administration, the compound will be prepared as a solution or suspension capable of being administered by injection. In certain cases, it may be useful to formulate these compounds in suppository form or as an extended release formulation for deposit under the skin or intermuscular injection.

Other medicaments can be added to such topical formulation for such secondary purposes as treating skin dryness, providing protection against light; other medications for treating dermatoses, preventing infection, reducing irritation, inflammation and the like.

Treatment of dermatoses or any other indications known or discovered to be susceptible to treatment by retinoic acid-like compounds will be effected by administration of the therapeutically effective dose of one or more compounds of the instant invention. A therapeutic concentration will be that concentration which effects reduction of the particular condition, or retards its expansion. In certain instances, the drug potentially could be used in a prophylactic manner to prevent onset of a particular condition. A given therapeutic concentration will vary from condition to condition and in certain instances may vary with the severity of the condition being treated and the patient's susceptibility to treatment. Accordingly, a given therapeutic concentration will be best determined at the time and place through routine experimentation. However, it is anticipated that in the treatment of, for example, acne, or other such dermatoses, that a formulation containing between 0.001 and 5 percent by weight, preferably about 0.01 to 1% will usually constitute a therapeutically effective concentration. If administered systemically, an amount between 0.01 and 100 mg per kg body weight per day, but preferably about 0.1 to 10 mg/kg, will effect a therapeutic result in most instances.

The retionic acid like activity of these compounds was confirmed through the classic measure of retionic acid activity involving the effects of retionic acid on ornithine decarboxylase. The original work on the correlation between retinoic acid and decrease in cell proliferation was done by Verma & Boutwell, *Cancer Research*, 1977, 37, 2196–2201. That reference discloses that ornithine decarboxylase (ODC) activity increased precedent to polyamine biosynthesis. It has been established elsewhere that increases in polyamine synthesis can be correlated or associated with cellular proliferation. Thus, if ODC activity could be inhibited, cell hyperproliferation could be modulated. Although all causes for ODC activity increase are unknown, it is known that 12-0-tetradecanoylphorbol-13-acetate (TPA) induces ODC activity. Retinoic acid inhibits this induction of ODC activity by TPA. The compounds of this invention also inhibit TPA induction of ODC as demonstrated by an assay essentially following the procedure set out in *Cancer Res.*, 35: 1662–1670, 1975.

By way of example of retinoic acid-like activity it is noted that in the assay conducted essentially in accordance with the method of Verma & Boutwell, ibid, the following examples of the preferred compounds of the present invention (Compounds 1, 2, 3 and 4) attained an 80% inhibition of TPA induced ODC activity at the following concentrations (IC_{80}):

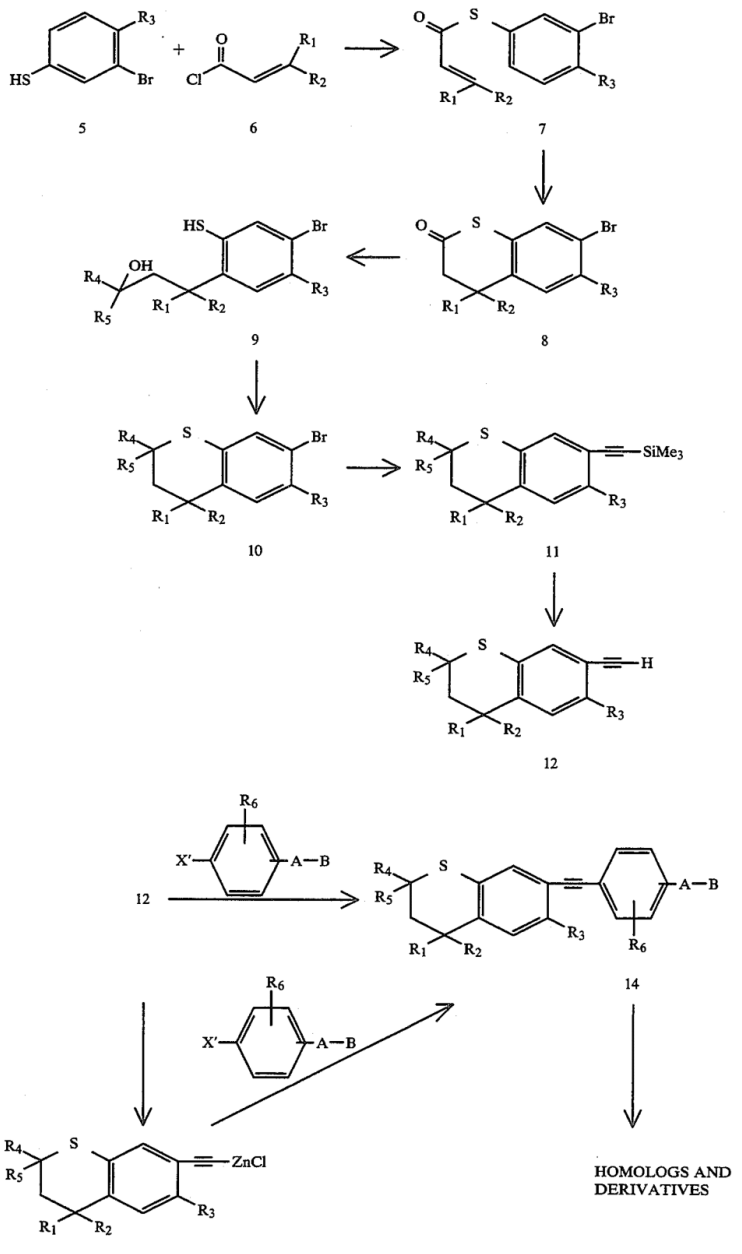
Compound	IC_{80} conc (nmols)
1	0.81
2	1.54
3	1.23
4	0.35

SPECIFIC EMBODIMENTS

The compounds of this invention can be made by a number of different synthetic chemical pathways. To illustrate this invention, there is here outlined a series of steps which have been proven to provide the compounds of Formula 1 when such synthesis is followed in fact and in spirit. The synthetic chemist will readily appreciate that the conditions set out here are specific embodiments which can be generalized to any and all of the compounds represented by Formula 1. Furthermore, the synthetic chemist will readily appreciate that the herein described synthetic steps may be varied and or adjusted by those skilled in the art without departing from the scope and spirit of the invention.

Compounds of Formula 1 where X is —S— and R_4 and R_5 are hydrogen or lower alkyl, are prepared as per Reaction Scheme 1

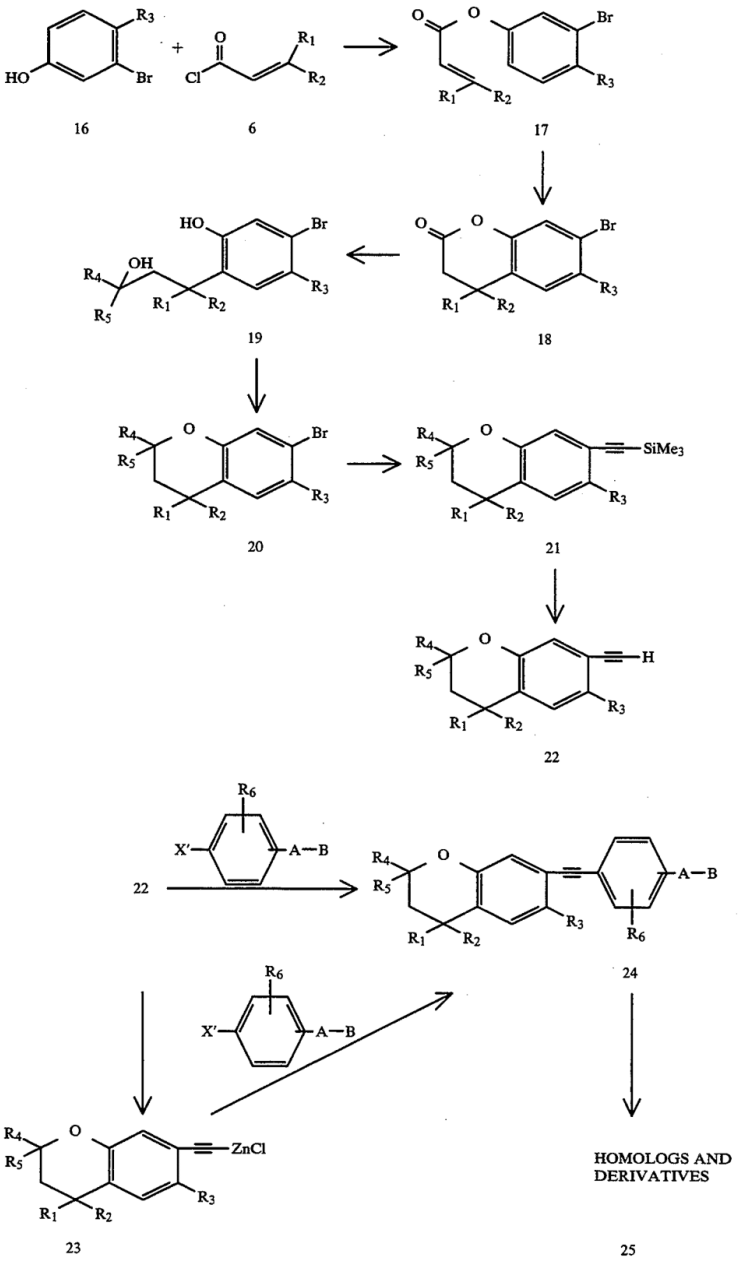
Reaction Scheme 1



In Reaction Scheme 1, R_1–R_5 are hydrogen or a lower alkyl group, R_6 is defined as above in connection with Formula 1, A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ where n is 0–5 and B is H, or a protected acid, alcohol, aldehyde or ketone. X' is Cl, Br or I when n is 0 but preferably is Br or I when n is 1–5.

Compounds of Formula 1 where X is oxygen and R_4 and R_5 are hydrogen or lower alkyl, are prepared as per Reaction Scheme 2.

Reaction Scheme 2



In Reaction Scheme 2 the definitions of R_1–R_6, A, B and X' are the same as in Reaction Scheme 1.

A general description of the synthetic steps outlined in Reaction Schemes 1 and 2 is as follows.

In Reaction Scheme 1 the 3-bromo-thiophenol (Compound 5) is acylated with an acylating agent, such as an acid chloride (Compound 6) derived from an appropriately substituted acrylic acid. The acylation is conducted in an inert solvent (such as tetrahydrofuran) in the presence of strong base (for example sodium hydride). The resulting thioester (Compound 7) which contains the olefinic bond of the acrylic acid moiety is ring closed in the presence of a Friedel Crafts type

catalyst (such as aluminum chloride) by stirring in a suitable solvent such as methylene chloride. The resulting 2-oxo-7-bromothiochroman (Compound 8) is usually isolated in crystalline form.

The R_4 and/or R_5 substituents are introduced by treating the 2-oxo-7-bromo-thiochroman (Compound 8) with a Grignard reagent in the presence of $CeCl_3$, bearing the alkyl substituents R_4 and R_5 (such as methylmagnesium bromide when R_4 and R_5 are methyl). When the Grignard reagent (such as methylmagnesium bromide) is in excess, the thiochroman ring is opened and the tertiary alcohol derivative of the 3-bromo thiophenol (Compound 9) is formed.

Ring closure of the thiophenol derivative (Compound 9) which has the desired R_1, R_2, R_3, R_4 and R_5 substituents, is affected by heating in acidic conditions, preferably by heating Compound 9 in aqueous acid. The resulting 7-bromothiochroman which bears the desired alkyl (or hydrogen) substituents, R_1, R_2, R_3, R_4 and R_5 is shown as Compound 10 in Reaction Scheme 1.

To introduce the acetylene (ethyne) portion into the molecule, the substituted 7-bromothiochroman (Compound 10) is reacted with trimethylsilylacetylene in the presence of cuprous iodide and a suitable catalyst, typically having the formula $Pd(PQ_3)_2Cl_2$ (Q is phenyl). The reaction is typically conducted in the presence of bis(triphenylphosphine) palladium (II) chloride catalyst and an acid acceptor (such as triethylamine) under an inert gas (argon) atmosphere, by heating in a sealed tube. The resulting 7-trimethylsilylethynylthiochroman is shown as Compound 11 in Reaction Scheme 1.

As is shown on Reaction Scheme 1, the trimethylsilyl moiety is removed from the 7-trimethylsilylethynyl-thiochroman (Compound 11) in the next synthetic step, to provide the ring substituted 7-ethynyl-thiochroman derivative (Compound 12). The latter reaction is conducted under basic conditions, preferably under an inert gas atmosphere.

In order to introduce the phenyl or substituted phenyl substituent on the acetylene (ethyne) portion of Compound 12, Compound 12 is coupled with the reagent X'—Q—A—B (Formula 3, Q is a di- or multi-substituted phenyl residue) where the symbols A, X' and B have the same meaning as defined in connection with Formula 3. In other words, the phenyl or substituted phenyl substituent is introduced into the 7-ethynyl-thiochroman (Compound 12) by reacting the latter with a halogen substituted phenyl compound (Formula 3) in which the benzene nucleus either has the desired substituent [A-B] or wherein the actual substituent A-B can be readily converted to the desired substituent by means of organic reactions well known in the art.

Coupling of the 7-ethynyl-thiochroman (Compound 12) with the reagent X'—Q—A—B is affected directly in the presence of cuprous iodide, a suitable catalyst, typically of the formula $Pd(PQ_3)_2Cl_2$ and an acid acceptor, such as triethylamine, by heating in a sealed tube under an inert gas (argon) atmosphere.

The resulting disubstituted acetylene compound (Compound 14) may be the target compound made in accordance with the invention, or maybe readily converted into the target compound by such steps as salt formation, esterification, deesterification, homologation, amide formation and the like. These steps are further discussed below.

Compound 14 may also be obtained by first converting the 7-ethynyl-thiochroman derivative (Compound 12) into a corresponding metal salt, such as a zinc salt, (Compound 13) and thereafter coupling the salt (Compound 13) with the reagent X'—Q—A—B (Formula 3 Q is phenyl or substituted phenyl residue) in the presence of a catalyst having the formula $Pd(PQ_3)_4$ (Q is phenyl), or similar complex.

Derivatization of Compound 14 is indicated in Reaction Scheme 1 as conversion to "homologs and derivatives" Compounds 15.

More specifically with respect to either derivatization or deblocking of protected functionalities in Compound 14, or with respect to the preparation of phenyl derivatives of the formula X'—Q—A—B, (which after coupling either directly yield the compounds of the invention, or are readily converted into them) the following is noted.

Where a protected phenyl derivative is needed to couple with the compounds of Formula 2 (Compounds 12 in Reaction Scheme 1), such may be prepared from their corresponding acids, alcohols, ketones or aldehydes. These starting materials, the protected acids, alcohols, aldehydes or ketones, are all available from chemical manufacturers or can be prepared by published methods. Carboxylic acids are typically esterified by refluxing the acid in a solution of the appropriate alcohol in the presence of an acid catalyst such as hydrogen chloride or thionyl chloride. Alternatively, the carboxylic acid can be condensed with the appropriate alcohol in the presence of dicyclohexylcarbodiimide and dimethylaminopyridine. The ester is recovered and purified by conventional means. Acetals and ketals are readily made by the method described in March, "Advanced Organic Chemistry," 2nd Edition , McGraw-Hill Book Company, p 810). Alcohols, aldehydes and ketones all may be protected by forming respectively, ethers and esters, acetals or ketals by known methods such as those described in McOmie, Plenum Publishing Press, 1973 and *Protecting Groups*, Ed. Greene, John Wiley & Sons, 1981.

To increase the value of n before effecting a coupling reaction, where such compounds are not available from a commercial source, the phenyl derivatives where B is —COOH are subjected to homologation by successive treatment under Arndt-Eistert conditions or other homologation procedures. Alternatively, phenyl derivatives where B is different from COOH, may also be homologated by appropriate procedures. The homologated acids can then be esterified by the general procedure outlined in the preceding paragraph.

An alternative means for making compounds where A is $(CH_2)_n$n is 1–5 is to subject the compounds of Formula 1, where B is an acid or other function, to homologation, using the Arndt-Eistert method referred to above, or other homologation procedures.

Compounds of Formula 1 where A is an alkenyl group having one or more double bonds can be made for example, by having the requisite number of double bonds incorporated into the intermediate of Formula 3; that is by using for the intermediate of Formula 3 an unsaturated aromatic compound bearing the X' leaving group (preferably halogen) in the phenyl nucleus. 4-Bromo or 4-iodo-cinnamic acid ethyl ester serve as concrete examples. Generally speaking, the compounds of Formula 3 where A is an unsaturated carbon chain can be obtained by synthetic schemes well known to the practicing organic chemist; for example by Wittig and like reactions, or by introduction of a double bond by elimination of halogen from an alpha-halo-phenylalkyl-carboxylic acid, ester or like carboxaldehyde. Com-

pound of Formula 1 where the A group has a triple (acetylenic) bond can be made by using the corresponding intermediate of Formula 3. Such intermediate can be obtained by reactions well known in the art, for example, by reaction of a corresponding phenyl-methyl ketone with strong base, such as lithium diisopuopyl amide.

The acids and salts derived from Formula 1 are readily obtainable from the corresponding esters. Basic saponification with an alkali metal base will provide the acid. For example, an ester of Formula 1 may be dissolved in a polar solvent such as an alkanol, preferably under an inert atmosphere at room temperature, with about a three molar excess of base, for example, potassium hydroxide. The solution is stirred for an extended period of time, between 15 and 20 hours, cooled, acidified and the hydrolysate recovered by conventional means.

The amide may be formed by any appropriate amidation means known in the art from the corresponding esters or carboxylic acids. One way to prepare such compounds is to convert an acid to an acid chloride and then treat that compound with ammonium hydroxide or an appropriate amine. For example, the acid is treated with an alcoholic base solution such as ethanolic KOH (in approximately a 10% molar excess) at room temperature for about 30 minutes. The solvent is removed and the residue taken up in an organic solvent such as diethyl ether, treated with a dialkyl formamide and then a 10-fold excess of oxalyl chloride. This is all effected at a moderately reduced temperature between about −10 degrees and +10 degrees C. The last mentioned solution is then stirred at the reduced temperature for 1–4 hours, preferably 2 hours. Solvent removal provides a residue which is taken up in an inert inorganic solvent such as benzene, cooled to about 0 degrees C. and treated with concentrated ammonium hydroxide. The resulting mixture is stirred at a reduced temperature for 1–4 hours. The product is recovered by conventional means.

Alcohols are made by converting the corresponding acids to the acid chloride with thionyl chloride or other means (J. March "Advanced Organic Chemistry" 2nd Edition, McGraw-Hill Book Company), then reducing the acid chloride with sodium borohydride (March, Ibid, pg. 1124), which gives the corresponding alcohols. Alternatively, esters may be reduced with lithium aluminum hydride at reduced temperatures. Alkylating these alcohols with appropriate alky halides under Williamson reaction conditions (March, Ibid, pg. 357) gives the corresponding ethers. These alcohols can be converted to esters by reacting them with appropriate acids in the presence of acid catalysts or dicyclohexlcarbodiimide and dimethlaminopyridine.

Aldehydes can be prepared from the corresponding primary alcohols using mild oxidizing agents such as pyridinium dichromate in methylene chloride (Corey, E. J., Schmidt, G., *Tet. Lett.*, 399, 1979), or dimethyl sulfoxide/oxalyl chloride in methylene chloride (Omura, K., Swern, D., *Tetrahedron,* 1978, 34, 1651).

Ketones can be prepared from an appropriate aldehyde by treating the aldehyde with an alkyl Grignard reagent or similar reagent followed by oxidation.

Acetals or ketals can be prepared from the corresponding aldehyde or ketone by the method described in March, Ibid, p 810.

Compounds where B is H can be prepared from the corresponding halogenated benzene compounds, preferably where the halogen is I.

With reference to Reaction Scheme 2, 3-bromophenol, or a 3-bromophenol substituted in the 4-(para) position by an alkyl substituent (R_3) (Compound 16) is acylated with an acylating agent, such as an acid chloride (Compound 6) derived from an appropriately substituted acrylic acid. In Reaction Scheme 2, just as in Reaction Scheme 1, the R_1 and R_2 substituents of the target compounds are introduced through this acrylic acid derivative (Compound 6). The acylation with the acid chloride (Compound 6) is preferably conducted in the presence of a strong base (e.g. sodium hydride) in an inert solvent (such as tetrahydrofuran). The resulting substituted phenyl-acrylate is shown in Reaction Scheme 2 as Compound 17.

The substituted phenyl-acrylate 17 is ring closed under Friedel Crafts type reaction conditions ($AlCl_3$ catalyst, in an inert solvent, such as methylene chloride) to provide the 2-oxo-7-bromo-chroman compound (Compound 18) which bears, in the 4-position, the R_1 and R_2 substituents and in the 6-position the R_3 substituent (as applicable). Just like the analogous 2-oxo-thiochroman (Compound 8) in Reaction Scheme 1, the 2-oxo-chroman 18 of Reaction Scheme 2 is treated with a Grignard reagent to introduce the R_4 and R_5 substituents. When R_4 and R_5 are methyl, the Grignard reagent is preferably methylmagnesium chloride (dissolved in tetrahydrofuran, THF). A solution of Compound 18 in a suitable solvent, for example in dry diethylether is added to this Grignard reagent. The resulting phenol containing a tertiary alcohol side chain, (that is a molecule in which the chroman ring had been opened) is shown in Reaction Scheme 2 as Compound 19.

Compound 19 which already has the desired R_1, R_2, R_3, R_4 and R_5 substituents, is ring closed under acidic conditions, (e.g. by heating in aqueous sulfuric acid) to provide the chroman derivative (Compound 20). To introduce the acetylene (ethyne) portion into the molecule, the substituted 7-bromo chroman (Compound 20) is reacted with trimethylsilyl acetylene in the presence of cuprous iodide and a suitable catalyst, typically having the formula $Pd(PQ_3)_2Cl_2$ (Q is phenyl), as defined for the 7-bromo-thiochroman compound in Reaction Scheme 1. The resulting 7-trimethylsilyl-ethynyl-chroman is shown as Compound 21 in Reaction Scheme 2.

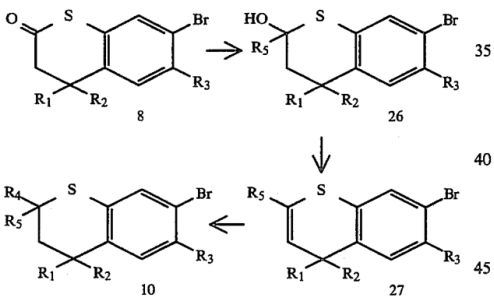
In Reaction Scheme 2, just as in Reaction Scheme 1, the trimethylsilyl moiety is removed from the 7-trimethylsilylethynyl-chroman (Compound 21) under basic conditions, to provide the ring substituted 7-ethynyl-chroman derivative (Compound 22).

Referring still to Reaction Scheme 2, the 7-ethynyl-chroman derivative (Compound 22) may be converted into the target compounds of the invention in synthetic steps which are analogous to the conversion of 7-ethynyl-thiochromans (Compound 12) into the corresponding target thiochroman derivatives (See Reaction Scheme 1). Briefly, Compound 22 is preferably heated with a reagent X'—Q—A—B (Formula 3, Q is phenyl or substituted phenyl residue) in the presence of cuprous iodide, a suitable catalyst, typically of the formula $Pd(PQ_3)_2Cl_2$ (Q is phenyl or the like) and an acid acceptor, such as triethylamine. This coupling reaction, yields the target chroman compounds, (Compound 24) or such derivatives which are readily converted into the target compounds by protection, deprotection, esterification, homologation etc., as is discussed in connection with

Reaction Scheme 1. The homologs are indicated, as a group, as Compound 25 in Reaction Scheme 2.

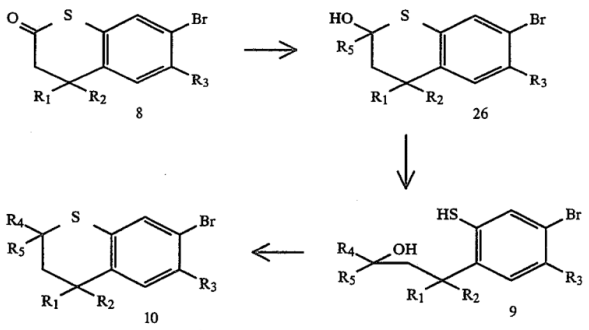
Alternatively, the 7-ethynyl-chroman compounds (Compound 22) may first be converted to the corresponding metal (zinc) salt (Compound 23) and thereafter coupled with the reagent X'—Q—A—B (Formula 3, Q is phenyl or substituted phenyl residue) under conditions which are similar to the conditions described in Reaction Scheme 1 for coupling of Compounds 13 with the same reagent.

Reaction Scheme 3



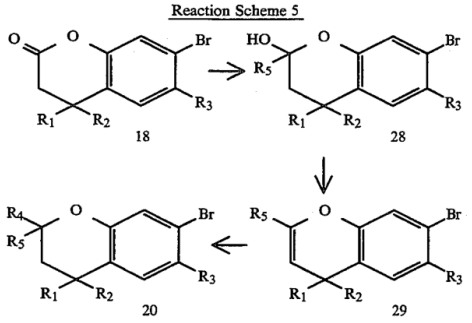
Referring to Reaction Scheme 3, the substituted 7-bromothiochroman (Compound 10), where one of the R_4 or R_5 substituents is alkyl and the other is hydrogen, can be made by treating the 2-oxo-7-bromo-thiochroman (Compound 8) with a Grignard reagent. As in Reaction Scheme 1 the 2-oxo-thiochroman (Compound 8) is subjected to an excess of Grignard reagent, bearing the alkyl substituents R_4 or R_5 (such as methylmagnesium bromide when R_4 or R_5 is methyl). However, the reaction temperature is controlled and maintained at a relatively low temperature (such as −14 degrees C.) and the duration of the reaction is kept relatively short (0.5 hours). A hemiacetal derivative of 3-bromothiophenol (Compound 26) is formed in this controlled Grignard reaction, as shown in Reaction Scheme 3. Compound 26 is converted by heating in acidic conditions, preferably with aqueous acid, to the unsaturated derivative (Compound 27). Compound 27 is reduced by hydrogenation in the presence of palladium sulfide-on-carbon catalyst at increased pressure (approximately 30 psi). The resulting 7-bromo-thiochroman which bears the desired hydrogen and alkyl substituents R_1, R_2, R_3, R_4 and R_5 with one of R_4 or R_5 being hydrogen, is shown as Compound 10.

Reaction Scheme 4



To obtain the 7-bromothio chroman (Compound 10), (Reaction Scheme 4) where the R_4 and R_5 substituents both are alkyl but not identical with one another, the hemiacetal derivative (Compound 26) is treated with a different Grignard reagent than previously used, as shown in Scheme 4. In this Grignard reaction the thiochroman ring is opened and the tertiary alcohol derivative of 3-bromo-thiophenol (Compound 9), is formed. Ring closure of the thiophenol derivative (Compound 9) which has the desired R_1, R_2, R_3, R_4 and R_5 substituents, is affected by heating in acidic conditions, preferably by heating with aqueous acid. The resulting 7-bromo thiochroman which bears the desired alkyl and hydrogen substituents R_1, R_2, R_3, R_4 and R_5 is shown as Compound 10.

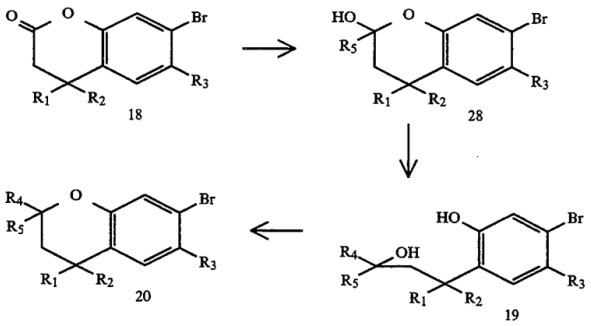
Reaction Scheme 5



In Reaction Scheme 5, just as in Reaction Scheme 3, one of the R_4 or R_5 substituents is alkyl and the other is hydrogen. Just like the analogous 2-oxo-thiochroman (Compound 8) in Reaction Scheme 3, the 2-oxochroman (Compound 18) of Reaction Scheme 5 is treated with a Grignard reagent to introduce the R_4 and R_5 substituents. With controlled reaction temperature and time, the resulting hemiacetal derivative can be isolated as Compound 28, as shown in Reaction Scheme 5. Under acidic conditions, (e.g. by heating in aqueous acid) the hemiacetal (Compound 28) is cyclized to form the corresponding unsaturated derivative (Compound 29). The unsaturated derivative can then be reduced using the same conditions as described in connection with Reaction Scheme 3 for the reduction of Com-

pound 26, or by a more general reducing procedure. The resulting chroman derivative is shown as Compound 20 in Reaction Scheme 5.

Reaction Scheme 6



Referring to Reaction Scheme 6, in Compound 20 of that scheme the R_4 and R_5 substituents are alkyl but are not identical. The R_4 and R_5 alkyl substituents are introduced by treating Compound 28 with a different Grignard reagent than previously used, to form the tertiary alcohol (Compound 19). The tertiary alcohol (Compound 19) which already has the desired R_1, R_2, R_3, R_4 and R_5 substituents, is ring closed under acidic conditions, as described above, to provide the chroman derivative (Compound 20).

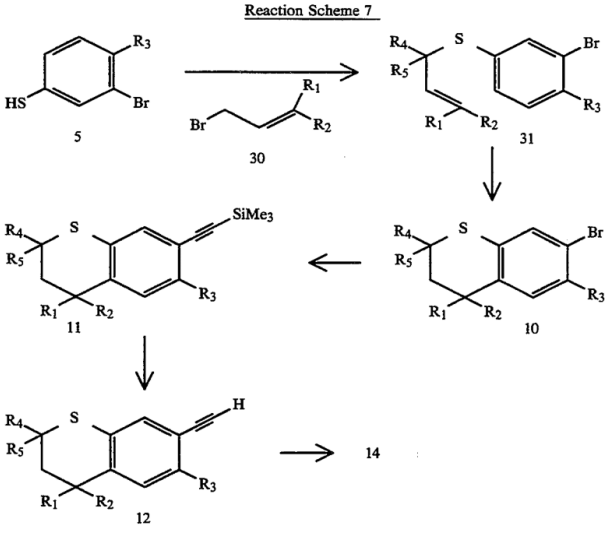
In order to obtain the final acetylenic products where one of the R_4 or R_5 substituent is alkyl and the other is hydrogen, or where the R_4 and R_5 substituents are alkyl but not identical to one another, Compounds 10 and 20 are subjected to substantially the same reaction procedures as outlined in Reaction Scheme 1 and Reaction Scheme 2.

With reference to the compounds of Formula 1, Reaction Scheme 7 illustrates an example of their synthesis when X=S and R_4 and R_5 are both hydrogen.

Thus, with reference to Reaction Scheme 7, the 3-bromothiophenol (Compound 5) (which maybe alkyl substituted in the 4-position) is alkylated with Compound 30. The resulting 3-bromo phenyl sulfides (Compound 31) are ring closed under Friedel Crafts (or like) conditions by refluxing in an inert solvent such as benzene or toluene, in the presence of phosphorus pentoxide and phosphoric acid. The resulting thiochroman (Compound 10) made in accordance with Reaction Scheme 7 has R_4 and R_5 as hydrogen, and preferably in accordance with this reaction scheme, R_1 and R_2 are methyl and R_3 is hydrogen.

To introduce the acetylene (ethyne) portion into the molecule, the substituted 7-bromothiochromane (Compound 10) is reacted with trimethylsilyl-acetylene in the presence of cuprous iodide and a suitable catalyst, typically having the formula $Pd(PQ_3)_2Cl_2$ (Q is phenyl). The reaction is typically conducted in the presence of bis(triphenylphosphine) palladium (II) chloride catalyst, an acid acceptor, (such as triethylamine) under an inert gas (argon) atmosphere, by heating in a sealed tube. The resulting 7-trimethylsilylethynylthiochroman, is shown as Compound 11 in Reaction Scheme 7.

Reaction Scheme 7

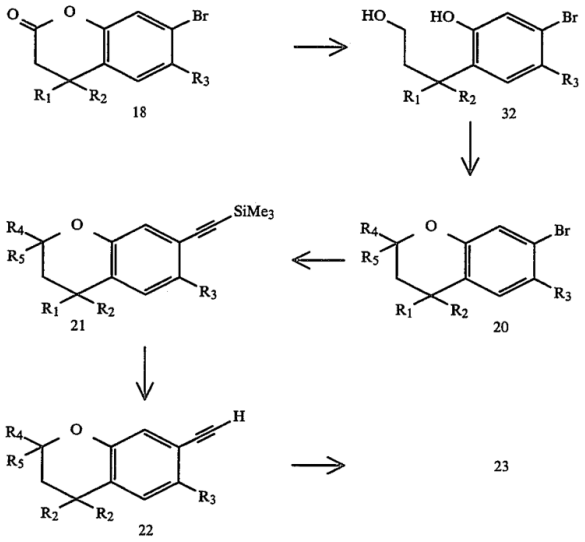


As is further shown on Reaction Scheme 7, the trimethylsilyl moiety is removed from the 7-trimethylsilylethynyl-thiochroman (Compound 11) in the next synthetic step, to provide the ring substituted 7-ethynyl-thiochroman derivative (Compound 12). The latter reaction is conducted under basic conditions, preferably under an inert gas atmosphere.

The 7-ethynyl-thiochroman (Compound 12) can be utilized directly in the coupling reaction set forth in Reaction Scheme 1, or prior to coupling can be converted to the corresponding ZnCl salt, as is described above.

Turning to compounds of Formula 1 where X=O and where R_4 and R_5 are H, (that is turning to chromans substituted in the 4, and possibly in the 6 position) the compounds can be made as indicated in Reaction Scheme 8.

Reaction Scheme 8



In Reaction Scheme 8 R_1 and R_2 are hydrogen or lower alkyl having 1 to 6 carbons, and R_3 is defined as above in connection with Formula 1.

The 2-oxo-chroman (Compound 18) of Reaction Scheme 2 is reduced with lithium aluminum hydride (or by a similar reducing agent) to provide the diol (Compound 32). The primary hydroxyl group of Compound 32 is selectively mesylated over the phenolic hydroxyl, followed by intramolecular displacement of the mesylate group under basic conditions, to give a 7-bromochroman derivative (Compound 20) which bears the desired alkyl substituents at R_1 and R_2 and where R_4 and R_5 are both hydrogen. The acetylenic (ethyne) function is introduced into the 4,4 disubstituted (and optionally 6-substituted) chroman Compound 20 in a sequence of reaction steps which are described in Reaction Scheme 7 in connection with the analogous thiochroman compounds.

As is further shown on Reaction Scheme 8 and in analogy to the reaction sequence shown on Reaction Scheme 7, the trimethylsilyl moiety is removed from 7-trimethylsilyl-ethynyl-thiochroman (Compound 21) under basic conditions, preferably under an inert gas atmosphere.

The 7-ethynyl-thiochroman (Compound 22) can be utilized directly in the coupling reaction as described for Reaction Scheme 2, or prior to coupling can be converted to the corresponding ZnCl salt as is described above.

SPECIFIC EXAMPLES

S-(3-Bromophenyl) 3,3-dimethyl-thio acrylate (Compound 33)

To an ice-bath cooled solution of 4.5 g (112.5 mmol) of sodium hydride (60% suspension in mineral oil) in 50 ml of dry THF was added slowly under argon a solution of 20 g (105.8 mmol) of 3-bromothiophenol in 80 ml of dry THF. The mixture was stirred at 0° C. for 30 minutes and then treated with a solution of 14 g (118 mmol) of dimethylacryloyl chloride in 30 ml of dry THF. The reaction mixture was allowed to stir at room temperature for 24 hours. The reaction mixture was poured onto 300 ml of water containing 5 ml of glacial acetic acid and the organic layer was separated. The aqueous layer was extracted with 2×200 ml ether. The organic extracts were combined and washed with 100 ml of water and 100 ml of saturated NaCl solution and then dried ($MgSO_4$). The solvent was removed in vacuo and the residue was kugelrohr distilled to give the title compound as a pale yellow oil.

PMR ($CDCl_3$); & 1.90 (3H, s), 2.14 (3H, s), 6.04 (1H, s), 7.26 (1H, t, J~7.8 Hz), 7.36 (1H, d, J~4 Hz), 7.5 (1H, dd, J~7.8 Hz, J~1.7 Hz), 7.59 (1H, d, J~1.7 Hz)

4,4 Dimethyl-7-bromo-2-oxo-thiochroman (Compound 34)

To a stirred, ice cooled suspension of 20 g (150 mmol) of aluminum chloride in 250 ml of methylene chloride was added a solution of 17 g (89.5 mmol) of S-(3-bromopenyl) 3-3 dimethylthio acrylate (Compound 33) in 100 ml of methylene chloride. The mixture was

stirred at room temperature for 24 hours and then poured nto 200 ml of an ice and brine mixture. The organic layer was separated and the aqueous layer was extracted with 150 ml of ether. The organic extracts were combined and then washed with water and saturated NaCl solution and dried ($MgSO_4$). The solvent was removed in vacuo and the residue purified by flash column chromatography (silica; 2% ethyl acetate, hexanes) to give the title compound as a white solid.

PMR ($CDCl_3$); & 1.38 (6H, s), 2.65 (2H, s) 7.33 (3H, s).

5-Bromo-2-(1,1,3-trimethyl-3 hydroxy butyl)-thiophenol (Compound 35)

To 132 g (354.3 mmol) of Cerium chloride (dried under vacuum at 135° C. for 2 days) was added 200 ml of dry THF, and the suspension was stirred at room temperature for 20 hours. The reaction mixture was then cooled to 0° C. and treated with 103 ml (309 mmol) of a 3.0M solution of methyl magnesium chloride in THF. The mixture was stirred at room temperature for 4 hours, cooled to 0° C., and treated with a solution of 9.6 g (35.4 mmol) of 4-4 dimethyl-7-bromo-2-oxo-thiochroman (Compound 34) in 60 ml of dry THF. The reaction mixture was allowed to stir at room temperature for 18 hours and then poured into 200 ml of ice containing 2 ml of sulfuric acid. The mixture was extracted with 500 ml of ether. The ether extracts were combined and washed with 300 ml of water and 300 ml of saturated NaCl solution and then dried ($MgSO_4$). The solvent was removed in vacuo to give the title compound as a pale yellow oil.

PMR ($CDCl_3$); & 1.08 (6H, s), 1.54 (6H, s), 2.31 (2H, s), 7.24 (1H, dd, J~8.5 Hz, J~2 Hz), 7.30 (1H, d, J~8.5 Hz) 7.34, (1H, d, J~2 Hz).

2,2,4,4 Tetramethyl-7-bromo-thiochroman (Compound 36)

A mixture of 10 g (33 mmol) of 5-bromo-2-(1,1,3-trimethyl-3-hydroxybutyl) thiophenol (Compound 35) and 100 ml of 20 percent aqueous sulfuric acid was heated at reflux for 48 hours. The mixture was cooled to room temperature and extracted with 2×50 ml of ether. The ether extracts were combined and washed with 25 ml of saturated sodium bicarbonate solution and 25 ml of saturated NaCl solution and dried ($MgSO_4$). The solvent was removed in vacuo and the residue purified by flash column chromatography (silica: 2% ethyl acetate in hexanes) followed by kugelrohr distillation to give the title compound as a clear oil.

PMR ($CDCl_3$); & 1.810 (6H, s), 1.282 (6H, s), 1.234 (2H, s), 7.047 (1H, dd, J~2 Hz, J~8.8 Hz) 7.114 (1H, d J~8.8 Hz) 7.16 (1H, d, J~2 Hz).

2,2,4,4 Tetramethyl-7-trimethylsilyl-ethynyl-thiochroman (Compound 37)

A solution of 3 g (10.5 mmol) of 2,2,4,4 tetramethyl-7-bromo-thiochroman (Compound 36) and 5.16 g (52.6 mmol) of trimethylsilylacetylene in 5 ml of triethylamine was placed in a heavy walled glass tube and degassed under nitrogen. The mixture was then treated, under nitrogen, with 184 mg (0.966 mmol) of cuprous iodide and 368 mg (0.524 mmol) of bis (triphenylphosphine) palladium (II) chloride, the reaction mixture was degassed again and placed under nitrogen and the tube was sealed. The mixture was heated at 60° C. for 24 hours, cooled to room temperature, and then filtered through celite. The solvent was removed in vacuo and the residue purified by flash column chromatography (silica; 100% hexanes) to give the title compound as a pale yellow solid.

PMR ($CDCl_3$); & 0.22 (9H, s), 1.35 (6H, s), 1.38 (6H, s), 1.93 (2H, s), 7.16 (1H, dd, J~8.1 Hz, J~1.74 Hz), 7.24 (1H, d, J~1.74 Hz) 7.30 (1H, J~8.1 Hz)

2,2,4,4-Tetramethyl-7-ethynyl-thiochroman (Compound 38)

To a solution of 1.04 g (3.4 mmole) of 2,2,4,4 tetramethlyl-7-trimethylsilylethynyl thiochroman (Compound 37) in 3 ml of isopropanol was added 5 ml of ethanolic KOH solution. The reaction mixture was stirred at room temperature for 24 hours and the alcohol was then removed in vacuo. The residue was extracted with ether (20 ml) and the combined ether layers were washed with water (15 ml) and saturated NaCl solution (20 ml) and dried ($MgSO_4$). The solvent was removed in vacuo and the residue purified by kugelhohr distillation to give the title compound as a clear oil.

PMR ($CDCl_3$); & 1.38 (6H, s), 1.42 (6H, s), 1.95 (2H, s), 3.02 (1H, s), 7.20 (1H, dd, J~8.1 Hz, 2.1 Hz), 7.29 (1H, d, J~2.1 Hz), 7.34 (1H, d, J~8.1 Hz)

Ethyl-4-(2,2,4,4-tetramethyl-7-thiochromanyl) ethynyl-benzoate (Compound 3)

A solution of 390 mg (1.7 mmol) of 2,2,4,4 tetramethyl-7-ethynyl-thio chroman (Compound 38) and 552 mg (2.0 mmol) of ethyl 4-iodobenzoate in 3 ml of triethylamine was placed in a heavy walled glass tube and degassed for 0.25 hours under nitrogen. The mixture was treated with 18 mg (0.256 mmol) of bis (triphenylphosphine) palladium (II) chloride and 8 mg (0.042 mmol) of cuprous iodide under nitrogen and stirred for 5 minutes. The mixture was treated again with the same amounts of bis (triphenylphosphine) palladium II chloride and cuprous iodide, and the mixture was degassed again. The tube was then sealed and the reaction mixture was heated at 45° C. for 70 hours and then cooled to room temperature. The reaction mixture was filtered through celite and the solvent was removed under vacuum. The residue was purified by flash column chromatography (silica; 1%/Ethylacetate in hexane) to give the title compound as a white solid.

PMR ($CDCl_3$): & 1.37–1.43 (15H, m), 1.96 (2H, s), 4.39 (2H, q, J~7.0 Hz), 7.26 (1H, dd, J~8.2 Hz, 1.8 Hz), 7.34 (1H, d, J~1.8 Hz), 7.37 (1H, d, J~8.2 Hz), 7.56 (2H, d, J ~8.0 Hz), 8.02 (2H, d, J~8.0 Hz).

3-Bromophenyl 3,3-dimethyl acrylate (Compound 39)

To an ice-cooled suspension of 4 g (100 mmol) of sodium hydride (60% in mineral oil) in 50 ml of dry THF was added dropwise a solution of 15.7 g (90.7 mmol) of 3-bromo phenol in 25 ml of dry THF. The mixture was stirred at 0 degrees C. for 0.5 hours and then treated with a solution of 10.65 g (90.0 mmol) of dimethyl acryloyl chloride in 30 ml of dry THF. The mixture was allowed to warm to room temperature and stirred for 24 hours. The reaction mixture was poured onto 200 ml of ice water containing 3 ml of glacial acetic acid. The mixture was extracted with 2×250 ml ether and the combined ether extracts were washed with 200 ml of water and 100 ml saturated NaCl solution and dried ($MgSO_4$). The solvent was removed in vacuo and the residue purified by kugelrohr distillation to give the title compound as a clear oil.

PMR ($CDCl_3$): & 2.02 (3H, s), 2.28 (3H, s), 5.94 (1H, broad s), 7.06–7.12 (1H, m), 7.28 (1H, t, J~8.0 Hz), 7.34 (1H, t, J~2.0 Hz), 7.37–7.42 (1H, m).

3-Bromo-2-(1,1,3-Trimethyl-3-hydroxybutyl)phenol (Compound 41)

To a stirred, ice-cooled suspension of 21 g (158 mmol) of aluminum chloride in 200 ml of methylene chloride was added slowly a solution of 23.74 g (93.1 mmol) of 5-bromo-phenyl-3,3-dimethyl acrylate (Compound 39) in 100 ml of methylene chloride. The mixture was warmed to room temperature and stirred for 52 hours. The mixture was poured into a mixture of ice and brine and the organic layer was separated. The aqueous layer was extracted with 2×100 ml ether. The organic extracts were combined and washed with 2×250 ml of water and 50 ml of saturated NaCl solution and dried ($MgSO_4$). The solvent was removed in vacuo and the residue was partially purified by flash column chromatography, (silica; 5% ethyl acetate/hexane) to give impure 4,4-dimethyl-7-bromo-2-oxochroman (Compound 40) as a yellow oil which was used in the next step without further purification. To an ice-cooled solution of 10 g of this impure 4,4,dimethyl-7-bromo-2-oxochroman (Compound 40) in 200 ml of dry THF was added under argon 39.2 ml of 3.0M methyl magnesium chloride (117.6 mmol) in THF. The reaction mixture was allowed to warm to room temperature and stirred for 5 hours. The reaction mixture was then poured into ice water containing 2 ml of sulfuric acid and the organic layer was separated. The aqueous layer was extracted with 200 ml of ether. The organic extracts were combined and washed with 200 ml of water and 200 ml of brine and dried ($MgSO_4$). The solvent was removed in vacuo and the residue purified by flash column chromatography (silica; 10% ethylacetate/hexanes) to give the title compound as a pale yellow oil.

PMR ($CDCl_3$): & 0.98 (6H, s), 1.36 (6H, s), 2.15 (2H, s), 6.82 (1H, d, J~1.9 Hz), 6.86 (1H, dd, J~8.3 Hz, 1.9 Hz), 7.04 (1H, d, J~8.3 Hz).

2,2,4,4-tetramethyl-7-bromochroman (Compound 42)

A mixture of 5.42 (18.9 mmol) of 3-bromo-2(1,1,3 trimethyl-3-hydroxy-butyl) phenol (Compound 41) and 50 ml of 20 percent aqueous sulfuric acid was heated at reflux for 24 hours. The reaction mixture was cooled to room temperature and treated with 100 ml of ether. The organic layer was separated and the aqueous layer was extracted with 50 ml of ether. The ether extracts were combined and washed with 100 ml of water and 100 ml saturated NaCl solution and dried ($MgSO_4$). The solvent was removed in vacuo and the residue was purified by Kugelrohr distillation to give the impure title compound as a pale yellow oil.

PMR ($CDCl_3$): & 1.22 (6H, s), 1.24 (6H, s), 1.72 (2H, s), 6.87 (1H, d, J~2.0 Hz), 6.92 (1H, dd, J~8.3 Hz, 2.0 Hz), 7.02 (1H, d, J~8.3 Hz) .

2,2,4,4-Tetramethyl-7-trimethylsilylethynyl-chroman (Compound 43)

A solution of 2 g (7.4 mmol) of 2,2,4,4 tetramethyl-7-bromochroman (Compound 42) and 3.63 g (37.0 mmol) of trimethylsilylacetylene in 5 ml of triethylamine was placed in a heavy walled glass tube and degassed under nitrogen. The mixture was then treated, under nitrogen, with 130 mg (0.6826 mmol) of cuprous iodide and 260 mg (0.3704 mmol) of bis (triphenyl phosphine) palladium (II) chloride. The reaction mixture was degassed again and placed under nitrogen and the tube was sealed. The mixture was heated to 60 degrees C. for 24 hours and then cooled to room temperature and filtered through celite. The solvent was removed in vacuo and the residue purified by flash column chromatography (silica; 2% ethyl acetate/hexane) to give the title compound as a pale yellow solid.

PMR ($CDCl_3$): & 0.23 (9H, s), 1.32 (12H, s), 1.82 (2H, s), 6.92 (1H, d, J~1.6 Hz) 7.00 (1H, dd, J~8.6 Hz, 1.6 Hz) , 7.19 (1H, J~8.6 Hz)

2,2,4,4-tetramethyl-7-ethynyl chroman (Compound 44)

To a solution of 1.16 g (4.1 mmol) of 2,2,4,4-tetramethyl-7-trimethylsilylethynyl-chroman (Compound 43) in 3 ml of isopropanol was added 5 ml of ethanolic KOH solution. The reaction mixture was stirred at room temperature for 24 hours and the alcohol was then removed under vacuum. The residue was extracted with 2×10 ml of ether and the combined ether extracts were washed with 15 ml of water and 20 ml of saturated NaCl solution and then dried ($MgSO_4$). The solvent was removed in vacuo and the residue purified by Kugelrohr distillation to give the title compound as a white crystalline solid.

PMR ($CDCl_3$): & 1.33 (6H, s), 1.34 (6H, s), 1.83 (2H, s), 2.99 (1H, s), 6.94 (1H, d, J~1.7 Hz), 7.04 (1H, dd, J~8.0 Hz, 1.7 Hz), 7.21 (1H, d, J~8.0 Hz).

Ethyl-4-(2,2,4,4-tetramethyl-7-chromanyl)-ethynyl-benzoate (Compound 4)

A solution of 270 mg (1.26 mmol) of 2,2,4,4 tetramethyl-7-ethynyl chroman (Compound 44) and 420 mg (1.52 mmol) of ethyl 4-iodobenzoate in 4 ml of triethylamine was placed in a heavy walled glass tube and degassed under nitrogen for 15 minutes. The mixture was treated with 6 mg (0.0315 mmol) of cuprous iodide and 13 mg (0.0185 mmol) of bis(triphenylphosphine) palladium (II) chloride under nitrogen, and stirred for 5 minutes. The mixture was treated again with the same amounts of bis (triphenyl phosphine) palladium II chloride and cuprous iodide. The mixture was degassed again and the tube was sealed. The reaction mixture was heated to 45° for 70 hours and then cooled to room temperature. The reaction mixture was filtered through celite and the solvent was removed under vacuum. The residue was purified by flash column chromatography (silica; 1% ethyl acetate in hexanes) to give the title compound as a white solid.

PMR ($CDCl_3$): & 1.32 (6H, s), 1.33 (6H, s), 1.40 (3H, t, J ~7.2 Hz), 1.84 (2H, s), 4.38 (2H, d, J~7.2 Hz), 7.00 (1H, d, J~1.4 Hz), 7.09 (1H, dd, J~7.9 Hz, 1.4 Hz), 7.25 (1H, d, J ~7.9 Hz), 7.56 (2H, d, J~8.3 Hz), 8.02 (2H, J~8.3 Hz).

3-Bromophenyl-3-methyl-but-2-enylsulfide (Compound 45)

A solution of 25 g (132 mmol) of 3-bromothiophenol in 100 ml of acetone was heated to reflux and then treated with 5.56 g (139 mmol) of powdered NaOH. The mixture was refluxed for a further 0.5 hour. The refluxing mixture was then treated with a solution of 19.7 g (132 mmol) of 1-bromo-3-methyl-2-butene in 30 ml of acetone and refluxed for a further 1.5 hours. The mixture was cooled and then solvent was removed in-vacuo. The residue was extracted with ether and the ether extract was washed with dilute NaOH solution, water, and saturated NaCl solution and thereafter dried ($CaCl_2$). After evaporation of the solvent, the residue

was purified by vacuum distillation to give the title compound as a white crystalline solid.

PMR ($CDCl_3$): & 1.61 (3H, s), 1.72 (3H, s), 3.52 (2H, d, J ~7.8 Hz), 5.27 (1H, t, J~7.8 Hz) 7.10 (1H, t, J~7.8 Hz), 7.21 (1H, dt, J~7.8 Hz, J~1.8 Hz), 7.27 (1H, dt, J~7.8 Hz, J~1.8 Hz), 7.44 (1H, t, J~1.8 Hz).

4,4,Dimethyl-7-trimethylsilylethynyl-Thiochroman (Compound 47)

To 3.63 g (14 mmol) of 3-bromophenyl-3-methyl-but-2-enyl sulfide (Compound 45) was added 15 g of a 1:10 P_2O_5, $MeSO_3H$ mixture, and stirred at room temperature for 4 hours. The mixture was treated with cool water followed by boiling water. The mixture was stirred for 10 minutes and cooled to room temperature. The reaction mixture was then extracted with ether and the combined ether extracts were washed with water and then saturated NaCl solution and dried ($CaCl_2$). The solvent was removed in vacuo and the residue purified by Kugelrohr distillation (140 degrees C./0.2 mm) to give impure 4,4-dimethyl-7-bromo-thiochroman (Compound 46) as a pale yellow solid. This was used in the next step without further purification. A solution of 2.03 g of this impure 4,4 dimethyl-7-bromo thiochroman (Compound 46) in 2 ml of triethylamine was placed in a heavy-walled tube and degassed and then treated under argon with 3.8 g (38.9 mmol) of trimethylsilylacetylene a powdered mixture of 100 mg of bis (triphenylphosphine) palladium (II) chloride and 50 mg of cuprous iodide. The reaction mixture was degassed again, then placed under argon and the tube was sealed. The mixture was heated at 60° C. for 12 hours. The mixture was cooled to room temperature and then filtered through celite. The solvent was removed in vacuo and the residue purified by flash chromatography (silica; hexanes) to give the title compound as a yellow oil.

PMR ($CDCl_3$): & 0.22 (9H, s), 1.3 (6H, s), 1.91–1.98 (2H, t, J~6.0 Hz), 2.99–3.2 (2H, t, J~6.0 Hz) 7.09 (1H, dd, J~1.8, J~8.2 Hz) 7.20 (1H, d, J~1.8 Hz) 7.26 (1H, d, J~8.2 Hz).

Ethyl-4-(4,4-dimethyl-7-thiochromanyl)-ethynyl-benzoate (Compound 1)

To a solution of 1 g (3.6 mmol) of 4,4-dimethyl-7-trimethylsilylethynyl-thio-chroman (Compound 47) in 10 ml of isopropyl alcohol was added 5 ml of 1N KOH solution. The reaction mixture was stirred at room temperature for 18 hours and the isopropanol was then removed under vacuum. The residue was extracted with ether and the ether extracts were combined and washed with dilute HCl solution, water and saturated NaCl solution, and were thereafter dried ($MgSO_4$). The solvent was removed in vacuo to give impure 4-4-dimethyl-7-ethynyl-thiochroman (Compound 48) as a pale yellow oil. This mixture was used in the next step without further purification. A solution of 281 mg of this impure 4,4-dimethyl-7-ethynylthiochroman (Compound 48) and 384 mg (1.4 mmol) of ethyl 4-iodo-benzoate in 1 ml of triethylamine was placed in a heavy walled glass tube and degassed under nitrogen. The mixture was treated with 60 mg of cuprous iodide and 30 mg of bis (triphenyl phosphine) palladium (II) chloride. The reaction mixture was degassed again under nitrogen, and the tube was sealed. The reaction mixture was stirred at room temperature for 18 hours. The solvent was removed under high vacuum and the residue purified by flash column chromatography (silica; 3% EtOAC/hexane) to give the title compound as an off-white solid.

PMR ($CDCl_3$): & 1.32 (6H, s), 1.40 (3H, t, J~7.1 Hz), 1.92–1.99 (2H, m), 2.99–3.06 (2H, m), 4.38 (2H, q, J~7.1 Hz, 7.17 (1H, dd, J~8.2 Hz, 1.7 Hz), 7.28 (1H, d, J~1.7 Hz), 7.34 (1H, d, J~8.2 Hz), 7.55 (2H, d, J~8.5 Hz), 8.01 (2H, d, J~8.5 Hz).

4-[4-4-Dimethyl-7-thiochromanyl)-ethynyl] benzoic acid (Compound 2)

To 150 mg (0.428 mmol) of ethyl-4-(4,4-dimethyl-7-thiochromanyl-ethynyl-benzoate (Compound 1) was added 5 ml of ethanolic KOH solution and the reaction mixture was stirred at room temperature for 24 hours. The ethanol was removed in vacuo and the residue was taken up in 3 ml of water and 3 ml of ether. The layers were separated and the aqueous layer was washed with ether. The aqueous layer was acidified to Ph=2 with 1N HCl and was extracted with 2×20 ml of ether. The combined ether extracts were washed successively with water and saturated NaCl solution and then dried ($MgSO_4$). The solvent was removed in vacuo to give the title compound as a white solid.

PMR ($CDCl_3$): & 1.30 (6H, s), 1.91–1.97 (2H, m), 2.99–3.04 (2H, m), 7.14 (1H, dd, J~8.1 Hz, 1.7 Hz), 7.24 (1H, d, J~1.7 Hz), 7.32 (1H, d, J~8.1 Hz), 7.52 (2H, d, J~8.3 Hz), 7.99 (2H, d, J~8.3 Hz).

4-(2,2,4,4-Tetramethyl-7-thiochromanyl)-ethynyl-benzoic acid (Compound 49)

To 99 mg (0.262 mmol) of ethyl-4-(2,2,4,4-tetramethyl-7-thiochromanyl) ethynyl-benzoate (Compound 3) was added 5 ml of an ethanolic KOH solution. The reaction mixture was stirred at room temperature for 48 hours and the solvent was then removed in vacuo. The residue was taken up with water and ether and the layers were separated. The aqueous layer was acidified to Ph~2 with 1N HCl and extracted with ether. The ether layer was then washed with water and saturated sodium chloride and then dried ($MgSo_4$). The solvent was removed in vacuo to give the title compound as a white solid PMR ($CDCl_3$): & 1.40 (6H, s), 1.42 (6H, s), 1.98 (2H, s), 7.23–7.29 (2H, m), 7.43 (1H, d, J~8.6 Hz), 7.57 (2H, d, J~8.5 Hz), 8.01 (2H, d, J~8.5 Hz).

4- (2,2,4,4-tetramethyl-7-chromanyl)-ethynyl benzoic acid (Compound 50)

To 207.2 mg (0.572 mmol) of ethyl-4-(2,2,4,4-tetramethyl-7-chromanyl) ethynyl-benzoate (Compound 4) was added 5 ml of ethanolic KOH solution. The reaction mixture was stirred at room temperature for 48 hours and the solvent was then removed in vacuo. The residue was taken up with water and ether and the layers were separated. The aqueous layer was acidified to Ph~2 with 1N HCl and extracted with ether. The ether layer was then washed with water and saturated sodium chloride and then dried ($MgSO_4$). The solvent was removed in vacuo to give the title compound as a pale yellow solid.

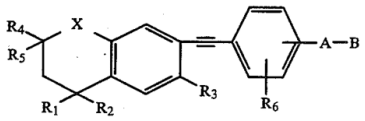
PMR ($CDCl_3$): & 1.37 (12H, s), 1.87 (2H, s), 6.95 (1H, d, J~1.6 Hz), 7.09 (1H, dd, J~8.0 Hz, 1.6 Hz), 7.30 (1H, d, J ~8.0 Hz), 7.57 (2H, d, J~8.3 Hz), 8.02 (2H, d, J~8.3 Hz).

What is claimed is:

1. A compound of the formula

This section list the patent claims in detail.

This is an example of an independent patent claim. It has no reference to another claim so is therefore classified as an independent claim. The total number of independent claims in this section gives the value for the variable independent claims (IC).



where X is S or O;

R_1, R_2, R_4 and R_5 independently are hydrogen or lower alkyl of 1 to 3 carbons, and R_3 is hydrogen or lower alkyl;

A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ and where n is an integer from 0–5;

R_6 is hydrogen, lower alkyl, lower alkenyl or lower cycloalkyl having 1 to 6 carbons, or halogen; and

B is COOH or a pharmaceutically acceptable salt thereof, $COOR_8$, $COONR_9R_{10}$, where R_8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R_8 is phenyl or lower alkylphenyl, R_9 and R_{10} independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl.

2. A compound of claim **1** where X is S.

3. A compound of claim **2** where A is $(CH_2)_n$ and n is 0, 1, or 2.

4. A compound of claim **3** where R_4 is H and R_5 is lower alkyl.

5. A compound of claim **3** where R_4 is the same alkyl group as R_5.

6. A compound of claim **3** where R_4 is lower alkyl and R_5 is lower alkyl and R_4 and R_5 are different.

7. A compound of claim **3** where R_4 and R_5 are both hydrogen.

8. A compound of claim **1** where X is O.

9. A compound of claim **8** where A is $(CH_2)_n$ and n is 0, 1, or 2.

10. A compound of claim **9** where R_4 is H and R_5 is lower alkyl.

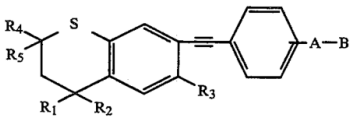
11. A compound of claim **9** where R_4 is lower alkyl and R_5 is lower alkyl and R_4 and R_5 are different.

12. A compound of claim **9** where R_4 is the same alkyl group as R_5.

13. A compound of claim **10** where R_4 and R_5 are both H.

14. A pharmaceutical composition comprising one or more compounds set forth in claim **1**, the composition including a pharmaceutically acceptable excipient.

15. A compound of the formula



where R_1, R_2, R_4 and R_5 independently are hydrogen or lower alkyl of 1 to 3 carbons and R_3 is hydrogen or lower alkyl;

A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ where n is an integer between 0 to 5; and

B is COOH or a pharmaceutically acceptable salt thereof, $COOR_8$, $COONR_9R_{10}$, where R_8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R_8 is phenyl or lower alkylphenyl, R_9 and R_{10} independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl.

16. A compound of claim **15** wherein A is $(CH_2)_n$ and n is 0.

17. A compound of claim **16** where R_1–R_5 are independently H or methyl.

18. A compound of claim **17** where B is $COOC_2H_5$.

19. The compound of claim **18** where R_3 is H and R_1, R_2, R_4 and R_5 are methyl.

20. The compound of claim **18** where R_1, R_2 are methyl and R_3, R_4 and R_5 are hydrogen.

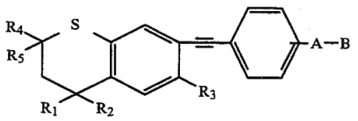
21. A compound of claim **17** where B is COOH.

22. A compound of claim **21** where R_3 is H and R_1, R_2, R_4 and R_5 are methyl.

23. A compound of claim **21** where R_1 and R_2 are methyl and R_3, R_4 and R_5 are H.

24. The compound of claim **16** where B is COOH or a pharmaceutically acceptable salt thereof, R_1, R_2 are methyl, and R_3, R_4 and R_5 are hydrogen.

25. A compound of the formula



where R_1, R_2, R_4 and R_5 independently are hydrogen or lower alkyl of 1 to 3 carbons and R_3 is hydrogen or lower alkyl;

A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ where n is an integer between 0 to 5, and

B is COOH or a pharmaceutically acceptable salt thereof, $COOR_8$, $COONR_9R_{10}$, where R_8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R_8 is phenyl or lower alkylphenyl, R_9 and R_{10} independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl.

26. A compound of claim **25** where A is $(CH_2)_n$ and n is 0.

27. A compound of claim **26** where B is $COOC_2H_5$.

28. The compound of claim **27** where R_3 is hydrogen, R_1, R_2, R_4 and R_5 are methyl.

29. A compound of claim **26** where B is COOH.

30. A compound of claim **29** where R_3 is hydrogen, R_1, R_2, R_4 and R_5 are methyl.

* * * * *

This is an example of a dependent claim. It reference s another claim (claim 29) so is therefore classified as a dependent claim. The total number of dependent claims in this section gives the value for the variable dependent claims (DC).

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 5,346,915
DATED : September 13, 1994
INVENTOR(S) : Roshantha A.S. Chandraratna

It is certified that error appears in the above-indentified patent and that said Letters Patent is hereby corrected as shown below:

Title page, item [54] and col. 1, line 3, in the title, "RETINOLD-LIKE" should be --RETINOID-LIKE--;
Column 21, line 2, "nto" should be --into--;
Column 26, line 5, before "7.17", "Hz" should be --Hz)--;
Column 26, line 41, "Mgso4" should be --$MgSO_4$--.

Signed and Sealed this

Second Day of January, 1996

Attest:

BRUCE LEHMAN

Attesting Officer

Commissioner of Patents and Trademarks

Patent U.S. Patent 7211600, Front Page Only

(12) **United States Patent**
Lipson et al.

(10) **Patent No.: US 7,211,600 B2**
(45) **Date of Patent: May 1, 2007**

(54) **METHODS OF MODULATING *C-KIT* TYROSINE PROTEIN KINASE FUNCTION WITH INDOLINONE COMPOUNDS**

(75) Inventors: **Ken Lipson**, San Mateo, CA (US); **Gerald McMahon**, Kenwood, CA (US)

(73) Assignee: **Sugen Inc.**, South San Francisco, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/205,474**

(22) Filed: **Aug. 16, 2005**

(65) **Prior Publication Data**

US 2005/0288353 A1 Dec. 29, 2005

Related U.S. Application Data

(63) Continuation of application No. 10/600,868, filed on Jun. 23, 2003, now abandoned, which is a continuation of application No. 09/741,842, filed on Dec. 22, 2000, now abandoned.

(60) Provisional application No. 60/171,963, filed on Dec. 22, 1999.

(51) **Int. Cl.**
A01N 43/38 (2006.01)

(52) **U.S. Cl.** **514/418**; 514/312; 514/414; 514/404; 514/469

(58) **Field of Classification Search** ..... None
See application file for complete search history.

Section 65 lists the prior publication date of the patent application. This patent was published prior to its grant date.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,376,110	A	3/1983	David et al.	
4,966,849	A	10/1990	Vallee et al.	
5,217,999	A	6/1993	Levitzki et al.	
5,330,992	A	7/1994	Eissenstat et al.	
5,792,783	A	8/1998	Tang et al.	
5,883,116	A	3/1999	Tang et al.	
5,905,149	A	5/1999	Battistini et al.	
6,339,100	B1	1/2002	Longley	
6,395,734	B1	5/2002	Tang et al.	
6,514,981	B1	2/2003	Tang et al.	
6,689,806	B1 *	2/2004	Tang et al.	514/418
6,878,733	B1	4/2005	Shenoy et al.	

FOREIGN PATENT DOCUMENTS

WO	WO 91/15495	10/1991
WO	WO 92/20642	11/1992
WO	WO 92/21660	12/1992
WO	WO 94/03427	2/1994
WO	WO 94/10202	5/1994
WO	WO 94/14808	7/1994
WO	WO 96/22976	8/1996
WO	WO 98/07695	2/1998
WO	WO 99/06468	2/1999
WO	WO 99/10325	3/1999
WO	WO 99/61422	12/1999

OTHER PUBLICATIONS

Akbasak, A., "Oncogenes: Cause Or Consequence In The Development Of Glial Tumors," *Journal of the Neurological Sciences*, 1992, 119-133, vol. 111.

Andoh, A., "Rapid Intestinal Ischaemia-Reperfusion Injury Is Suppressed In Genetically Mast Cell-Deficient Ws/Ws Rats," *Clinical Experimental Immunology*, 1999, 90-93, vol. 116.

Beck, D., "Expression Of Stem Cell Factor And Its Receptor By Human Neuroblastoma Cells And Tumors," *Blood*, 1995, 3132-3138, vol. 86, No. 8.

Bedi, A., "BCR-ABL-Mediated Inhibition Of Apoptosis With Delay Of G2/M Transition After DNA Damage: A Mechanism Of Resistance To Multiple Anticancer Agents," *Blood*, 1995, 1148-1158, vol. 86, No. 3.

Bellone, G., "Growth Stimulation Of Colorectal Carcinoma Cells Via The c-Kit Receptor Is Inhibited By TGF-β1," *Journal of Cellular Physiology*, 1997, 1-11, vol. 172.

Berdel, W., "Recombinant Human Stem Cell Factor Stimulates Growth Of A Human Glioblastoma Cell Line Expressing c-Kit Protooncogene," *Cancer Research*, 1992, 3498-3502, vol. 52.

Blume-Jensen, P., "Kit/Stem Cell Factor Receptor-Induced Activation Of Phosphatidylinositol 3'-Kinase is Essential for Male Fertility," Nature Genetics, 2000, 157-162, vol. 24.

Bokemeyer, C., "Expression Of Stem-Cell Factor And Its Receptor c-Kit Protein In Normal Testicular Tissue And Malignant Germ-Cell Tumours," Journal of Cancer Research and Clinical Oncology, 1996, 301-306, vol. 122.

Broudy, V., "Stem Cell Factor and Hematopoiesis," *Blood*, 1997, 1345-1364, vol. 90, No. 4.

Carpino, N., "p62^{dok} A Constitutively Tyrosine-Phosphorylated, GAP-Associated Protein In Chronic Myelogenous Leukemia Progenitor Cells," *Cell*, 1997, 197-204, vol. 88.

Castells, M., "The Presence Of Membrane-Bound Stem Cell Factor On Highly Immagture Nonmetachromatic Mast Cells In The Peripheral Blood Of A Patient With Aggressive Systemic Mastocytosis," *Journal of Allergy and Clinical Immunology*, 1996, 831-840, vol. 98, No. 4.

Chao, M., "Growth Factor Signaling: Where Is The Specificity?" Cell, 1992, 995-997, vol. 68.

Cohen, P., "Expression Of Stem Cell Factor And c-Kit In Human Neuroblastoma," *Blood*, 1994, 3465-3472, vol. 84, No. 10.

Columbo, M., "The Human Recombinant c-Kit Receptor Ligand, rhSCF, Induces Mediator Release From Human Cutaneous Mast Cells And Enhances IgE-Dependent Mediator Release From Both Skin Mast Cells And Peripheral Blood Basophils," *The Journal of Immunology*, 1992, 599-608, vol. 149, No. 2.

Costa, J., "The Cells Of The Allergic Response," *JAMA*, 1997, 1815-1822, vol. 278, No. 22.

Dastych, J., "Stem Cell Factor Induces Mast Cell Adhesion To Fibronectin," Journal of Immunology, 1994, 213-219, vol. 152.

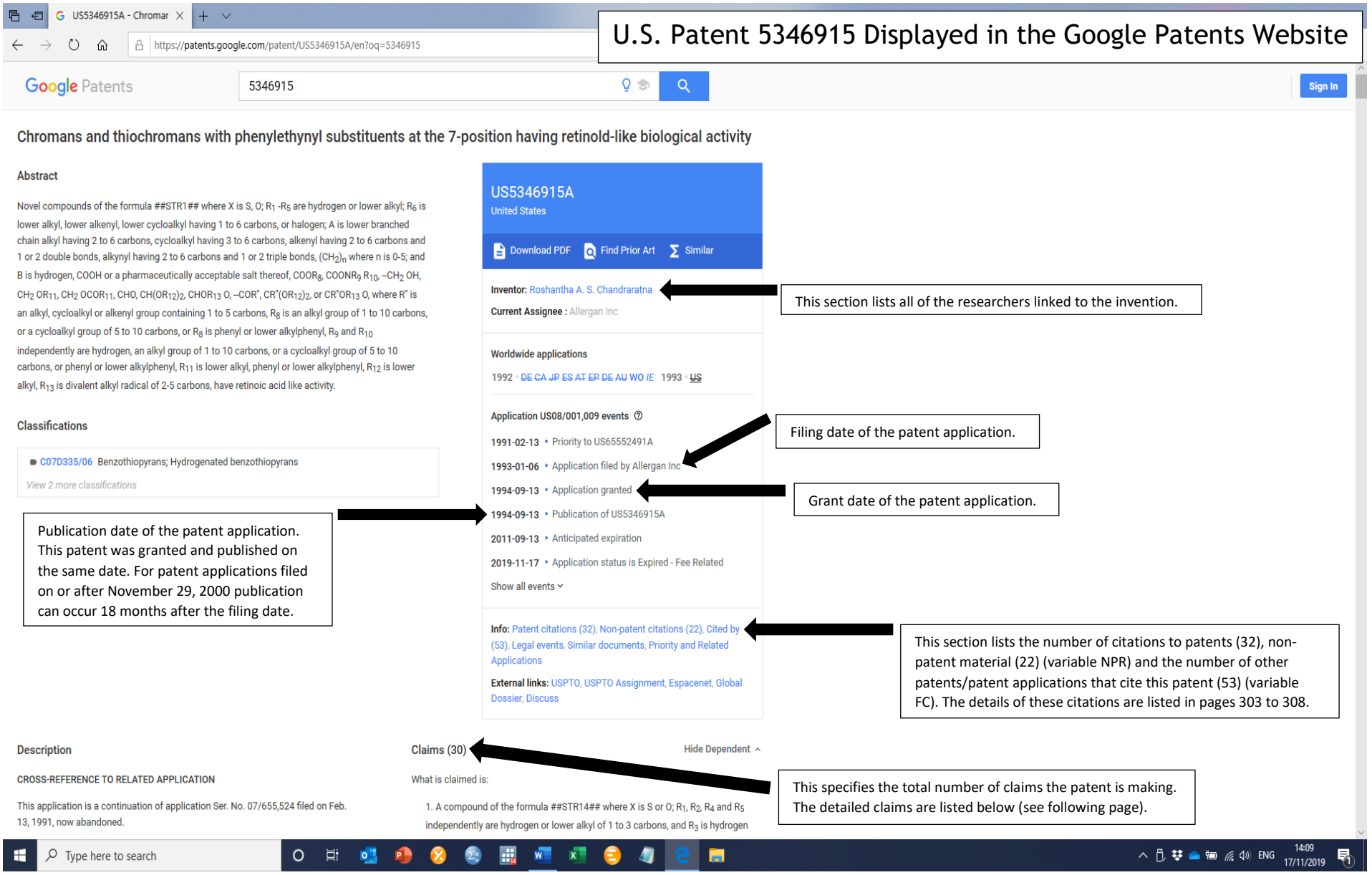
(Continued)

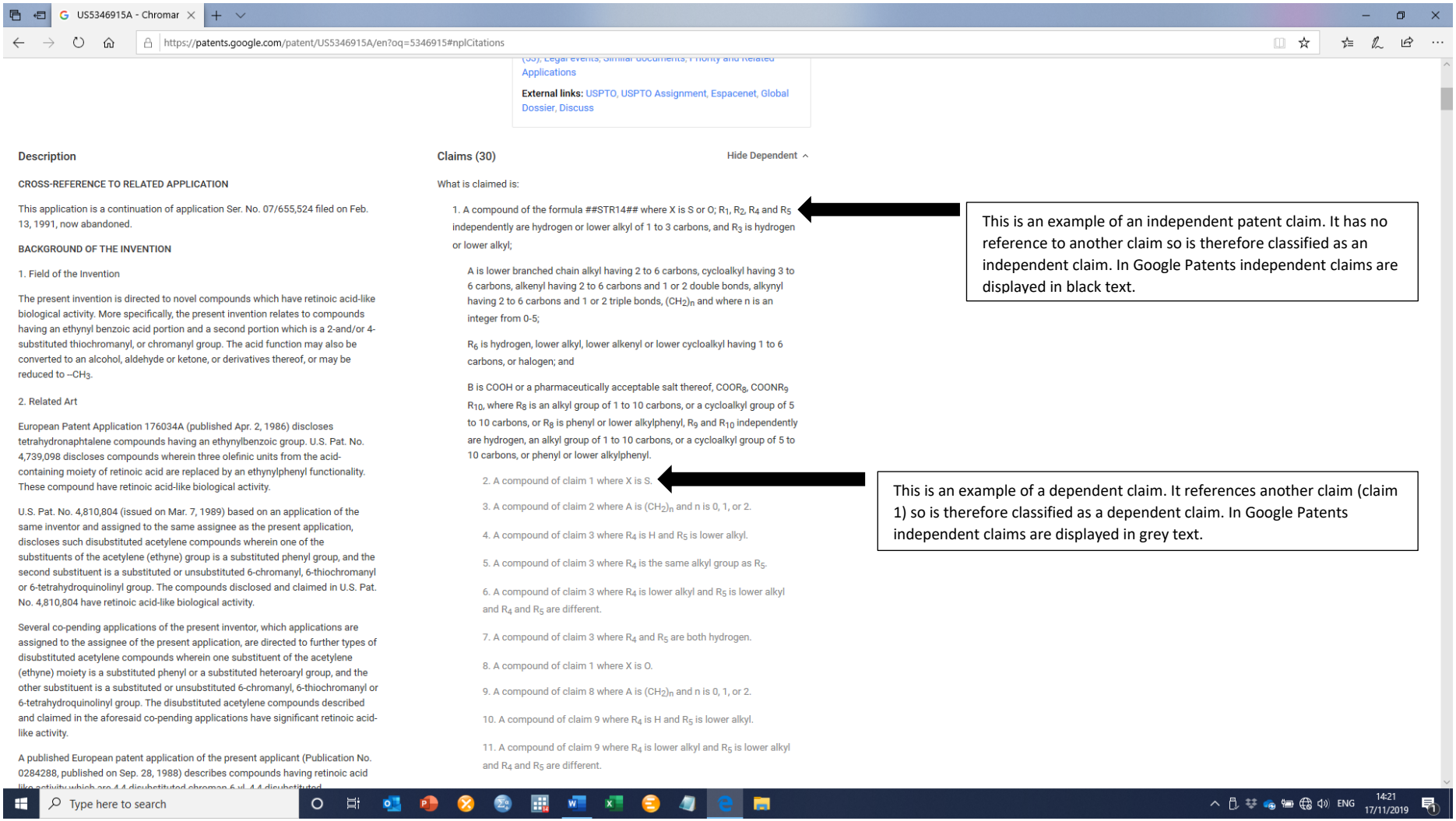
Primary Examiner—Ardin H. Marschel
Assistant Examiner—Shirley V. Gembeh
(74) *Attorney, Agent, or Firm*—Bryan C. Zielinski; Vincent P. Liptak

(57) **ABSTRACT**

The present invention concerns compounds and their use to inhibit the activity of a receptor tyrosine kinase. The invention is preferably used to treat cell proliferative disorders such as cancers characterized by over-activity or inappropriate activity c-kit kinase.

4 Claims, 2 Drawing Sheets


U.S. Patent 5346915 Displayed in the Google Patents Website
https://patents.google.com/patent/US5346915A/en?oq=5346915
Google Patents
5346915
Sign In
Chromans and thiochromans with phenylethynyl substituents at the 7-position having retinold-like biological activity
Abstract
Novel compounds of the formula ##STR1## where X is S, O; R1 -R5 are hydrogen or lower alkyl; R6 is lower alkyl, lower alkenyl, lower cycloalkyl having 1 to 6 carbons, or halogen; A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, (CH2)n where n is 0-5; and B is hydrogen, COOH or a pharmaceutically acceptable salt thereof, COOR8, COONR9 R10, --CH2 OH, CH2 OR11, CH2 OCOR11, CHO, CH(OR12)2, CHOR13 O, --COR", CR"(OR12)2, or CR"OR13 O, where R" is an alkyl, cycloalkyl or alkenyl group containing 1 to 5 carbons, R8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R8 is phenyl or lower alkylphenyl, R9 and R10 independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl, R11 is lower alkyl, phenyl or lower alkylphenyl, R12 is lower alkyl, R13 is divalent alkyl radical of 2-5 carbons, have retinoic acid like activity.
Classifications
C07D335/06 Benzothiopyrans; Hydrogenated benzothiopyrans
View 2 more classifications
US5346915A
United States
Download PDF
Find Prior Art
Similar
Inventor: Roshantha A. S. Chandraratna
Current Assignee : Allergan Inc
This section lists all of the researchers linked to the invention.
Worldwide applications
1992 · DE CA JP ES AT EP DE AU WO IE 1993 · US
Application US08/001,009 events
1991-02-13 • Priority to US65552491A
1993-01-06 • Application filed by Allergan Inc
Filing date of the patent application.
1994-09-13 • Application granted
Grant date of the patent application.
1994-09-13 • Publication of US5346915A
Publication date of the patent application. This patent was granted and published on the same date. For patent applications filed on or after November 29, 2000 publication can occur 18 months after the filing date.
2011-09-13 • Anticipated expiration
2019-11-17 • Application status is Expired - Fee Related
Show all events
Info: Patent citations (32), Non-patent citations (22), Cited by (53), Legal events, Similar documents, Priority and Related Applications
External links: USPTO, USPTO Assignment, Espacenet, Global Dossier, Discuss
This section lists the number of citations to patents (32), non-patent material (22) (variable NPR) and the number of other patents/patent applications that cite this patent (53) (variable FC). The details of these citations are listed in pages 303 to 308.
Description
Claims (30)
Hide Dependent
CROSS-REFERENCE TO RELATED APPLICATION
This application is a continuation of application Ser. No. 07/655,524 filed on Feb. 13, 1991, now abandoned.
What is claimed is:
1. A compound of the formula ##STR14## where X is S or O; R1, R2, R4 and R5 independently are hydrogen or lower alkyl of 1 to 3 carbons, and R3 is hydrogen
This specifies the total number of claims the patent is making. The detailed claims are listed below (see following page).
Type here to search
ENG
14:09
17/11/2019


US5346915A - Chromar
https://patents.google.com/patent/US5346915A/en?oq=5346915#nplCitations
Applications
External links: USPTO, USPTO Assignment, Espacenet, Global Dossier, Discuss
Description
CROSS-REFERENCE TO RELATED APPLICATION
This application is a continuation of application Ser. No. 07/655,524 filed on Feb. 13, 1991, now abandoned.
BACKGROUND OF THE INVENTION
1. Field of the Invention
The present invention is directed to novel compounds which have retinoic acid-like biological activity. More specifically, the present invention relates to compounds having an ethynyl benzoic acid portion and a second portion which is a 2-and/or 4-substituted thiochromanyl, or chromanyl group. The acid function may also be converted to an alcohol, aldehyde or ketone, or derivatives thereof, or may be reduced to --CH3.
2. Related Art
European Patent Application 176034A (published Apr. 2, 1986) discloses tetrahydronaphtalene compounds having an ethynylbenzoic group. U.S. Pat. No. 4,739,098 discloses compounds wherein three olefinic units from the acid-containing moiety of retinoic acid are replaced by an ethynylphenyl functionality. These compound have retinoic acid-like biological activity.
U.S. Pat. No. 4,810,804 (issued on Mar. 7, 1989) based on an application of the same inventor and assigned to the same assignee as the present application, discloses such disubstituted acetylene compounds wherein one of the substituents of the acetylene (ethyne) group is a substituted phenyl group, and the second substituent is a substituted or unsubstituted 6-chromanyl, 6-thiochromanyl or 6-tetrahydroquinolinyl group. The compounds disclosed and claimed in U.S. Pat. No. 4,810,804 have retinoic acid-like biological activity.
Several co-pending applications of the present inventor, which applications are assigned to the assignee of the present application, are directed to further types of disubstituted acetylene compounds wherein one substituent of the acetylene (ethyne) moiety is a substituted phenyl or a substituted heteroaryl group, and the other substituent is a substituted or unsubstituted 6-chromanyl, 6-thiochromanyl or 6-tetrahydroquinolinyl group. The disubstituted acetylene compounds described and claimed in the aforesaid co-pending applications have significant retinoic acid-like activity.
A published European patent application of the present applicant (Publication No. 0284288, published on Sep. 28, 1988) describes compounds having retinoic acid
Claims (30)
Hide Dependent
What is claimed is:
1. A compound of the formula ##STR14## where X is S or O; R1, R2, R4 and R5 independently are hydrogen or lower alkyl of 1 to 3 carbons, and R3 is hydrogen or lower alkyl;
A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, (CH2)n and where n is an integer from 0-5;
R6 is hydrogen, lower alkyl, lower alkenyl or lower cycloalkyl having 1 to 6 carbons, or halogen; and
B is COOH or a pharmaceutically acceptable salt thereof, COOR8, CONR9 R10, where R8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R8 is phenyl or lower alkylphenyl, R9 and R10 independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl.
2. A compound of claim 1 where X is S.
3. A compound of claim 2 where A is (CH2)n and n is 0, 1, or 2.
4. A compound of claim 3 where R4 is H and R5 is lower alkyl.
5. A compound of claim 3 where R4 is the same alkyl group as R5.
6. A compound of claim 3 where R4 is lower alkyl and R5 is lower alkyl and R4 and R5 are different.
7. A compound of claim 3 where R4 and R5 are both hydrogen.
8. A compound of claim 1 where X is O.
9. A compound of claim 8 where A is (CH2)n and n is 0, 1, or 2.
10. A compound of claim 9 where R4 is H and R5 is lower alkyl.
11. A compound of claim 9 where R4 is lower alkyl and R5 is lower alkyl and R4 and R5 are different.
This is an example of an independent patent claim. It has no reference to another claim so is therefore classified as an independent claim. In Google Patents independent claims are displayed in black text.
This is an example of a dependent claim. It references another claim (claim 1) so is therefore classified as a dependent claim. In Google Patents independent claims are displayed in grey text.
Type here to search
ENG
14:21
17/11/2019

and claimed in the aforesaid co-pending applications have significant retinoic acid-like activity.

A published European patent application of the present applicant (Publication No. 0284288, published on Sep. 28, 1988) describes compounds having retinoic acid like activity which are 4,4 disubstituted chroman-6-yl, 4,4 disubstituted--thiochroman-6-yl acetylenes also substituted by a substituted heteroaryl group.

Retinoic acid-like activity has been generally recognized in the art to be associated with useful biological activity. Specifically, compounds having retinoic acid-like activity are useful as regulators of cell proliferation and differentiation, and particularly as agents for treating dermatoses, such as acne, Darier's disease, psoriasis, icthyosis, eczema, atopic dermatitis and epithelial cancers, for treating arthritic diseases and other immunological disorders (e.g. lupus erythematosus) for promoting wound healing, for treating dry eye syndrome and for reversing and preventing the effects of sun damage to skin.

With respect to the synthetic processes of the present invention which involve either the formation of an acetylenic (ethynyl) function in the compounds of the invention, or the coupling of the compounds of the invention which already have the ethynyl function with a halogen substituted phenyl group, the following articles comprise background information: A General Synthesis of Terminal and Internal Arylalkynes by the Palladium-Catalyzed Reaction of Alkynylzinc Reagents with Aryl Halides by Anthony O. King and Ei-ichi Negishi, J. Org. Chem. 43 1978 p 358; Conversion of Methyl Ketones into Terminal Acetylenes and (E)-Trisubstituted Olefins of Terpenoid Origin by Ei-ichi, Anthony O. King, and William L. Klima, J. Org. Chem. 45 1980 p.2526, and A Convenient Synthesis of Ethynylarenes and Diethynylarenes by S. Takahashi, Y. Kuroyama, K. Sonogashira, N. Hagihara, Synthesis 1980 p 627-630.

SUMMARY OF THE INVENTION

This invention covers compounds of Formula 1 ##STR2## wherein X is S, O; R_1 -R_5 are hydrogen or lower alkyl; R_6 is lower alkyl, lower alkenyl, lower cycloalkyl having 1 to 6 carbons, or halogen; A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ where n is 0-5; B is hydrogen, COOH or a pharmaceutically acceptable salt thereof, $COOR_8$, $CONR_9R_{10}$, --CH_2OH, CH_2OR_{11}, CH_2OCOR_{11}, CHO, $CH(OR_{12})_2$, $CHOR_{13}O$, --COR", $CR"(OR_{12})_2$, or $CR"OR_{13}O$, where R" is an alkyl, cycloalkyl or alkenyl group containing 1 to 5 carbons, R_8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R_8 is phenyl or lower alkylphenyl, R_9 and R_{10} independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl, R_{11} is lower alkyl, phenyl or lower alkylphenyl, R_{12} is lower alkyl, R_{13} is divalent alkyl radical of 2-5 carbons.

In a second aspect, this invention relates to the use of the compounds of Formula 1 for treating dermatoses, such as ache, Darier's disease, psoriasis, icthyosis, eczema, atopic dermatitis and epithelial cancers. These compounds are also

10. A compound of claim 9 where R_4 is H and R_5 is lower alkyl.

11. A compound of claim 9 where R_4 is lower alkyl and R_5 is lower alkyl and R_4 and R_5 are different.

12. A compound of claim 9 where R_4 is the same alkyl group as R_5.

13. A compound of claim 10 where R_4 and R_5 are both H.

14. A pharmaceutical composition comprising one or more compounds set forth in claim 1, the composition including a pharmaceutically acceptable excipient.

15. A compound of the formula ##STR15## where R_1, R_2, R_4 and R_5 independently are hydrogen or lower alkyl of 1 to 3 carbons and R_3 is hydrogen or lower alkyl;

A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ where n is an integer between 0 to 5; and

B is COOH or a pharmaceutically acceptable salt thereof, $COOR_8$, $CONR_9R_{10}$, where R_8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R_8 is phenyl or lower alkylphenyl, R_9 and R_{10} independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl.

16. A compound of claim 15 wherein A is $(CH_2)_n$ and n is 0.

17. A compound of claim 16 where R_1 -R_5 are independently H or methyl.

18. A compound of claim 17 where B is $COOC_2H_5$.

19. The compound of claim 18 where R_3 is H and R_1, R_2, R_4 and R_5 are methyl.

20. The compound of claim 18 where R_1, R_2 are methyl and R_3, R_4 and R_5 are hydrogen.

21. A compound of claim 17 where B is COOH.

22. A compound of claim 21 where R_3 is H and R_1, R_2, R_4 and R_5 are methyl.

23. A compound of claim 21 where R_1 and R_2 are methyl and R_3, R_4 and R_5 are H.

24. The compound of claim 16 where B is COOH or a pharmaceutically acceptable salt thereof, R_1, R_2 are methyl, and R_3, R_4 and R_5 are

(e.g. lupus erythematosus), in promoting wound healing, in treating dry eye syndrome and in reversing the effects of sun damage to skin.

This invention also relates to a pharmaceutical formulation comprising a compound of Formula 1 in admixture with a pharmaceutically acceptable excipient.

In another aspect, this invention relates to the process for making a compound of Formula 1 which process comprises reacting a compound of Formula 2 with a compound of Formula 3 in the presence of cuprous iodide and $Pd(PQ_3)_2 Cl_2$ (Q is phenyl) or a similar complex ##STR3## where R_1 -R_6 are the same as described above, X' is a halogen, preferably I; and A is the same as defined above; and B is H, or a protected acid, alcohol, aldehyde or ketone, giving the corresponding compound of Formula 1; or to the process of making a compound of Formula 1 which consists of reacting a zinc salt of Formula 4 with a compound of Formula 3 in the presence of $Pd(PQ_3)_4$ (Q is phenyl) or a similar complex. ##STR4## where R_1 -R_6, and X, are the same as defined above, giving the corresponding compound of Formula 1; or homologating a compound of the Formula 5 ##STR5## where n is 0-4 to give an acid of Formula 1; or converting an acid of Formula 1 to a salt; or

forming an acid addition salt;

converting an acid of Formula 1 to an ester; or

converting an acid of Formula 1 to an amide; or

reducing an acid of Formula 1 to an alcohol or aldehyde; or

converting an alcohol of Formula 1 to an ether or ester; or

oxidizing an alcohol of Formula 1 to an aldehyde; or

converting an aldehyde of Formula 1 to an acetal; or

converting a ketone of Formula 1 to a ketal.

GENERAL EMBODIMENTS

Definitions

The term "ester" as used here refers to and covers any compound falling within the definition of that term as classically used in organic chemistry. Where B (of Formula 1) is --COOH, this term covers the products derived from treatment of this function with alcohols, preferably with aliphatic alcohols having 1-6 carbons. Where the ester is derived from compounds where B is $--CH_2$ OH, this term covers compounds of the formula $--CH_2$ OOCR where R is any substituted or unsubstituted aliphatic, aromatic or aliphatic-aromatic group, preferably with 1-6 carbons in the aliphatic portions.

Preferred esters are derived from the saturated aliphatic alcohols or acids of ten or fewer carbon atoms or the cyclic or saturated aliphatic cyclic alcohols and acids of 5 to 10 carbon atoms. Particularly preferred aliphatic esters are those derived from lower alkyl acids or alcohols. Here, and where ever else used, lower alkyl means having 1-6 carbon atoms and includes straight as well as branched chain alkyl

25. A compound of the formula ##STR16## where R_1, R_2, R_4 and R_5 independently are hydrogen or lower alkyl of 1 to 3 carbons and R_3 is hydrogen or lower alkyl;

A is lower branched chain alkyl having 2 to 6 carbons, cycloalkyl having 3 to 6 carbons, alkenyl having 2 to 6 carbons and 1 or 2 double bonds, alkynyl having 2 to 6 carbons and 1 or 2 triple bonds, $(CH_2)_n$ where n is an integer between 0 to 5, and

B is COOH or a pharmaceutically acceptable salt thereof, $COOR_8$, $CONR_9R_{10}$, where R_8 is an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or R_8 is phenyl or lower alkylphenyl, R_9 and R_{10} independently are hydrogen, an alkyl group of 1 to 10 carbons, or a cycloalkyl group of 5 to 10 carbons, or phenyl or lower alkylphenyl.

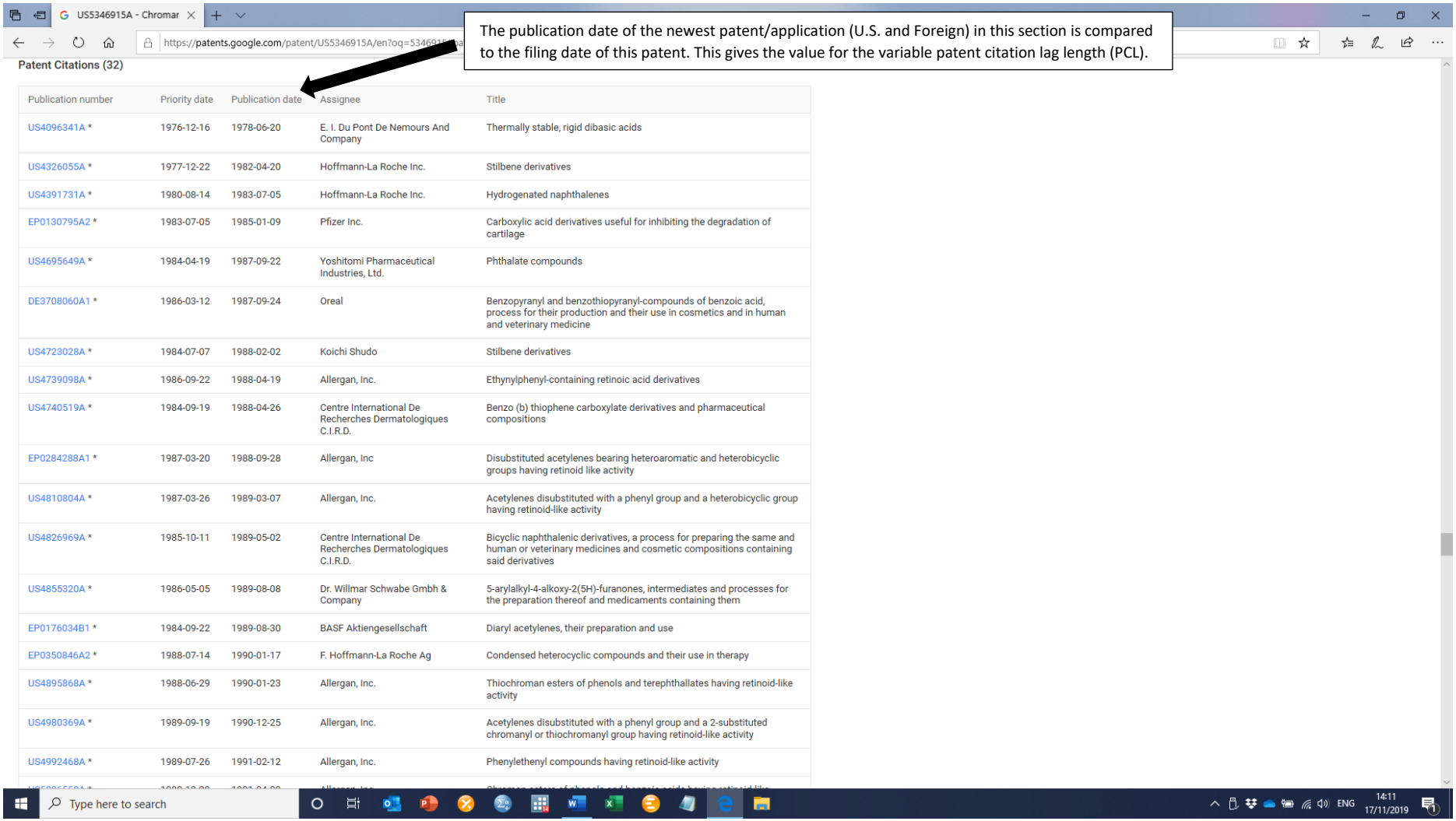
26. A compound of claim 25 where A is $(CH_2)_n$ and n is 0.

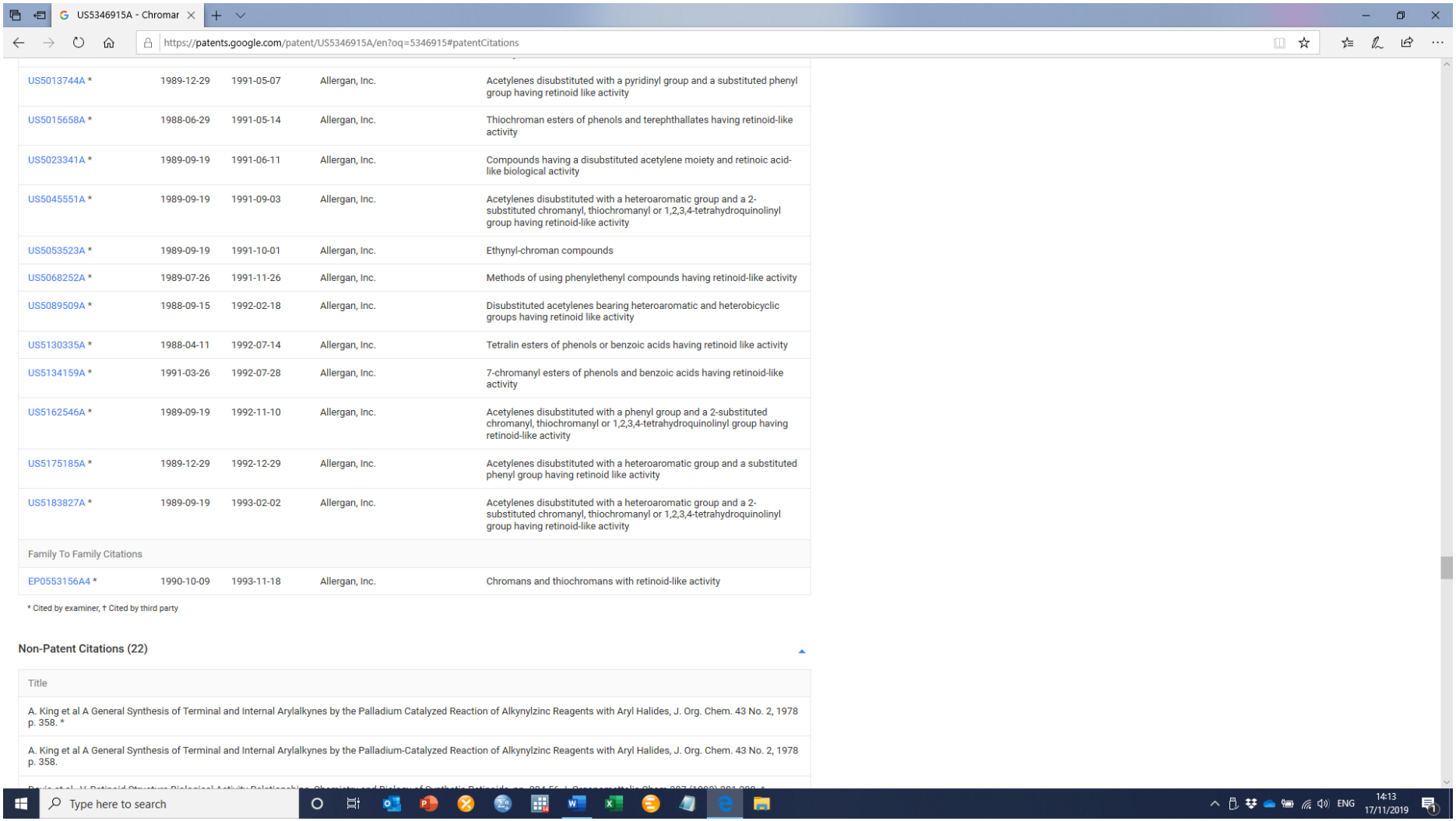
27. A compound of claim 26 where B is $COOC_2H_5$.

28. The compound of claim 27 where R_3 is hydrogen, R_1, R_2, R_4 and R_5 are methyl.

29. A compound of claim 26 where B is COOH.

30. A compound of claim 29 where R_3 is hydrogen, R_1, R_2, R_4 and R_5 are methyl.


US5346915A - Chromar
https://patents.google.com/patent/US5346915A/en?oq=5346915
The publication date of the newest patent/application (U.S. and Foreign) in this section is compared to the filing date of this patent. This gives the value for the variable patent citation lag length (PCL).
Patent Citations (32)
Publication number | Priority date | Publication date | Assignee | Title
US4096341A * | 1976-12-16 | 1978-06-20 | E. I. Du Pont De Nemours And Company | Thermally stable, rigid dibasic acids
US4326055A * | 1977-12-22 | 1982-04-20 | Hoffmann-La Roche Inc. | Stilbene derivatives
US4391731A * | 1980-08-14 | 1983-07-05 | Hoffmann-La Roche Inc. | Hydrogenated naphthalenes
EP0130795A2 * | 1983-07-05 | 1985-01-09 | Pfizer Inc. | Carboxylic acid derivatives useful for inhibiting the degradation of cartilage
US4695649A * | 1984-04-19 | 1987-09-22 | Yoshitomi Pharmaceutical Industries, Ltd. | Phthalate compounds
DE3708060A1 * | 1986-03-12 | 1987-09-24 | Oreal | Benzopyranyl and benzothiopyranyl-compounds of benzoic acid, process for their production and their use in cosmetics and in human and veterinary medicine
US4723028A * | 1984-07-07 | 1988-02-02 | Koichi Shudo | Stilbene derivatives
US4739098A * | 1986-09-22 | 1988-04-19 | Allergan, Inc. | Ethynylphenyl-containing retinoic acid derivatives
US4740519A * | 1984-09-19 | 1988-04-26 | Centre International De Recherches Dermatologiques C.I.R.D. | Benzo (b) thiophene carboxylate derivatives and pharmaceutical compositions
EP0284288A1 * | 1987-03-20 | 1988-09-28 | Allergan, Inc | Disubstituted acetylenes bearing heteroaromatic and heterobicyclic groups having retinoid like activity
US4810804A * | 1987-03-26 | 1989-03-07 | Allergan, Inc. | Acetylenes disubstituted with a phenyl group and a heterobicyclic group having retinoid-like activity
US4826969A * | 1985-10-11 | 1989-05-02 | Centre International De Recherches Dermatologiques C.I.R.D. | Bicyclic naphthalenic derivatives, a process for preparing the same and human or veterinary medicines and cosmetic compositions containing said derivatives
US4855320A * | 1986-05-05 | 1989-08-08 | Dr. Willmar Schwabe Gmbh & Company | 5-arylalkyl-4-alkoxy-2(5H)-furanones, intermediates and processes for the preparation thereof and medicaments containing them
EP0176034B1 * | 1984-09-22 | 1989-08-30 | BASF Aktiengesellschaft | Diaryl acetylenes, their preparation and use
EP0350846A2 * | 1988-07-14 | 1990-01-17 | F. Hoffmann-La Roche Ag | Condensed heterocyclic compounds and their use in therapy
US4895868A * | 1988-06-29 | 1990-01-23 | Allergan, Inc. | Thiochroman esters of phenols and terephthallates having retinoid-like activity
US4980369A * | 1989-09-19 | 1990-12-25 | Allergan, Inc. | Acetylenes disubstituted with a phenyl group and a 2-substituted chromanyl or thiochromanyl group having retinoid-like activity
US4992468A * | 1989-07-26 | 1991-02-12 | Allergan, Inc. | Phenylethenyl compounds having retinoid-like activity
Type here to search
ENG
14:11
17/11/2019

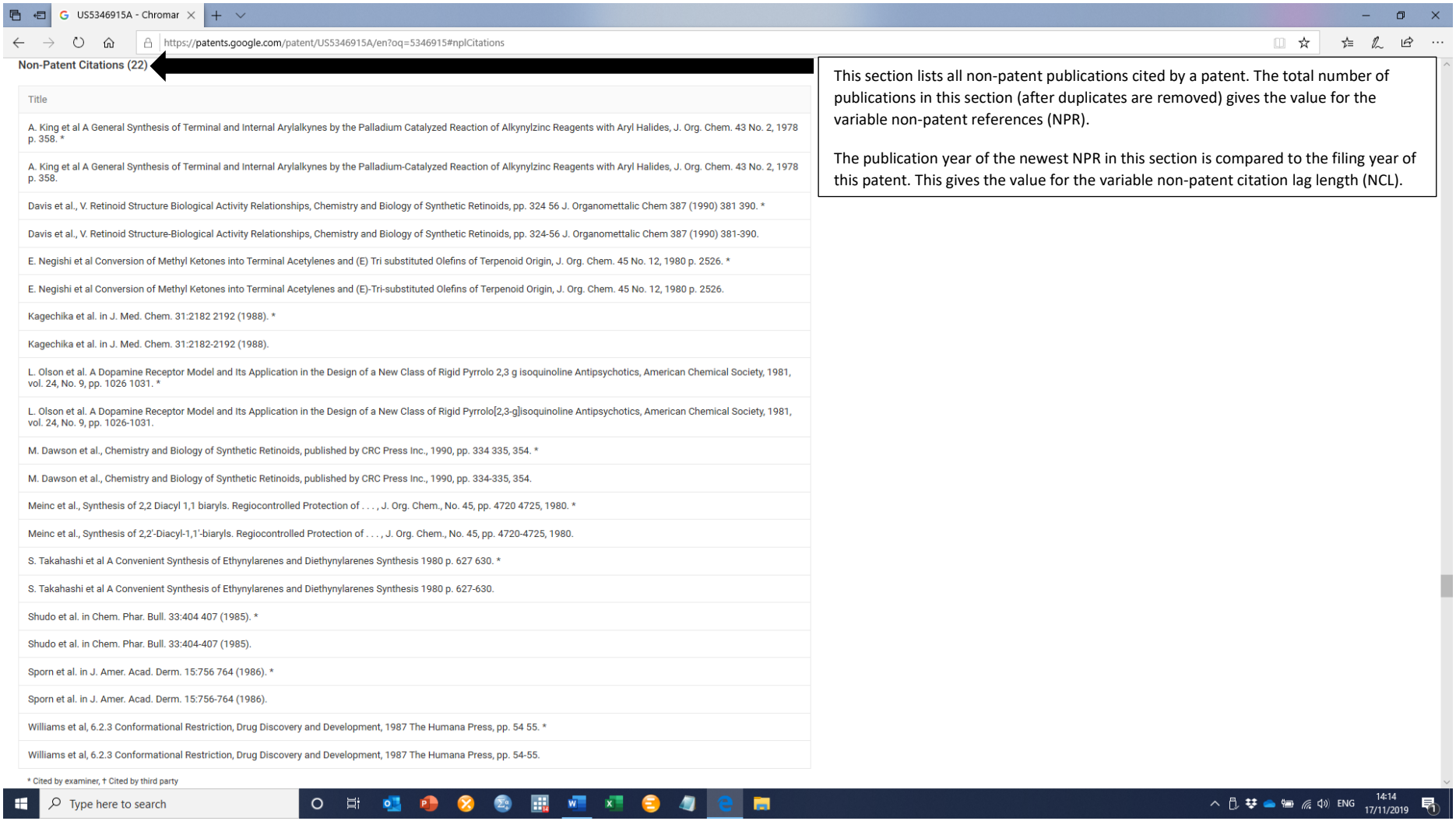


US5013744A *	1989-12-29	1991-05-07	Allergan, Inc.	Acetylenes disubstituted with a pyridinyl group and a substituted phenyl group having retinoid like activity
US5015658A *	1988-06-29	1991-05-14	Allergan, Inc.	Thiochroman esters of phenols and terephthallates having retinoid-like activity
US5023341A *	1989-09-19	1991-06-11	Allergan, Inc.	Compounds having a disubstituted acetylene moiety and retinoic acid-like biological activity
US5045551A *	1989-09-19	1991-09-03	Allergan, Inc.	Acetylenes disubstituted with a heteroaromatic group and a 2-substituted chromanyl, thiochromanyl or 1,2,3,4-tetrahydroquinolinyl group having retinoid-like activity
US5053523A *	1989-09-19	1991-10-01	Allergan, Inc.	Ethynyl-chroman compounds
US5068252A *	1989-07-26	1991-11-26	Allergan, Inc.	Methods of using phenylethenyl compounds having retinoid-like activity
US5089509A *	1988-09-15	1992-02-18	Allergan, Inc.	Disubstituted acetylenes bearing heteroaromatic and heterobicyclic groups having retinoid like activity
US5130335A *	1988-04-11	1992-07-14	Allergan, Inc.	Tetralin esters of phenols or benzoic acids having retinoid like activity
US5134159A *	1991-03-26	1992-07-28	Allergan, Inc.	7-chromanyl esters of phenols and benzoic acids having retinoid-like activity
US5162546A *	1989-09-19	1992-11-10	Allergan, Inc.	Acetylenes disubstituted with a phenyl group and a 2-substituted chromanyl, thiochromanyl or 1,2,3,4-tetrahydroquinolinyl group having retinoid-like activity
US5175185A *	1989-12-29	1992-12-29	Allergan, Inc.	Acetylenes disubstituted with a heteroaromatic group and a substituted phenyl group having retinoid like activity
US5183827A *	1989-09-19	1993-02-02	Allergan, Inc.	Acetylenes disubstituted with a heteroaromatic group and a 2-substituted chromanyl, thiochromanyl or 1,2,3,4-tetrahydroquinolinyl group having retinoid-like activity
Family To Family Citations				
EP0553156A4 *	1990-10-09	1993-11-18	Allergan, Inc.	Chromans and thiochromans with retinoid-like activity

* Cited by examiner, † Cited by third party

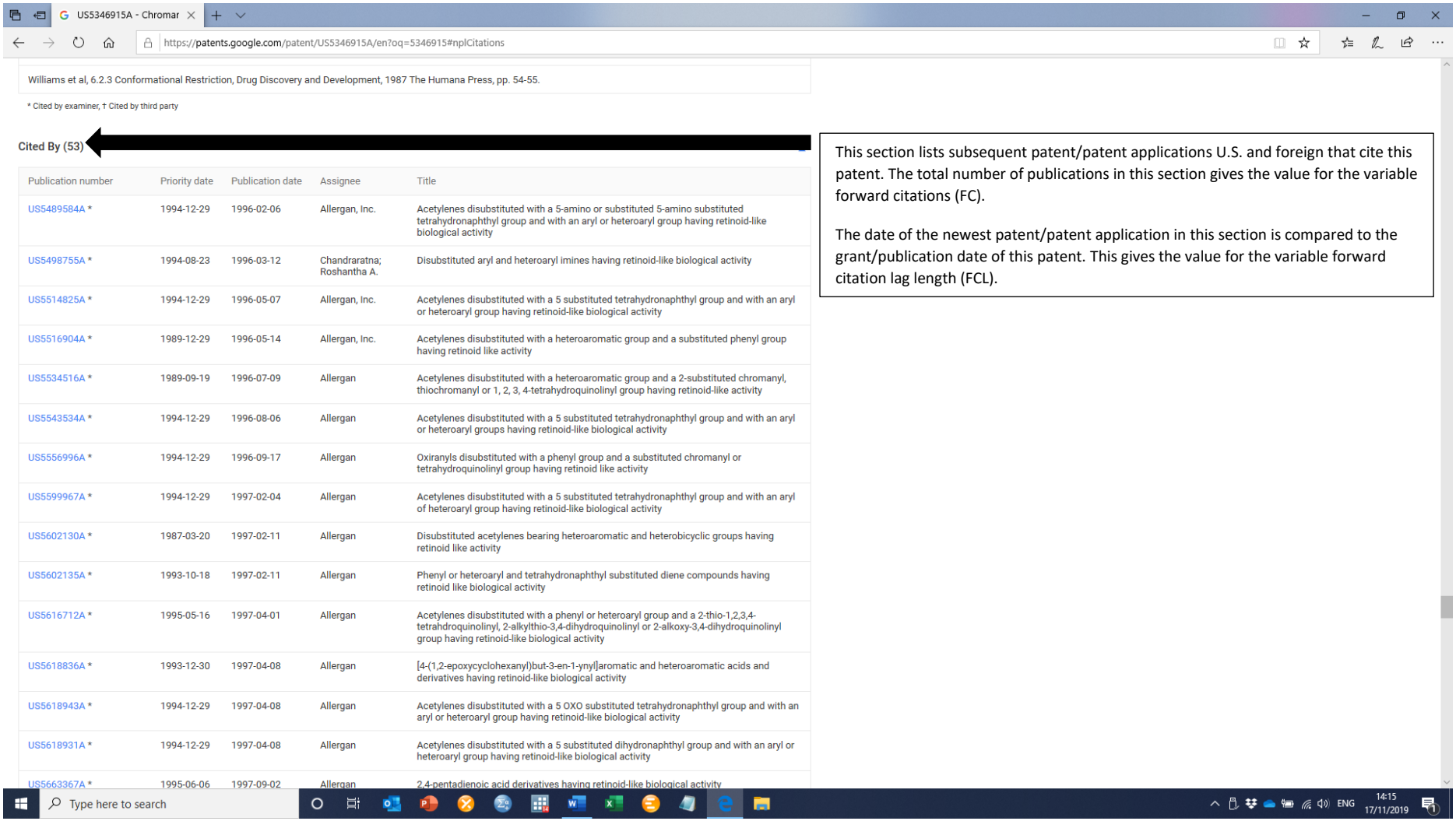
Non-Patent Citations (22)

Title
A. King et al A General Synthesis of Terminal and Internal Arylalkynes by the Palladium Catalyzed Reaction of Alkynylzinc Reagents with Aryl Halides, J. Org. Chem. 43 No. 2, 1978 p. 358. *
A. King et al A General Synthesis of Terminal and Internal Arylalkynes by the Palladium-Catalyzed Reaction of Alkynylzinc Reagents with Aryl Halides, J. Org. Chem. 43 No. 2, 1978 p. 358.



This section lists all non-patent publications cited by a patent. The total number of publications in this section (after duplicates are removed) gives the value for the variable non-patent references (NPR).

The publication year of the newest NPR in this section is compared to the filing year of this patent. This gives the value for the variable non-patent citation lag length (NCL).



Williams et al, 6.2.3 Conformational Restriction, Drug Discovery and Development, 1987 The Humana Press, pp. 54-55.

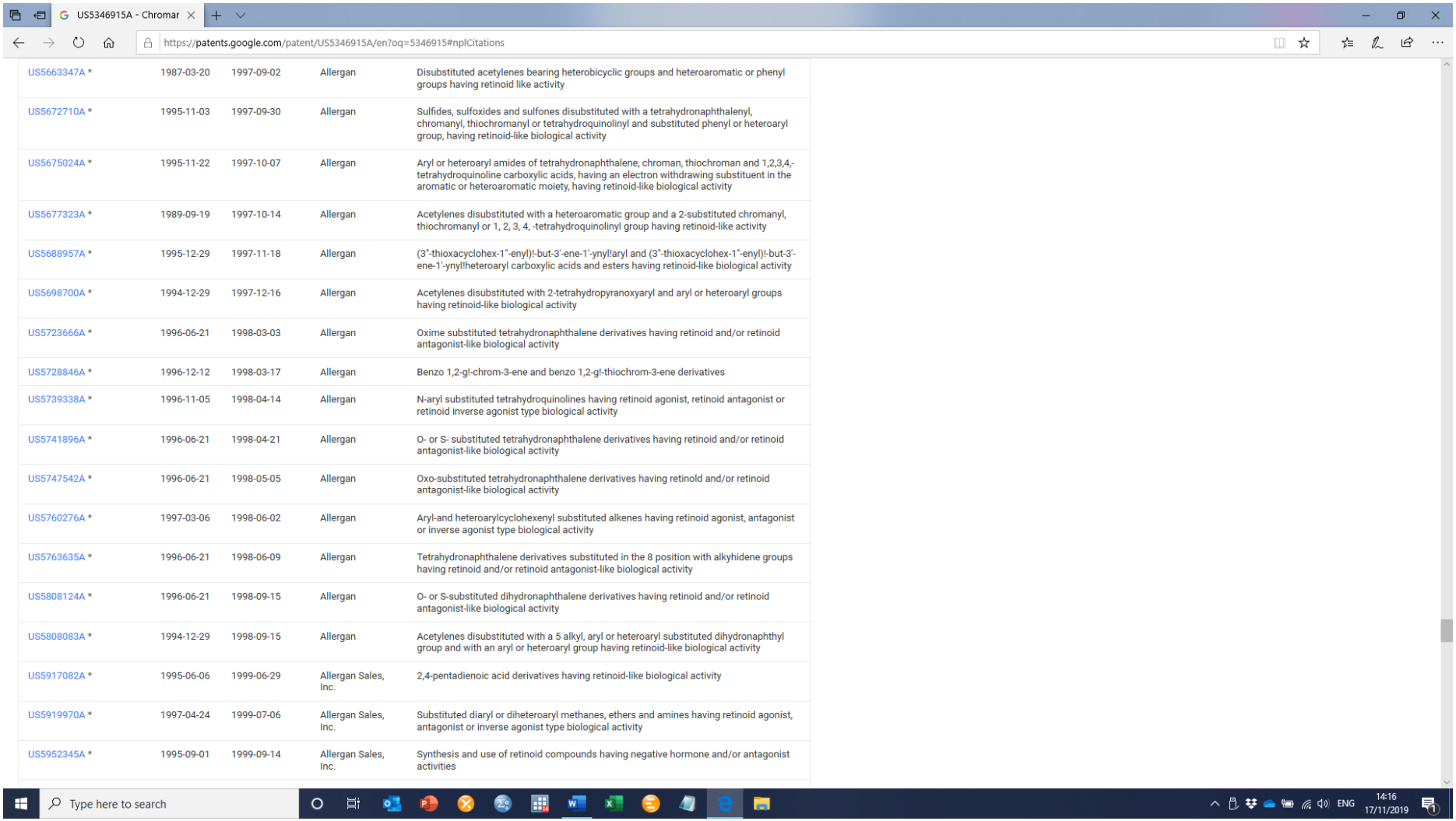
* Cited by examiner, † Cited by third party

Cited By (53)

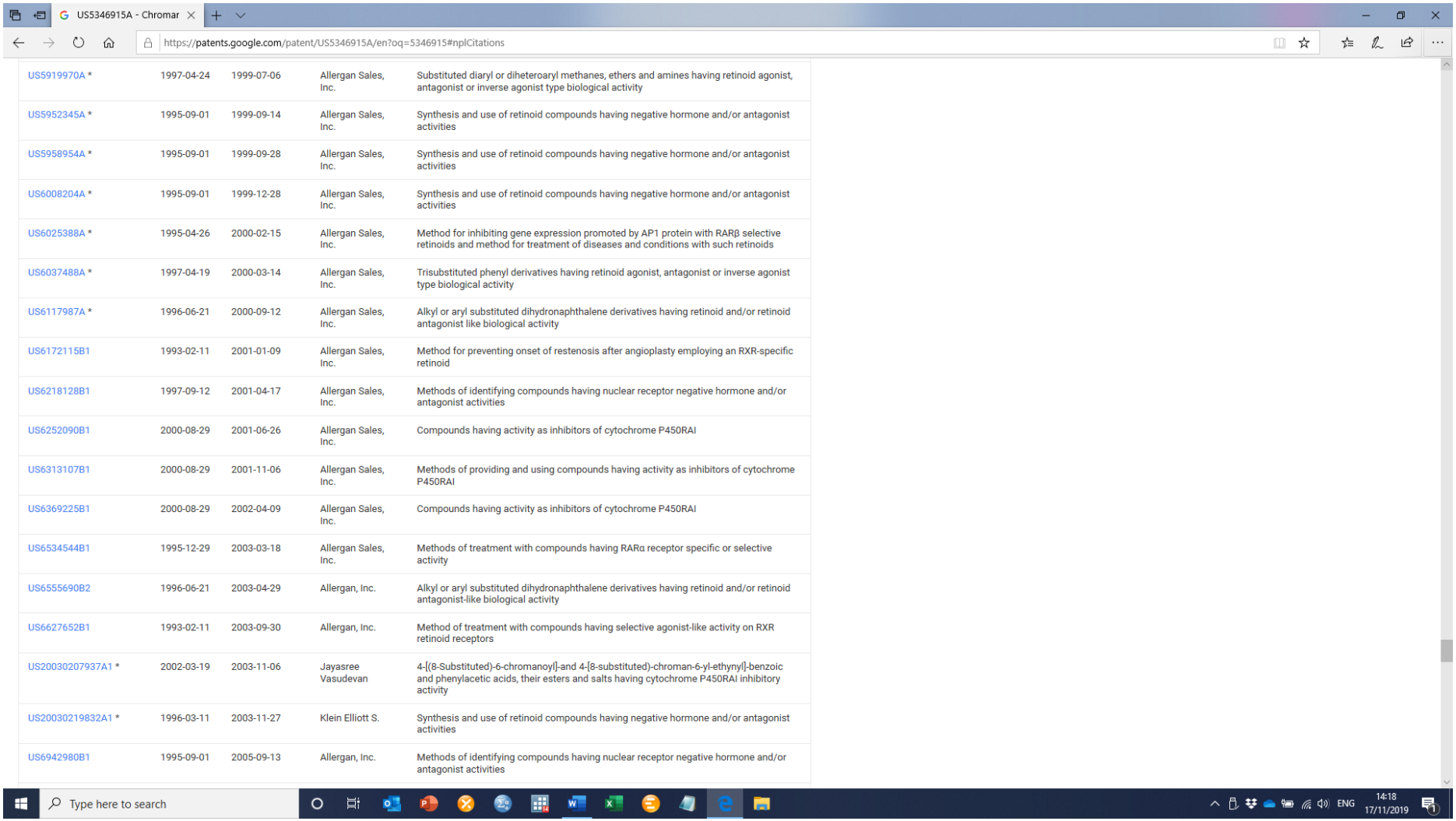
Publication number	Priority date	Publication date	Assignee	Title
US5489584A *	1994-12-29	1996-02-06	Allergan, Inc.	Acetylenes disubstituted with a 5-amino or substituted 5-amino substituted tetrahydronaphthyl group and with an aryl or heteroaryl group having retinoid-like biological activity
US5498755A *	1994-08-23	1996-03-12	Chandraratna; Roshantha A.	Disubstituted aryl and heteroaryl imines having retinoid-like biological activity
US5514825A *	1994-12-29	1996-05-07	Allergan, Inc.	Acetylenes disubstituted with a 5 substituted tetrahydronaphthyl group and with an aryl or heteroaryl group having retinoid-like biological activity
US5516904A *	1989-12-29	1996-05-14	Allergan, Inc.	Acetylenes disubstituted with a heteroaromatic group and a substituted phenyl group having retinoid like activity
US5534516A *	1989-09-19	1996-07-09	Allergan	Acetylenes disubstituted with a heteroaromatic group and a 2-substituted chromanyl, thiochromanyl or 1, 2, 3, 4-tetrahydroquinolinyl group having retinoid-like activity
US5543534A *	1994-12-29	1996-08-06	Allergan	Acetylenes disubstituted with a 5 substituted tetrahydronaphthyl group and with an aryl or heteroaryl groups having retinoid-like biological activity
US5556996A *	1994-12-29	1996-09-17	Allergan	Oxiranyls disubstituted with a phenyl group and a substituted chromanyl or tetrahydroquinolinyl group having retinoid like activity
US5599967A *	1994-12-29	1997-02-04	Allergan	Acetylenes disubstituted with a 5 substituted tetrahydronaphthyl group and with an aryl of heteroaryl group having retinoid-like biological activity
US5602130A *	1987-03-20	1997-02-11	Allergan	Disubstituted acetylenes bearing heteroaromatic and heterobicyclic groups having retinoid like activity
US5602135A *	1993-10-18	1997-02-11	Allergan	Phenyl or heteroaryl and tetrahydronaphthyl substituted diene compounds having retinoid like biological activity
US5616712A *	1995-05-16	1997-04-01	Allergan	Acetylenes disubstituted with a phenyl or heteroaryl group and a 2-thio-1,2,3,4-tetrahdroquinolinyl, 2-alkylthio-3,4-dihydroquinolinyl or 2-alkoxy-3,4-dihydroquinolinyl group having retinoid-like biological activity
US5618836A *	1993-12-30	1997-04-08	Allergan	[4-(1,2-epoxycyclohexanyl)but-3-en-1-ynyl]aromatic and heteroaromatic acids and derivatives having retinoid-like biological activity
US5618943A *	1994-12-29	1997-04-08	Allergan	Acetylenes disubstituted with a 5 OXO substituted tetrahydronaphthyl group and with an aryl or heteroaryl group having retinoid-like biological activity
US5618931A *	1994-12-29	1997-04-08	Allergan	Acetylenes disubstituted with a 5 substituted dihydronaphthyl group and with an aryl or heteroaryl group having retinoid-like biological activity
US5663367A *	1995-06-06	1997-09-02	Allergan	2,4-pentadienoic acid derivatives having retinoid-like biological activity

This section lists subsequent patent/patent applications U.S. and foreign that cite this patent. The total number of publications in this section gives the value for the variable forward citations (FC).

The date of the newest patent/patent application in this section is compared to the grant/publication date of this patent. This gives the value for the variable forward citation lag length (FCL).



US5663347A *	1987-03-20	1997-09-02	Allergan	Disubstituted acetylenes bearing heterobicyclic groups and heteroaromatic or phenyl groups having retinoid like activity
US5672710A *	1995-11-03	1997-09-30	Allergan	Sulfides, sulfoxides and sulfones disubstituted with a tetrahydronaphthalenyl, chromanyl, thiochromanyl or tetrahydroquinolinyl and substituted phenyl or heteroaryl group, having retinoid-like biological activity
US5675024A *	1995-11-22	1997-10-07	Allergan	Aryl or heteroaryl amides of tetrahydronaphthalene, chroman, thiochroman and 1,2,3,4,-tetrahydroquinoline carboxylic acids, having an electron withdrawing substituent in the aromatic or heteroaromatic moiety, having retinoid-like biological activity
US5677323A *	1989-09-19	1997-10-14	Allergan	Acetylenes disubstituted with a heteroaromatic group and a 2-substituted chromanyl, thiochromanyl or 1, 2, 3, 4, -tetrahydroquinolinyl group having retinoid-like activity
US5688957A *	1995-12-29	1997-11-18	Allergan	(3'-thioxacyclohex-1"-enyl)!-but-3'-ene-1'-ynyl!aryl and (3"-thioxacyclohex-1"-enyl)!-but-3'-ene-1'-ynyl!heteroaryl carboxylic acids and esters having retinoid-like biological activity
US5698700A *	1994-12-29	1997-12-16	Allergan	Acetylenes disubstituted with 2-tetrahydropyranoxyaryl and aryl or heteroaryl groups having retinoid-like biological activity
US5723666A *	1996-06-21	1998-03-03	Allergan	Oxime substituted tetrahydronaphthalene derivatives having retinoid and/or retinoid antagonist-like biological activity
US5728846A *	1996-12-12	1998-03-17	Allergan	Benzo 1,2-g!-chrom-3-ene and benzo 1,2-g!-thiochrom-3-ene derivatives
US5739338A *	1996-11-05	1998-04-14	Allergan	N-aryl substituted tetrahydroquinolines having retinoid agonist, retinoid antagonist or retinoid inverse agonist type biological activity
US5741896A *	1996-06-21	1998-04-21	Allergan	O- or S- substituted tetrahydronaphthalene derivatives having retinoid and/or retinoid antagonist-like biological activity
US5747542A *	1996-06-21	1998-05-05	Allergan	Oxo-substituted tetrahydronaphthalene derivatives having retinold and/or retinoid antagonist-like biological activity
US5760276A *	1997-03-06	1998-06-02	Allergan	Aryl-and heteroarylcyclohexenyl substituted alkenes having retinoid agonist, antagonist or inverse agonist type biological activity
US5763635A *	1996-06-21	1998-06-09	Allergan	Tetrahydronaphthalene derivatives substituted in the 8 position with alkyhidene groups having retinoid and/or retinoid antagonist-like biological activity
US5808124A *	1996-06-21	1998-09-15	Allergan	O- or S-substituted dihydronaphthalene derivatives having retinoid and/or retinoid antagonist-like biological activity
US5808083A *	1994-12-29	1998-09-15	Allergan	Acetylenes disubstituted with a 5 alkyl, aryl or heteroaryl substituted dihydronaphthyl group and with an aryl or heteroaryl group having retinoid-like biological activity
US5917082A *	1995-06-06	1999-06-29	Allergan Sales, Inc.	2,4-pentadienoic acid derivatives having retinoid-like biological activity
US5919970A *	1997-04-24	1999-07-06	Allergan Sales, Inc.	Substituted diaryl or diheteroaryl methanes, ethers and amines having retinoid agonist, antagonist or inverse agonist type biological activity
US5952345A *	1995-09-01	1999-09-14	Allergan Sales, Inc.	Synthesis and use of retinoid compounds having negative hormone and/or antagonist activities



US5919970A *	1997-04-24	1999-07-06	Allergan Sales, Inc.	Substituted diaryl or diheteroaryl methanes, ethers and amines having retinoid agonist, antagonist or inverse agonist type biological activity
US5952345A *	1995-09-01	1999-09-14	Allergan Sales, Inc.	Synthesis and use of retinoid compounds having negative hormone and/or antagonist activities
US5958954A *	1995-09-01	1999-09-28	Allergan Sales, Inc.	Synthesis and use of retinoid compounds having negative hormone and/or antagonist activities
US6008204A *	1995-09-01	1999-12-28	Allergan Sales, Inc.	Synthesis and use of retinoid compounds having negative hormone and/or antagonist activities
US6025388A *	1995-04-26	2000-02-15	Allergan Sales, Inc.	Method for inhibiting gene expression promoted by AP1 protein with RARβ selective retinoids and method for treatment of diseases and conditions with such retinoids
US6037488A *	1997-04-19	2000-03-14	Allergan Sales, Inc.	Trisubstituted phenyl derivatives having retinoid agonist, antagonist or inverse agonist type biological activity
US6117987A *	1996-06-21	2000-09-12	Allergan Sales, Inc.	Alkyl or aryl substituted dihydronaphthalene derivatives having retinoid and/or retinoid antagonist like biological activity
US6172115B1	1993-02-11	2001-01-09	Allergan Sales, Inc.	Method for preventing onset of restenosis after angioplasty employing an RXR-specific retinoid
US6218128B1	1997-09-12	2001-04-17	Allergan Sales, Inc.	Methods of identifying compounds having nuclear receptor negative hormone and/or antagonist activities
US6252090B1	2000-08-29	2001-06-26	Allergan Sales, Inc.	Compounds having activity as inhibitors of cytochrome P450RAI
US6313107B1	2000-08-29	2001-11-06	Allergan Sales, Inc.	Methods of providing and using compounds having activity as inhibitors of cytochrome P450RAI
US6369225B1	2000-08-29	2002-04-09	Allergan Sales, Inc.	Compounds having activity as inhibitors of cytochrome P450RAI
US6534544B1	1995-12-29	2003-03-18	Allergan Sales, Inc.	Methods of treatment with compounds having RARα receptor specific or selective activity
US6555690B2	1996-06-21	2003-04-29	Allergan, Inc.	Alkyl or aryl substituted dihydronaphthalene derivatives having retinoid and/or retinoid antagonist-like biological activity
US6627652B1	1993-02-11	2003-09-30	Allergan, Inc.	Method of treatment with compounds having selective agonist-like activity on RXR retinoid receptors
US20030207937A1 *	2002-03-19	2003-11-06	Jayasree Vasudevan	4-[(8-Substituted)-6-chromanoyl]-and 4-[8-substituted)-chroman-6-yl-ethynyl]-benzoic and phenylacetic acids, their esters and salts having cytochrome P450RAI inhibitory activity
US20030219832A1 *	1996-03-11	2003-11-27	Klein Elliott S.	Synthesis and use of retinoid compounds having negative hormone and/or antagonist activities
US6942980B1	1995-09-01	2005-09-13	Allergan, Inc.	Methods of identifying compounds having nuclear receptor negative hormone and/or antagonist activities

References

Aboody, David, and Baruch Lev, 1998, The Value Relevance of Intangibles: The Case of Software Capitalization, *Journal of Accounting Research* 36, 161-191.

Acs, Zoltan J., and David B. Audretsch, 1989, Patents as a Measure of Innovative Activity, *Kyklos* 42, 171-180.

Aggarwal, Vikas A., and David H. Hsu, 2014, Entrepreneurial Exits and Innovation, *Management Science* 60, 867-887.

Agrawal, Ajay K., 2001, University-to-Industry Knowledge Transfer: Literature Review and Unanswered Questions, *International Journal of Management Reviews* 3, 285-302.

Ahuja, Gautam, and Curba Morris Lampert, 2001, Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions, *Strategic Management Journal* 22, 521-543.

Albert, Michael B., Daniel Avery, Francis Narin, and Paul McAllister, 1991, Direct Validation of Citation Counts as Indicators of Industrially Important Patents, *Research Policy* 20, 251-259.

Allison, John R., Mark A. Lemley, Kimberly A. Moore, and R. Derek Trunkey, 2004, Valuable Patents, *Georgetown Law Journal* 92, 435.

Allison, John R., Mark A. Lemley, and Joshua Walker, 2011, Patent Quality and Settlement Among Repeat Patent Litigants, *Georgetown Law Journal* 99, 677-712.

Allison, John R., and Thomas W. Sager, 2007, Valuable Patents Redux: On the Enduring Merit of Using Patent Characteristics to Identify Valuable Patents, *Texas Law Review* 85, 1769.

Allison, John R., and Emerson H. Tiller, 2003, The Business Method Patent Myth, *Berkeley Technology Law Journal* 18, 987.

Almeida, Paul, Jan Hohberger, and Pedro Parada, 2011, Individual Scientific Collaborations and Firm-level Innovation, *Industrial and Corporate Change* 20, 1571-1599.

Altman, Edward I., 1968, Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *The Journal of Finance* 23, 589-609.

Alvesson, Mats, 1995. *Cultural Perspectives on Organizations* (Cambridge University Press Archive).

Amir, Eli, and Baruch Lev, 1996, Value-relevance of Nonfinancial Information: The Wireless Communications Industry, *Journal of Accounting and Economics* 22, 3-30.

Amit, Raphael, and Paul J. H. Schoemaker, 1993, Strategic Assets and Organizational Rent, *Strategic Management Journal* 14, 33-46.

Arora, Ashish, and Alfonso Gambardella, 1994, The Changing Technology of Technological Change: General and Abstract Knowledge and the Division of Innovative Labour, *Research Policy* 23, 523-532.

Arts, Sam, and Reinhilde Veugelers, 2012, The Technological Origins and Novelty of Breakthrough Inventions, Working Paper. Faculty of Economics and Business (KU Leuven, Leuven, Belgium).

Arundal, Anthony, and Isabelle Kabla, 1998, What Percentage of Innovations are Patented? Empirical Estimates for European Firms, *Research Policy* 27, 127-141.

Ashley, Richard, Clive W. J. Granger, and Richard Schmalensee, 1980, Advertising and Aggregate Consumption: An Analysis of Causality, *Econometrica: Journal of the Econometric Society* 1149-1167.

Austin, David H., 1993, An Event-study Approach to Measuring Innovative Output: The Case of Biotechnology, *The American Economic Review* 83, 253-258.

Balogun, Julia, and Mark Jenkins, 2003, Re-conceiving Change Management: A Knowledge-based Perspective, *European Management Journal* 21, 247-257.
Barber, Brad M., and John D. Lyon, 1997, Detecting Long-run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics, *Journal of Financial Economics* 43, 341-372.
Barber, Brad M., and Terrance Odean, 2007, All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *The Review of Financial Studies* 21, 785-818.
Barnes, Paul, 1998, Can Takeover Targets be Identified by Statistical Techniques? Some UK Evidence, *Journal of the Royal Statistical Society: Series D (The Statistician)* 47, 573-591.
Barnes, Paul, 1999, Predicting UK Takeover Targets: Some Methodological Issues and an Empirical Study, *Review of Quantitative Finance and Accounting* 12, 283-302.
Barnett, William P., 1990, The Organizational Ecology of a Technological System, *Administrative Science Quarterly* 31-60.
Barney, Jonathan A., 2002, A Study of Patent Mortality Rates: Using Statistical Survival Analysis to Rate and Value Patent Assets, *American Intellectual Property Law Association* (*AIPLA) Quarterly Journal* 30, 317.
Barney, Jay B., 1986, Strategic Factor Markets: Expectations, Luck, and Business Strategy, *Management Science* 32, 1231-1241.
Barney, Jay B., 1991, Firm Resources and Sustained Competitive Advantage, *Journal of Management* 17, 99-120.
Barney, Jay B., 2001, Resource-based Theories of Competitive Advantage: A Ten-year Retrospective on the Resource-based View, *Journal of Management* 27, 643-650.
Barzel, Yoram, 1968, Optimal Timing of Innovations, *The Review of Economics and Statistics* 50, 348-355.
Batelle Technology Partnership Practice, 2013, The Economic Impact of the U.S. Biopharmaceutical Industry, (Columbus, Ohio).
Baum, Joel A. C., Tony Calabrese, and Brian S. Silverman, 2000, Don't Go It Alone: Alliance Network Composition and Startups' Performance in Canadian Biotechnology, *Strategic Management Journal* 21, 267-294.
Ben-Zion, Uri, 1984, The R&D and Investment Decision and its Relationship to the Firm's Market Value: Some Preliminary Results, in *R&D, Patents, and Productivity* (University of Chicago Press).
Bierly, Paul, and Alok Chakrabarti, 1996, Generic Knowledge Strategies in the US Pharmaceutical Industry, *Strategic Management Journal* 17, 123-135.
Blundell, Richard, Rachel Griffith, and John Van Reenen, 1995, Dynamic Count Data Models of Technological Innovation, *The Economic Journal* 105, 333-344.
Blundell, Richard, Rachel Griffith, and John Van Reenen, 1999, Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms, *The Review of Economic Studies* 66, 529-554.
Bollinger, Christopher R., and Amitabh Chandra, 2005, Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data, *Journal of Labor Economics* 23, 235-257.
Bontis, Nick, 2002. *The Strategic Management of Intellectual Capital and Organizational Knowledge* (Oxford University, Oxford).
Booth, Bruce, and Rodney Zemmel, 2004, Opinion: Prospects for Productivity, *Nature Reviews Drug Discovery* 3, 451.
Bosch, Jean-Claude, and Insup Lee, 1994, Wealth Effects of Food and Drug Administration (FDA) Decisions, *Managerial and Decision Economics* 15, 589-599.
Bosworth, Derek, and Mark Rogers, 2001, Market Value, R&D and Intellectual Property: An Empirical Analysis of Large Australian Firms, *Economic Record* 77, 323-337.

Bosworth, Derek Leonard, and Hamid Mahdian, 1999, Returns to Intellectual Capital in the Pharmaceuticals Sector, *Economie Applique* 52, 69-93.
Bound, John, Clint Cummins, Zvi Griliches, Bronwyn H. Hall, and Adam B. Jaffe, 1982, Who Does R&D and Who Patents?, NBER *Working Paper* 908.
Bouwman, Christa H. S., Kathleen Fuller, and Amrita S. Nain, 2007, Market Valuation and Acquisition Quality: Empirical Evidence, *The Review of Financial Studies* 22, 633-679.
Branstetter, Lee, and Yoshiaki Ogura, 2005, Is Academic Science Driving a Surge in Industrial Innovation? Evidence from Patent Citations, NBER *Working Paper* 11561.
Breitzman, Anthony, and Patrick Thomas, 2015, Inventor Team Size as a Predictor of the Future Citation Impact of Patents, *Scientometrics* 103, 631-647.
Brooks, Chris, 2008. *Introductory Econometrics for Finance* (Cambridge University Press, Cambridge).
Callaert, Julie, Joris Grouwels, and Bart Van Looy, 2012, Delineating the Scientific Footprint in Technology: Identifying Scientific Publications within Non-patent References, *Scientometrics* 91, 383-398.
Callaert, Julie, Bart Van Looy, Arnold Verbeek, Koenraad Debackere, and Bart Thijs, 2006, Traces of Prior Art: An Analysis of Non-patent References Found in Patent Documents, *Scientometrics* 69, 3-20.
Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In Search of Distress Risk, *The Journal of Finance* 63, 2899-2939.
Carlucci, Daniela, Bernard Marr, and Gianni Schiuma, 2004, The Knowledge Value Chain: How Intellectual Capital Impacts on Business Performance, *International Journal of Technology Management* 27, 575-590.
Carneiro, Alberto, 2000, How Does Knowledge Management Influence Innovation and Competitiveness?, *Journal of Knowledge Management* 4, 87-98.
Carpenter, M., M. Cooper, and Francis Narin, 1980, Linkage Between Basic Research Literature and Patents, *Research Management* 23, 30-35.
Carpenter, Mark P., and Francis Narin, 1983, Validation Study: Patent Citations as Indicators of Science and Foreign Dependence, *World Patent Information* 5, 180-185.
Cassiman, Bruno, and Reinhilde Veugelers, 2006, In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition, *Management Science* 52, 68-82.
Cassiman, Bruno, Reinhilde Veugelers, and Pluvia Zuniga, 2008, In Search of Performance Effects of (In) Direct Industry Science Links, *Industrial and Corporate Change* 17, 611-646.
Chaney, Paul K., Timothy M. Devinney, and Russell S. Winer, 1991, The Impact of New Product Introductions on the Market Value of Firms, *Journal of Business* 64, 573-610.
Chang, K. C., J. Hao, C. Chen, and C. C. Yuan, 2014, The Determinants of Valuable Patents, Proceedings of PICMET '14 Conference: Portland International Center for Management of Engineering and Technology; Infrastructure and Service Integration.
Chang, Su-Chao, and Ming-Shing Lee, 2008, The Linkage Between Knowledge Accumulation Capability and Organizational Innovation, *Journal of Knowledge Management* 12, 3-20.
Chen, Yu-Shan, and Ke-Chiun Chang, 2010, Exploring the Nonlinear Effects of Patent Citations, Patent Share and Relative Patent Position on Market Value in the US Pharmaceutical Industry, *Technology Analysis & Strategic Management* 22, 153-169.

Chesbrough, Henry, and Richard S. Rosenbloom, 2002, The Role of the Business Model in Capturing Value from Innovation: Evidence from Xerox Corporation's Technology Spin-Off Firms, *Industrial and Corporate Change* 11, 529-555.

Chou, Shih-Wei, 2005, Knowledge Creation: Absorptive Capacity, Organizational Mechanisms, and Knowledge Storage/Retrieval Capabilities, *Journal of Information Science* 31, 453-465.

Clark, Kim B., and Takahiro Fujimoto, 1991. *Product Development Performance: Strategy, Organization, and Management in the World Auto Industry* (Harvard Business School Press, Boston, MA, Harvard).

Cockburn, Iain, and Zvi Griliches, 1988, The Estimation and Measurement of Spillover Effects of R&D Investment, *American Economic Review* 78, 419-423.

Cockburn, Iain M., and Rebecca M. Henderson, 1998, Absorptive Capacity, Coauthoring Behaviour, and the Organization of Research and Drug Discovery, *The Journal of Industrial Economics* 46, 157-182.

Cockburn, Iain M., and Rebecca M. Henderson, 2001, Scale and Scope in Drug Development: Unpacking the Advantages of Size in Pharmaceutical Research, *Journal of Health Economics* 20, 1033-1057.

Cockburn, Iain M., and Megan J. MacGarvie, 2009, Patents, Thickets, and the Financing of Early-stage Firms: Evidence from the Software Industry, *Journal of Economics & Management Strategy* 18, 729-773.

Cohen, Jacob, Patricia Cohen, S. G. West, and L. S. Aiken, 2003. *Applied Multiple Regression/Correlation Analysis for The Behavioral Sciences* (Lawrence Erlbaum, Mahwah, NJ).

Cohen, Lauren, and Dong Lou, 2012, Complicated Firms, *Journal of Financial Economics* 104, 383-400.

Cohen, Wesley M., Akira Goto, Akiya Nagata, Richard R. Nelson, and John P. Walsh, 2002, R&D Spillovers, Patents and the Incentives to Innovate in Japan and the United States, *Research Policy* 31, 1349-1367.

Cohen, Wesley M., and Daniel A. Levinthal, 1989, Innovation and Learning: The Two Faces of R&D, *The Economic Journal* 99, 569-596.

Cohen, Wesley M., and Daniel A. Levinthal, 1990, Absorptive Capacity: A New Perspective on Learning and Innovation, *Administrative Science Quarterly* 35, 128-152.

Cohen, Wesley M., and David Mowery, 1984, Firm Heterogeneity and R&D: An Agenda for Research, *Strategic Management of Industrial R&D* 107, 32.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh, 2000, Protecting their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (Or Not), NBER *Working Paper* 7522.

Collier, Roger, 2011, Bye, Bye Blockbusters, Hello Niche Busters, *Canadian Medical Association Journal* 183 (11), E697-E698.

Collins, Peter, and Suzanne Wyatt, 1988, Citations in Patents to the Basic Research Literature, *Research Policy* 17, 65-74.

Connolly, Robert A., Barry T. Hirsch, and Mark Hirschey, 1986, Union Rent Seeking, Intangible Capital, and Market Value of the Firm, *The Review of Economics and Statistics* 68, 567-577.

Connolly, Robert A., and Mark Hirschey, 1988, Market Value and Patents, *Economics Letters* 27, 83-87.

Connolly, Robert A., and Mark Hirschey, 1990, Firm Size and R&D Effectiveness, *Economics Letters* 32, 277-281.

Coombs, Joseph E., and David L. Deeds, 2000, International Alliances as Sources of Capital: Evidence from the Biotechnology Industry, *The Journal of High Technology Management Research* 11, 235-253.

Cox, D. R., 1970, *The Analysis of Binary Data* (Methuen, London).
Cramer, Jan Salomon, 1999, Predictive Performance of the Binary Logit Model in Unbalanced Samples, *Journal of the Royal Statistical Society: Series D (The Statistician)* 48, 85-94.
Criscuolo, Paola, Linus Dahlander, Thorsten Grohsjean, and Ammon Salter, 2017, Evaluating Novelty: The Role of Panels in the Selection of R&D Projects, *Academy of Management Journal* 60, 433-460.
Criscuolo, Paola, and Bart Verspagen, 2008, Does it Matter Where Patent Citations Come From? Inventor vs. Examiner Citations in European Patents, *Research Policy* 37, 1892-1908.
Czarnitzki, Dirk, 2009, The Virtue of Industry-Science Collaborations, *EIB Papers* 14, 120-143.
Czarnitzki, Dirk, Wolfgang Glänzel, and Katrin Hussinger, 2007, Patent and Publication Activities of German Professors: An Empirical Assessment of their Co-activity, *Research Evaluation* 16, 311-319.
Czarnitzki, Dirk, Bronwyn H. Hall, and Raffaele Oriani, 2006. *The Market Valuation of Knowledge Assets in US and European Firms* (Edward Elgar, Cheltenham).
Danbolt, Jo, Antonios Siganos, and Abongeh Tunyi, 2016, Abnormal Returns from Takeover Prediction Modelling: Challenges and Suggested Investment Strategies, *Journal of Business Finance & Accounting* 43, 66-97.
Darby, Michael R., Qiao Liu, and Lynne G. Zucker, 2004, High Stakes in High Technology: High-tech Market Values as Options, *Economic Inquiry* 42, 351-369.
Dasgupta, Partha, and Joseph Stiglitz, 1980, Industrial Structure and the Nature of Innovative Activity, *The Economic Journal* 90, 266-293.
De Carolis, D., 2002, The Role of Social Capital and Organizational Knowledge in Enhancing Entrepreneurial Opportunities in High-technology Environments, in Choo C. and Bontis N., Ed.: *The Strategic Management of Intellectual Capital and Organizational Knowledge* (Oxford University Press, New York).
Dedman, Elisabeth, Stephen W.-J. Lin, Arun J. Prakash, and Chun-Hao Chang, 2008, Voluntary Disclosure and its Impact on Share Prices: Evidence from the UK Biotechnology Sector, *Journal of Accounting and Public Policy* 27, 195-216.
Deng, Yi, 2008, The Value of Knowledge Spillovers in the U.S. Semiconductor Industry, *International Journal of Industrial Organization* 26, 1044-1058.
Deng, Zhen, Baruch Lev, and Francis Narin, 1999, Science and Technology as Predictors of Stock Performance, *Financial Analysts Journal* 55, 20-32.
Dierickx, Ingemar, and Karel Cool, 1989, Asset Stock Accumulation and Sustainability of Competitive Advantage, *Management Science* 35, 1504-1513.
DiMasi, Joseph A., 1995, Success Rates for New Drugs Entering Clinical Testing in the United States, *Clinical Pharmacology & Therapeutics* 58, 1-14.
DiMasi, Joseph A., 2001, Risks in New Drug Development: Approval Success Rates for Investigational Drugs, *Clinical Pharmacology & Therapeutics* 69, 297-307.
DiMasi, Joseph A., 2014, Pharmaceutical R&D Performance by Firm Size: Approval Success Rates and Economic Returns, *American Journal of Therapeutics* 21, 26-34.
DiMasi, Joseph A., Lanna Feldman, A. Seckler, and A. Wilson, 2010, Trends in Risks Associated with New Drug Development: Success Rates for Investigational Drugs, *Clinical Pharmacology & Therapeutics* 87, 272-277.
DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen, 2016, Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs, *Journal of Health Economics* 47, 20-33.
DiMasi, Joseph A., Henry G. Grabowski, and John Vernon, 1995, R&D Costs, Innovative Output and Firm Size in the Pharmaceutical Industry, Int*ernational Journal of the Economics of Business* 2, 201-219.

DiMasi, Joseph A., Ronald W. Hansen, and Henry Grabowski, 2003, The Price of Innovation: New Estimates of Drug Development Costs, *Journal of Health Economics* 22, 151-185.
DiMasi, Joseph A., Ronald W. Hansen, Henry G. Grabowski, and Louis Lasagna, 1991, Cost of Innovation in the Pharmaceutical Industry, *Journal of Health Economics* 10, 107-142.
DiMasi, Joseph A., J. C. Hermann, K. Twyman, R. K. Kondru, S. Stergiopoulos, K. A. Getz, and W. Rackoff, 2015, A Tool for Predicting Regulatory Approval After Phase II Testing of New Oncology Compounds, *Clinical Pharmacology & Therapeutics* 98, 506-513.
DiMasi, Joseph A., Jennifer Kim, and Kenneth A. Getz, 2014, The Impact of Collaborative and Risk-sharing Innovation Approaches on Clinical and Regulatory Cycle Times, *Therapeutic Innovation & Regulatory Science* 48, 482-487.
DiMasi, Joseph A., J. M. Reichert, L. Feldman, and A. Malins, 2013, Clinical Approval Success Rates for Investigational Cancer Drugs, *Clinical Pharmacology & Therapeutics* 94, 329-335.
Dosi, Giovanni, 1982, Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change, *Research Policy* 11, 147-162.
Dranove, David, and David Meltzer, 1994, Do Important Drugs Reach the Market Sooner?, *The RAND Journal of Economics* 25, 402-423.
Eccles, Robert, Robert Herz, Mary Keegan, and David Phillips, 2001, The Risk of Risk, *Balance Sheet* 9, 28-33.
Egan, Edward J., 2012, The Economics of Patent Citations: Startup Commercialization Strategy, Value, and Success (Online), University of California, Berkeley Electronic Theses and Dissertations, Available: https://escholarship.org/uc/item/7fz268rp (Accessed on 9th May 2013).
Eisenhardt, Kathleen M., and Behnam N. Tabrizi, 1995, Accelerating Adaptive Processes: Product Innovation in the Global Computer Industry, *Administrative Science Quarterly* 40, 84-110.
Ely, Kirsten, Paul J. Simko, and L. G. Thomas, 2003, The Usefulness of Biotechnology Firms' Drug Development Status in the Evaluation of Research and Development Costs, *Journal of Accounting, Auditing and Finance* 18, 163-196.
Escribano, Alvaro, Andrea Fosfuri, and Josep A. Tribó, 2009, Managing External Knowledge Flows: The Moderating Role of Absorptive Capacity, *Research Policy* 38, 96-105.
Espahbodi, Hassan, and Pouran Espahbodi, 2003, Binary Choice Models and Corporate Takeover, *Journal of Banking & Finance* 27, 549-574.
Espinosa, Monica, Miles Gietzmann, and Ivana Raonic, 2009, US Institutional Investors Response to the News Flow of Intangibles Intensive European Stocks: A Study of European Biotech and Pharma Stocks, *European Accounting Review* 18, 63-92.
Fabrizio, Kira R., 2009, Absorptive Capacity and the Search for Innovation, *Research Policy* 38, 255-267.
Fama, Eugene F., and Kenneth R. French, 1998, Value Versus Growth: The International Evidence, *The Journal of Finance* 53, 1975-1999.
Fischer, Timo, and Jan Leidinger, 2014, Testing Patent Value Indicators on Directly Observed Patent Value—An Empirical Analysis of Ocean Tomo Patent Auctions, *Research Policy* 43, 519-529.
Fischer, Timo, and Philipp Ringler, 2014, What Patents are Used as Collateral?—An Empirical Analysis of Patent Reassignment Data, *Journal of Business Venturing* 29, 633-650.

Fleming, Lee, 2001, Recombinant Uncertainty in Technological Search, *Management Science* 47, 117-132.
Fleming, Lee, and Olav Sorenson, 2004, Science as a Map in Technological Search, *Strategic Management Journal* 25, 909–928.
Fortmann-Roe, Scott, 2012, Accurately Measuring Model Prediction Error (Online), Available: http://scott.fortmann-roe.com/docs/MeasuringError.html (Accessed on 18th March 2017).
Gambardella, Alfonso, 1992, Competitive Advantages from In-house Scientific Research: The US Pharmaceutical Industry in the 1980s, *Research Policy* 21, 391-407.
Gambardella, Alfonso, Dietmar Harhoff, and Bart Verspagen, 2008, The Value of European Patents, *European Management Review* 5, 69-84.
Gambardella, Alfonso, Dietmar Harhoff, and Bart Verspagen, 2011, The Determinants of the Private Value of Patented Inventions, (Mimeo, Bocconi University).
Gao, Wenlian, and Julia Chou, 2015, Innovation Efficiency, Global Diversification, and Firm Value, *Journal of Corporate Finance* 30, 278-298.
Gittelman, Michelle, 2007, Does Geography Matter for Science-based Firms? Epistemic Communities and the Geography of Research and Patenting in Biotechnology, *Organization Science* 18, 724-741.
Gittelman, Michelle, and Bruce Kogut, 2003, Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns, *Management Science* 49, 366-382.
Giuri, Paola, Myriam Mariani, Stefano Brusoni, Gustavo Crespi, Dominique Francoz, Alfonso Gambardella, Walter Garcia-Fontes, Aldo Geuna, Raul Gonzales, and Dietmar Harhoff, 2007, Inventors and Invention Processes in Europe: Results from the Patval-EU Survey, *Research Policy* 36, 1107-1127.
Grabowski, Henry, 2008, Follow-on Biologics: Data Exclusivity and the Balance Between Innovation and Competition, *Nature Reviews Drug Discovery* 7, 479-488.
Grabowski, Henry G., and John Vernon, 1990, A New Look at the Returns and Risks to Pharmaceutical R&D, *Management Science* 36, 804-821.
Grabowski, Henry G., and John Vernon, 2000, The Distribution of Sales Revenues from Pharmaceutical Innovation, *Pharmacoeconomics* 18, 21-32.
Grabowski, Henry G., and John M Vernon, 1994, Returns to R&D on New Drug Introductions in the 1980s, *Journal of Health Economics* 13, 383-406.
Graham, Stuart J. H., Robert Merges, Pam Samuelson, and Ted Sichelman, 2009, High Technology Entrepreneurs and the Patent System: Results of the 2008 Berkeley Patent Survey, *Berkeley Technology Law Journal* 24, 1255-1327.
Grandi, Alessandro, Bronwyn H. Hall, and Raffaele Oriani, 2009. *R&D and Financial Investors* (Edward Elgar, Cheltenham).
Grant, Robert M., 1996, Toward a Knowledge-based Theory of the Firm, *Strategic Management Journal* 17, 109-122.
Gratton, Lynda, and Sumantra Ghoshal, 2003, Managing Personal Human Capital: New Ethos for the 'Volunteer' Employee, *European Management Journal* 21, 1-10.
Greene, William H., 2003. *Econometric Analysis* (Pearson Education, India).
Griliches, Zvi, 1981, Market Value, R&D, and Patents, *Economics Letters* 7, 183-187.
Griliches, Zvi, 1984. *R&D, Patents, and Productivity* (Chicago University Press, Chicago).
Griliches, Zvi, 1990, Patent Statistics as Economic Indicators: A Survey, *Journal of Economic Literature* 28, 1661-1707.
Griliches, Zvi, Bronwyn H Hall, and Ariel Pakes, 1988, R&D, Patents, and Market Value Revisited: Is There Evidence of a Second Technological Opportunity Related Factor?, NBER *Working Paper* 2624.

Griliches, Zvi, Bronwyn H. Hall, and Ariel Pakes, 1991, R&D, Patents, and Market Value: Is There A Second (Technological Opportunity) Factor?, *Economics of Innovation and New Technology* 1, 183-201.
Griliches, Zvi, Ariel Pakes, and Bronwyn H. Hall, 1986, The Value of Patents as Indicators of Inventive Activity, NBER *Working Paper* 2083.
Grossman, Jerome H., Proctor P. Reid, and Robert P. Morgan, 2001, Contributions of Academic Research to Industrial Performance in Five Industry Sectors, *The Journal of Technology Transfer* 26, 143-152.
Gu, Feng, 2005, Innovation, Future Earnings, and Market Efficiency, *Journal of Accounting, Auditing and Finance* 20, 385-418.
Gujarati, Damodar N., and D. Porter, 2009. *Basic Econometrics McGraw-Hill International Edition* (McGraw-Hill, New York, N.Y.; London).
Guler, Isin, and Atul Nerkar, 2012, The Impact of Global and Local Cohesion on Innovation in the Pharmaceutical Industry, *Strategic Management Journal* 33, 535-549.
Haeussler, Carolin, Dietmar Harhoff, and Elisabeth Mueller, 2009, To Finance or Not....-The Role of Patents for Venture Capital Financing, *CEPR Discussion Paper* 7115.
Hall, Bronwyn H., 1993, The Value of Intangible Corporate Assets: An Empirical Study of the Components of Tobin's Q, NBER *Working Paper* 93-207.
Hall, Bronwyn H., 1999, Innovation and Market Value, NBER *Working Paper* 6984.
Hall, Bronwyn H., and Dietmar Harhoff, 2012, Recent Research on the Economics of Patents, NBER *Working Paper* 17773.
Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, 2000, Market Value and Patent Citations: A First Look, NBER *Working Paper* 7741.
Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, 2001, The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools, NBER *Working Paper* 8498.
Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, 2005, Market Value and Patent Citations, *RAND Journal of Economics* 36, 16-38.
Hall, Bronwyn H., Grid Thoma, and Salvatore Torrisi, 2007, The Market Value of Patents and R&D: Evidence from European Firms, NBER *Working Paper* 13426.
Hall, Bronwyn H., and Rosemarie Ham Ziedonis, 2001, The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995, *RAND Journal of Economics* 32, 101-128.
Halperin, Michael R., and Alok K. Chakrabarti, 1987, Firm and Industry Characteristics Influencing Publications of Scientists in Large American Firms, *R&D Management* 17, 167-173.
Hamill, Philip A, Mark Hutchinson, Quang Minh Nhi Nguyen, and Mark Mulcahy, 2018, FDA Approval Announcements: Attention-grabbing or Event-day Misspecification?, *Economics Letters* 170, 171-174.
Hamill, Philip A., Philip McIlkenny, and Kwaku K. Opong, 2013, Valuation Implications of Pharmaceutical Firms' R&D Regulatory Approval Notifications, *The British Accounting Review* 45, 99-111.
Harhoff, Dietmar, Francis Narin, Frederic M. Scherer, and Katrin Vopel, 1999, Citation Frequency and the Value of Patented Inventions, *The Review of Economics and Statistics* 81, 511-515.
Harhoff, Dietmar, and Markus Reitzig, 2004, Determinants of Opposition Against EPO Patent Grants-The Case of Biotechnology and Pharmaceuticals, *International Journal of Industrial Organization* 22, 443-480.
Harhoff, Dietmar, Frederic M. Scherer, and Katrin Vopel, 2003, Citations, Family Size, Oppositions, and the Value of Patent Rights, *Research Policy* 32, 1343-1364.

Hay, Michael, David W. Thomas, John L. Craighead, Celia Economides, and Jesse Rosenthal, 2014, Clinical Development Success Rates for Investigational Drugs, *Nature Biotechnology* 32, 40-51.
Heemstra, Harald E., Remco L. de Vrueh, Sonja van Weely, Hans A. Büller, and Hubert G. M. Leufkens, 2008, Predictors of Orphan Drug Approval in the European Union, *European Journal of Clinical Pharmacology* 64, 545-552.
Heemstra, Harald E., Hubert G. M. Leufkens, R. P. Channing Rodgers, Kui Xu, Bettie C. G. Voordouw, and M. Miles Braun, 2011, Characteristics of Orphan Drug Applications that Fail to Achieve Marketing Approval in the USA, *Drug Discovery Today* 16, 73-80.
Helmers, Christian, and Mark Rogers, 2011, Does Patenting Help High-tech Start-ups?, *Research Policy* 40, 1016-1027.
Hemphill, C. Scott, and Bhaven N. Sampat, 2011, When Do Generics Challenge Drug Patents?, *Journal of Empirical Legal Studies* 8, 613-649.
Henard, David H., and David M. Szymanski, 2001, Why Some New Products Are More Successful Than Others, *Journal of Marketing Research* 38, 362-375.
Henderson, Rebecca, 1993, Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry, *The RAND Journal of Economics* 24, 248-270.
Henderson, Rebecca, and Iain Cockburn, 1994, Measuring Competence? Exploring Firm Effects in Pharmaceutical Research, *Strategic Management Journal* 15, 63-84.
Henderson, Rebecca M., and Iain M. Cockburn, 1996, Scale, Scope and Spillovers: The Determinants of Research Productivity in Drug Discovery, *RAND Journal of Economics* 27, 32-59.
Hess, Andrew M., and Frank T. Rothaermel, 2011, When are Assets Complementary? Star Scientists, Strategic Alliances, and Innovation in the Pharmaceutical Industry, *Strategic Management Journal* 32, 895-909.
Hicks, D., and Francis Narin, 2001, Strategic Research Alliances and 360 Degree Bibliometric Indicators, in John E. Jankowski, Albert N. Link, and Nicholas S. Vonortas, Eds.: NSF Workshop on Strategic Research Partnerships (National Science Foundation, Division of Science Resources Studies, Arlington, VA.).
Himmelmann, Achim, 2009, R&D Valuation and Stock-market Reactions to Product Innovations – Evidence from the Pharmaceutical and Biotech Industry (Online), Available at: http://www.socialpolitik.ovgu.de/sozialpolitik_media/papers/Himmelmann_Achim_uid813_pid735-p-260.pdf (Accessed on 26th January 2013).
Hingorani, Archana, Karen A. Shastri, and Kuldeep Shastri, 1994, Product Regulation and Stock Prices: The Case of the US Food and Drug Administration, *Journal of the Economics of Business* 1, 163-178.
Hirschey, Mark, and Vernon J. Richardson, 2004, Are Scientific Indicators of Patent Quality Useful to Investors?, *Journal of Empirical Finance* 11, 91-107.
Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li, 2013, Innovative Efficiency and Stock Returns, *Journal of Financial Economics* 107, 632-654.
Hitt, Michael A., Leonard Bierman, Katsuhiko Shimizu, and Rahul Kochhar, 2001, Direct and Moderating Effects of Human Capital on Strategy and Performance in Professional Service Firms: A Resource-based Perspective, *Academy of Management Journal* 44, 13-28.
Hoetker, Glenn, and Rajshree Agarwal, 2007, Death Hurts, But It Isn't Fatal: The Postexit Diffusion of Knowledge Created by Innovative Firms, *Academy of Management Journal* 50, 446-467.
Holak, Susan L., and Donald R. Lehmann., 1990, Purchase Intentions and the Dimensions of Innovation: An Exploratory Model, *Journal of Product Innovation Management:*

An International Publication of the Product Development & Management Association 7, 59-73.

Holthausen, Robert W., and David F. Larcker, 1992, The Prediction of Stock Returns Using Financial Statement Information, *Journal of Accounting and Economics* 15, 373-411.

Hong, Paul, William J. Doll, Abraham Y. Nahm, and Xiao Li, 2004, Knowledge Sharing in Integrated Product Development, *European Journal of Innovation Management* 7, 102-112.

Hoopes, David G., Tammy L. Madsen, and Gordon Walker, 2003, Guest Editors' Introduction to the Special Issue: Why is there a Resource-Based View? Toward a Theory of Competitive Heterogeneity, *Strategic Management Journal* 24, 889-902.

Horstmann, I., MacDonald G. M., and A. Slivinski, 1985, Patents as Information Transfer Mechanisms: To Patent or (Maybe) Not to Patent, *The Journal of Political Economy* 93, 837-858.

Hoskisson, Robert E., Michael A. Hitt, William P. Wan, and Daphne Yiu, 1999, Theory and Research in Strategic Management: Swings of a Pendulum, *Journal of Management* 25, 417-456.

Hosmer Jr, David W., Stanley Lemeshow, and Rodney X. Sturdivant, 2013. *Applied Logistic Regression* (John Wiley & Sons, Hoboken, New Jersey).

Howell, D. C., 2007, Resampling Statistics: Randomization and the Bootstrap (Online), Available at: http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html (Accessed on14th October 2016).

Hsu, David H., and Rosemarie Ham Ziedonis, 2011, Strategic Factor Markets and the Financing of Technology Startups: When Do Patents Matter More as Signaling Devices? *Working Paper*, The Wharton School, University of Pennsylvania.

Huizing, Ard, and Wim Bouman, 2002. Knowledge and Learning, Markets and Organizations: Managing the Information Transaction Space in Choo C. and Bontis N., Ed.: *The Strategic Management of Intellectual Capital and Organizational Knowledge* (Oxford University Press, New York)

Hung, Yu-Chung, Shi-Ming Huang, Quo-Pin Lin, and Mei-Ling Tsai, 2005, Critical Factors in Adopting a Knowledge Management System for the Pharmaceutical Industry, *Industrial Management & Data Systems* 105, 164-183.

Inoue, Atsushi, and Lutz Kilian, 2004, Bagging Time Series Models, *CEPR Discussion Paper No. 4333* (Online). Available at SSRN: https://ssrn.com/abstract=540262 (Accessed on 10th March 2017).

Jaffe, Adam B., 1986, Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value, *The American Economic Review* 76, 984-1001.

Jaffe, Adam B., and Manuel Trajtenberg, 1996, Flows of Knowledge from Universities and Federal Laboratories: Modeling the Flow of Patent Citations Over Time and Across Institutional and Geographic Boundaries, *Proceedings of the National Academy of Sciences* 93, 12671-12677.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty, 2000, Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors, *American Economic Review* 90, 215-218.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, 1993, Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics* 108, 577.

Jensen, Elizabeth J., 1987, Research Expenditures and the Discovery of New Drugs, *The Journal of Industrial Economics* 83-95.

Jibu, Mari, Yoshiyuki Osabe, and Katy Börner, 2015, Knowledge Flows and Delays in the Pharmaceutical Innovation System, International Conference on Scientometrics & Informetrics (ISSI) Paper (Online), Available at https://pdfs.semanticscholar.org/f2ae/2b0ca487b1a8756a3bbe11d372a86f06a993.pdf (Accessed 24th on September 2015).

Jones, Stewart, David Johnstone, and Roy Wilson, 2017, Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks, *Journal of Business Finance & Accounting* 44, 3-34.

Jong, Simcha, and Kremena Slavova, 2014, When Publications Lead to Products: The Open Science Conundrum in New Product Development, *Research Policy* 43, 645-654.

Kaitin, Kenneth I., and Catherine Cairns, 2003, The New Drug Approvals of 1999, 2000, and 2001: Drug Development Trends A Decade After Passage of the Prescription Drug User Fee Act of 1992, *Drug Information Journal* 37, 357-371.

Kaitin, Kenneth I., Patricia A. Dicerbo, and Louis Lasagna, 1991, The New Drug Approvals of 1987, 1988, and 1989: Trends in Drug Development, *The Journal of Clinical Pharmacology* 31, 116-122.

Kaitin, Kenneth I., and Elaine M. Healy, 2000, The New Drug Approvals of 1996, 1997, and 1998: Drug Development Trends in the User Fee Era, *Drug Information Journal* 34, 1-14.

Kaitin, Kenneth I., Michael Manocchia, Mark Seibring, and Louis Lasagna, 1994, The New Drug Approvals of 1990, 1991, and 1992: Trends in Drug Development, *The Journal of Clinical Pharmacology* 34, 120-127.

Kaitin, Kenneth I., Barbara W. Richard, and Louis Lasagna, 1987, Trends in Drug Development: The 1985-86 New Drug Approvals, *The Journal of Clinical Pharmacology* 27, 542-548.

Kamien, M. I., and N. L. Schwartz, 1976, On the Degree of Rivalry for Maximum Innovative Activity, *Quarterly Journal of Economics* 90, 245-260.

Kapoor, Rahul, and Thomas Klueter, 2015, Decoding the Adaptability–Rigidity Puzzle: Evidence from Pharmaceutical incumbents' Pursuit of Gene Therapy and Monoclonal Antibodies, *Academy of Management Journal* 58, 1180-1207.

Katila, Riitta, and Eric L. Chen, 2008, Effects of Search Timing on Innovation: The Value of Not Being in Sync with Rivals, *Administrative Science Quarterly* 53, 593-625.

Kesan, Jay P., 2009, Transferring Innovation, *Fordham Law Review* 77, 2169-2224.

Kesselheim, Aaron S., Jessica A. Myers, and Jerry Avorn, 2011, Characteristics of Clinical Trials to Support Approval of Orphan vs Nonorphan Drugs for Cancer, *Jama* 305, 2320-2326.

Khoury, Theodore A., and Erin G. Pleggenkuhle-Miles, 2011, Shared Inventions and the Evolution of Capabilities: Examining the Biotechnology Industry, *Research Policy* 40, 943-956.

Kim, W. Chan, and Renée Mauborgne, 1999, Strategy, Value Innovation, and the Knowledge Economy, *MIT Sloan Management Review* 40, 41.

Knight, H. Jackson, 1996. *Patent Strategy: For Researchers and Research Managers* (John Wiley Ann Sons, Chichester).

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2012, Technological Innovation, Resource Allocation, and Growth (Online). Available https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2193068 (Accessed on 16th January 2013).

Kogut, Bruce, and Udo Zander, 1992, Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology, *Organization Science* 3, 383-397.

Korkeamaki, Timo, and Tuomos Takalo, 2013, Valuation of Innovation and Intellectual Property: The Case of iPhone, *European Management Review* 10, 197–210.

Kothari, Sagar P., Jowell S. Sabino, and Tzachi Zach, 2005, Implications of Survival and Data Trimming for Tests of Market Efficiency, *Journal of Accounting and Economics* 39, 129-161.
Kuhn, Max, and Kjell Johnson, 2013. *Applied Predictive Modeling* (Springer, New York).
Lane, Peter J., and Michael Lubatkin, 1998, Relative Absorptive Capacity and Interorganizational Learning, *Strategic Management Journal* 19, 461-477.
Lanjouw, Jean O., 1998, Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value, *Review of Economic Studies* 65, 671-710.
Lanjouw, Jean O., and Iain M. Cockburn, 2001, New Pills for Poor People? Empirical Evidence After GATT, *World Development* 29, 265-289.
Lanjouw, Jean O., and Mark Schankerman, 1999, The Quality of Ideas: Measuring Innovation with Multiple Indicators, NBER *Working Paper* 7345.
Lanjouw, Jean O., and Mark Schankerman, 2001, Characteristics of Patent Litigation: A Window on Competition, *RAND Journal of Economics* 32, 129-151.
Lanjouw, Jean O., and Mark Schankerman, 2004, Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators, *The Economic Journal* 114, 441-565.
Leonard-Barton, Dorothy, 1992, Core Capabilities and Core Rigidities: A Paradox in Managing New Product Development, *Strategic Management Journal* 13, 111-125.
Lerner, Joshua, 1994, The Importance of Patent Scope: An Empirical Analysis, *RAND Journal of Economics* 25, 319-333.
Lerner, Joshua, 2010, The Litigation of Financial Innovations, *The Journal of Law and Economics* 53, 807-831.
Lev, Baruch, 2003, Remarks on the Measurement, Valuation, and Reporting of Intangible Assets, *Federal Reserve Bank of New York* (*FRBNY) Economic Policy Review* 9, 17-22.
Levin, Richard C., Alvin K. Klevorick, Richard R. Nelson, and Sidney G. Winter, 1987, Appropriating the Returns of Industrial Research and Development, *Brookings Papers on Economic Activity* 3, 783-831.
Liberti, Lawrence, Alasdair Breckenridge, Jarno Hoekman, Neil McAuslane, Pieter Stolk, and Hubert Leufkens, 2017, Factors Related to Drug Approvals: Predictors of Outcome?, *Drug Discovery Today* 22, 937-946.
Liebeskind, Julia Porter, Amalya Lumerman Oliver, Lynne Zucker, and Marilynn Brewer, 1996, Social Networks, Learning, and Flexibility: Sourcing Scientific Knowledge in New Biotechnology Firms, *Organization Science* 7, 428-443.
Lim, Kwanghui, 2009, The Many Faces of Absorptive Capacity: Spillovers of Copper Interconnect Technology for Semiconductor Chips, *Industrial and Corporate Change* 18, 1249-1284.
Lin, Lin, and Jenifer Piesse, 2004, Identification of Corporate Distress in UK Industrials: A Conditional Probability Analysis Approach, *Applied Financial Economics* 14, 73-82.
Liu, Kun, 2014, Human Capital, Social Collaboration, and Patent Renewal within US Pharmaceutical Firms, *Journal of Management* 40, 616-636.
Liu, Kun, Jonathan Arthurs, John Cullen, and Roger Alexander, 2008, Internal Sequential Innovations: How Does Interrelatedness Affect Patent Renewal?, *Research Policy* 37, 946-953.
Liu, Qiao, 2006, How Good Is Good News? Technology Depth, Book-to-Market Ratio, and Innovative Events, *Journal of Accounting, Auditing & Finance* 21, 293-321.
Lo, Andrew W., 1986, Logit Versus Discriminant Analysis: A Specification Test and Application to Corporate Bankruptcies, *Journal of Econometrics* 31, 151-178.

Loughran, Tim, and Jay R. Ritter, 2000, Uniformly Least Powerful Tests of Market Efficiency, *Journal of Financial Economics* 55, 361-389.
Lyon, John D., Brad M. Barber, and Chih-Ling Tsai, 1999, Improved Methods for Tests of Long-Run Abnormal Stock Returns, *The Journal of Finance* 54, 165-201.
Mackinlay, A. Craig, 1997, Event Studies in Economics and Finance, *Journal of Economic Literature* 35, 13-39.
Maes, Johan, and Luc Sels, 2014, SMEs' Radical Product Innovation: The Role of Internally and Externally Oriented Knowledge Capabilities, *Journal of Small Business Management* 52, 141-163.
Magazzini, Laura, Fabio Pammolli, and Massimo Riccaboni, 2008, Patent Value and R&D Competition (Online), Working Paper Series, Department of Economics, University of Verona, Available at http://dse.univr.it/workingpapers/WP51.pdf (Accessed on 24th September 2015).
Magazzini, Laura, Fabio Pammolli, and Massimo Riccaboni, 2012, Learning from Failures or Failing to Learn? Lessons from Pharmaceutical R&D, *European Management Review* 9, 45-58.
Mann, Ronald J., 2005, Do Patents Facilitate Financing in the Software Industry?, *Texas Law Review* 83, 961-1030.
Mann, Ronald J., and Thomas W. Sager, 2007, Patents, Venture Capital, and Software Start-ups, *Research Policy* 36, 193-208.
Mansfield, Edwin, 1981, Composition of R and D Expenditures: Relationship to Size of Firm, Concentration, and Innovative Output, *The Review of Economic and Statistics* 63, 610-615.
Mansfield, Edwin, 1995, Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing, *The Review of Economics and Statistics* 77, 55-65.
Mansfield, Edwin, 1998, Academic Research and Industrial Innovation: An Update of Empirical Findings, *Research Policy* 26, 773-776.
March, James G., 1991, Exploration and Exploitation in Organizational Learning, *Organization Science* 2, 71-87.
Marr, Bernard, and J.-C. Spender, 2004, Measuring Knowledge Assets–Implications of the Knowledge Economy for Performance Measurement, *Measuring Business Excellence* 8, 18-27.
Mattison, Nancy, A. Gene Trimble, and Louis Lasagna, 1988, New Drug Development in the United States, 1963 through 1984, *Clinical Pharmacology & Therapeutics* 43, 290-301.
McEvily, Susan K., and Bala Chakravarthy, 2002, The Persistence of Knowledge-Based Advantage: An Empirical Test for Product Performance and Technological Knowledge, *Strategic Management Journal* 23, 285-305.
McMillan, C. Steven, Alfredo Mauri, and Robert D. Hamilton, 2003, The Impact of Publishing and Patenting Activities on New Product Development and Firm Performance: The Case of the US Pharmaceutical Industry, *International Journal of Innovation Management* 7, 213-221.
McMillan, C. Steven, Francis Narin, and David L. Deeds, 2000, An Analysis of the Critical Role of Public Science in Innovation: The Case of Biotechnology, *Research Policy* 29, 1-8.
McNamara, Peter, and Charles Baden-Fuller, 2007, Shareholder Returns and the Exploration-exploitation Dilemma: R&D Announcements by Biotechnology Firms, *Research Policy* 36, 548-565.
Menell, Peter, and Suzanne Scotchmer, 2007. *Intellectual Property* (Elsevier, Amsterdam).
Mestre-Ferrandiz, Jorge, Jon Sussex, and Adrian Towse, 2012, The R&D Cost of a New Medicine, *Monographs*, Office of Health Economics, Number 000135.

Meyer, Martin, 2000, Does Science Push Technology? Patents Citing Scientific Literature, *Research Policy* 29, 409-434.
Meyer, M., and O. Persson, 1998, Nanotechnology - Interdisciplinarity, Patterns of Collaboration and Differences in Application, *Scientometrics* 42, 195-205.
Moore, Kimberly A., 2000, Judges, Juries, and Patent Cases - An Empirical Peek Inside the Black Box, *Michigan Law Review* 99, 365-409.
Moore, Kimberly A, 2005, Worthless Patents, *Berkeley Technology Law Journal* 20, 1521.
Mosteller, Frederick, and John Wilder Tukey, 1977. Data Analysis and Regression: A Second Course in Statistics, *Addison-Wesley Series in Behavioral Science: Quantitative Methods* (Addison-Wesley, Reading, Mass).
Munos, Bernard, 2009, Lessons from 60 Years of Pharmaceutical Innovation, *Nature Reviews Drug Discovery* 8, 959.
Nagaoka, Sadao, 2005, Patent Quality, Cumulative Innovation and Market Value: Evidence from Japanese Firm Level Panel Data, *Institute of Innovation Research Working Paper* (Hitotsubashi University, Tokyo).
Nagaoka, Sadao, 2007, Assessing The R&D Management of a Firm in Terms of Speed and Science Linkage: Evidence from the US Patents, *Journal of Economics & Management Strategy* 16, 129-156.
Nagaoka, Sadao, and John P. Walsh, 2009, Commercialization and Other Uses of Patents in Japan and the U.S.: Major Findings from the Rieti-Georgia Tech Inventor Survey, *RIETI Discussion Paper Series* 09-E-011.
Nahapiet, Janine, and Sumantra Ghoshal, 1998, Social Capital, Intellectual Capital, and the Organizational Advantage, *Academy of Management Review* 23, 242-266.
Narin, Francis, Mark P. Carpenter, and Patricia Woolf, 1984, Technological Performance Assessments Based on Patents and Patent Citations, *IEEE Transactions on Engineering Management* EM-31, 172-183.
Narin, Francis, Kimberly S. Hamilton, and Dominic Olivastro, 1997, The Increasing Linkage Between U.S. Technology and Public Science, *Research Policy* 26, 317-330.
Narin, Francis, and Elliot Noma, 1985, Is Technology Becoming Science?, *Scientometrics* 7, 369-381.
Narin, Francis, Elliot Noma, and Ross Perry, 1987, Patents as Indicators of Corporate Technological Strength, *Research Policy* 16, 143-155.
Narin, Francis, and Dominic Olivastro, 1992, Status Report: Linkage Between Technology and Science, *Research Policy* 21, 237-249.
Nelson, Richard R., 1991, Why Do Firms Differ, and How Does it Matter?, *Strategic Management Journal* 12, 61-74.
Nelson, Richard R., and Sidney G. Winter, 1982, The Schumpeterian Tradeoff Revisited, *The American Economic Review* 72, 114-132.
Nemet, Gregory F., and Evan Johnson, 2012, Do Important Inventions Benefit from Knowledge Originating in Other Technological Domains?, *Research Policy* 41, 190-200.
Nerkar, Atul, 2003, Old is Gold? The Value of Temporal Exploration in the Creation of New Knowledge, *Management Science* 49, 211-229.
Nicholls-Nixon, Charlene L., 1995, Responding to Technological Change: Why Some Firms Do and Others Die, *The Journal of High Technology Management Research* 6, 1-16.
Nohria, Nitin, and Ranjay Gulati, 1996, Is Slack Good or Bad for Innovation?, *Academy of Management Journal* 39, 1245-1264.
Nonaka, Ikujiro, 1994, A Dynamic Theory of Organizational Knowledge Creation, *Organization Science* 5, 14-37.

Nonaka, Ikujiro, and Hirotaka Takeuchi, 1995. *The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation*, (Oxford University Press, Oxford).
Novelli, Elena, 2015, An Examination of the Antecedents and Implications of Patent Scope, *Research Policy* 44, 493-507.
Oh, Jong-Min, 2017, Absorptive Capacity, Technology Spillovers, and the Cross-section of Stock Returns, *Journal of Banking & Finance* 85, 146-164.
Ohlson, James A., 1980, Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research* 18, 109-131.
Olson, Mary K., 2013, Eliminating the US Drug Lag: Implications for Drug Safety, *Journal of Risk and Uncertainty* 47, 1-30.
Omland, Nils, 2011. Valuing Patents Through Indicators, in the *Economic Valuation of Patents. Methods and Applications* (Edward Elgar Publisher, Cheltenham).
Organisation for Economic Co-operation and Development (OECD), 1999. *Managing National Innovation Systems* (OECD Publications, Paris, France).
Organisation for Economic Co-operation and Development (OECD), 2002. *Benchmarking Science-Industry Relationships* (OECD Publications, Paris, France).
Organisation for Economic Co-operation and Development (OECD), 2009. OECD *Patent Statistics Manual 2009* (OECD Publications, Paris, France).
Ouellette, Lisa Larrimore, 2010, How Many Patents Does It Take to Make A Drug? Follow-on Pharmaceutical Patents and University Licensing, *Michigan Telecommunications and Technology Law Review* 17, 299-336.
Pakes, Ariel, 1985, On Patents, R&D, and the Stock Market Rate of Return, *Journal of Political Economy* 93, 390-409.
Pakes, Ariel, 1986, Patents as Options: Some Estimates of the Value of Holding European Patent Stocks, *Econometrica* 54, 755-784.
Palepu, Krishna G., 1986, Predicting Takeover Targets: A Methodological and Empirical Analysis, *Journal of Accounting and Economics* 8, 3-35.
Palomeras, Neus, 2003, Sleeping Patents: Any Reason to Wake Up, *IESE Research Papers* No D/506.
Palomeras, Neus, and Eduardo Melero, 2010, Markets for Inventors: Learning-by-Hiring as a Driver of Mobility, *Management Science* 56, 881-895.
Pan, Yang, and Robert T. Jackson, 2008, Ethnic Difference in the Relationship Between Acute Inflammation and Serum Ferritin in US Adult Males, *Epidemiology & Infection* 136, 421-431.
Pandit, Shail, Charles E. Wasley, and Tzachi Zach, 2011, The Effect of Research and Development (R&D) Inputs and Outputs on the Relation Between the Uncertainty of Future Operating Performance and R&D Expenditures, *Journal of Accounting, Auditing and Finance* 26, 121-144.
Patel, Darshak, and Michael R. Ward, 2011, Using Patent Citation Patterns to Infer Innovation Market Competition, *Research Policy* 40, 886-894.
Pavitt, Keith, 1991, What Makes Basic Research Economically Useful?, *Research Policy* 20, 109-119.
Peteraf, M., 1993, The Cornerstones of Competitive Advantage: A Resource-based View, *Strategic Management Journal* 14, 179-191.
Pisano, Gary P., 1994, Knowledge, Integration, and the Locus of Learning: An Empirical Analysis of Process Development, *Strategic Management Journal* 15, 85-100.
Pisano, Gary P., 1997, R&D Performance, Collaborative Arrangements and the Market for Know-How: A Test of the "Lemons" Hypothesis in Biotechnology, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=41980 (Accessed on 15th March 2014).

Plumlee, Marlene, Yuan Xie, Meng Yan, and Jeff Jiewei Yu, 2015, Bank Loan Spread and Private Information: Pending Approval Patents, *Review of Accounting Studies* 20, 593-638.
Polanyi, Michael, 1966. *The Tacit Dimension* (Routledge and K. Paul, London).
Porter, Michael E., 1980, Industry Structure and Competitive Strategy: Keys to Profitability, *Financial Analysts Journal* 36, 30-41.
Powell, Ronan G., 2001, Takeover Prediction and Portfolio Performance: A Note, *Journal of Business Finance & Accounting* 28, 993-1011.
Powell, Ronan G., 2004, Takeover Prediction Models and Portfolio Strategies: A Multinomial Approach, *Multinational Finance Journal* 8, 35-72.
Powell, Walter W., Kenneth W. Koput, and Laurel Smith-Doerr, 1996, Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology, *Administrative Science Quarterly* 41, 116-145.
Press, S. James, and Sandra Wilson, 1978, Choosing Between Logistic Regression and Discriminant Analysis, *Journal of the American Statistical Association* 73, 699-705.
Putzeist, Michelle, Harald E. Heemstra, Jordi Llinares Garcia, Aukje K. Mantel-Teeuwisse, Christine C. Gispen-De Wied, Arno W. Hoes, and Hubert G. M. Leufkens, 2012, Determinants for Successful Marketing Authorisation of Orphan Medicinal Products in the EU, *Drug Discovery Today* 17, 352-358.
Reed, Richard, and Robert J. DeFillippi, 1990, Causal Ambiguity, Barriers to Imitation, and Sustainable Competitive Advantage, *Academy of Management Review* 15, 88-102.
Reitzig, Markus, 2002, The Private Values of 'Thicket' Patenting: An Empirical Study Across Industries and Patent Types, Academy of Management Conference, Denver, Colorado (Online), Available at https://s3.amazonaws.com/academia.edu.documents/32214923/REITZIG.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1552921803&Signature=RSsPWGdjCn0%2FUdDATFSmy7c1bUw%3D&response-content-disposition=inline%3B%20filename%3DThe_Private_Values_of_Thicket_Patenting.pdf (Accessed on 30th November 2018).
Reitzig, Markus 2003, What Determines Patent Value? Insights from the Semiconductor Industry, *Research Policy* 32, 13-26.
Reitzig, Markus, 2004, Improving Patent Valuations for Management Purposes—Validating New Indicators by Analyzing Application Rationales, *Research Policy* 33, 937-957.
Roach, Michael, and Wesley M. Cohen, 2013, Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research, *Management Science* 59, 504-525.
Roberts, Peter W., and Raphael Amit, 2003, The Dynamics of Innovative Activity and Competitive Advantage: The Case of Australian Retail Banking, 1981 To 1995, *Organization Science* 14, 107-122.
Rogerson, Peter, 2001. *Statistical Methods for Geography* (Sage Publications Ltd, New York).
Romer, Paul M., and Joel Kurtzman, 2004. The Knowledge Economy, in *Handbook on Knowledge Management 1* (Springer, Berlin, Heidelberg).
Roos, Johan, 1998, Exploring the Concept of Intellectual Capital (IC), *Long Range Planning* 31, 150-153.
Roos, Johan, Göran Roos, Nicola C. Dragonetti, and Leif Edvinsson, 1997. *Intellectual Capital: Navigating the New Business Landscape* (MacMillan Press Ltd, London).
Rosenberg, Nathan, 1990, Why do Firms do Basic Research (with their Own Money)?, *Research Policy* 19, 165-174.

Rothaermel, Frank T., and David L. Deeds, 2004, Exploration and Exploitation Alliances in Biotechnology: A System of New Product Development, *Strategic Management Journal* 25, 201-221.
Rothaermel, Frank T., and Andrew M. Hess, 2007, Building Dynamic Capabilities: Innovation Driven by Individual-, Firm-, and Network-level Effects, *Organization Science* 18, 898-921.
Sakakibara, Mariko, 2010, An Empirical Analysis of Pricing in Patent Licensing Contracts, *Industrial and Corporate Change* 19, 927-945.
Sampat, Bhaven N., and Arvids A. Ziedonis, 2004. Patent Citations and the Economic Value of Patents, A Preliminary Assessment, in H. F. Moed et al., Eds.: *Handbook of Quantitative Science and Technology Research* (Kluwer Academic Publishers, Netherlands).
Sanders, Barkev S., Joseph Rossman, and L. James Harris, 1958, The Economic Impact of Patents, *Patent, Trademark and Copyright Journal of Research and Education* 2, 340-362.
Sarkar, Salil K., and Pieter J. De Jong, 2006, Market Response to FDA Announcements, *The Quarterly Review of Economics and Finance* 46, 586-597.
Scannell, Jack W., Alex Blanckley, Helen Boldon, and Brian Warrington, 2012, Diagnosing the Decline in Pharmaceutical R&D Efficiency, *Nature Reviews Drug Discovery* 11, 191.
Schankerman, Mark, and Ariel Pakes, 1986, Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period, *The Economic Journal* 96, 1052-1076.
Scherer, Frederick M., 2010. Pharmaceutical Innovation, in Bronwyn H. Hall, and Nathan Rosenberg, Eds.: *Handbook of the Economics of Innovation* (Elsevier, Amsterdam: North-Holland).
Scherer, Frederick M., and David Ross, 1990. *Economic Performance* (Boston: Houghton-Mifflin).
Schmookler, Jacob, 1962, Determinants of Inventive Activity, *The American Economic Review* 52, 165-176.
Schmookler, Jacob, 1966. *Invention and Economic Growth* (Harvard University Press, Cambridge, MA).
Schoenmakers, Wilfred, and Geert Duysters, 2010, The Technological Origins of Radical Inventions, *Research Policy* 39, 1051-1059.
Schwartz, Eduardo S, 2004, Patents and R&D as Real Options, *Economic Notes* 33, 23-54.
Serrano, Carlos J., 2010, The Dynamics of the Transfer and Renewal of Patents, *RAND Journal of Economics* 41, 686-708.
Sertkaya, Aylin, Anna Birkenbach, Ayesha Berlind, and John Eyraud, 2014, Examination of Clinical Trial Costs and Barriers for Drug Development (Eastern Research Group, Inc, Lexington, MA).
Shan, Weijan, Gordon Walker, and Bruce Kogut, 1994, Interfirm Cooperation and Startup Innovation in the Biotechnology Industry, *Strategic Management Journal* 15, 387-394.
Sharma, Anurag, and Nelson Lacey, 2004, Linking Product Development Outcomes to Market Valuation of the Firm: The Case of the U.S. Pharmaceutical Industry, *The Journal of Product Innovation Management* 21, 297-308.
Sheck, Lorraine, Christopher Cox, Henry T. Davis, A. Gene Trimble, William M. Wardell, and Ronald Hansen, 1984, Success Rates in the United States Drug Development System, *Clinical Pharmacology & Therapeutics* 36, 574-583.
Sher, Peter J., and Phil Y. Yang, 2005, The Effects of Innovative Capabilities and R&D Clustering on Firm Performance: The Evidence of Taiwan's Semiconductor Industry, *Technovation* 25, 33-43.

Shumway, Tyler, 2001, forecasting Bankruptcy More Accurately: A Simple Hazard Model, *The Journal of Business* 74, 101-124.
Sichelman, Ted, and Stuart J. H. Graham, 2010, Patenting by Entrepreneurs: An Empirical Study, *Michigan Telecommunications and Technology Law Review* 17, 67-110.
Silverberg, Gerald, and Bart Verspagen, 2007, The Size Distribution of Innovations Revisited: An Application of Extreme Value Statistics to Citation and Value Measures of Patent Significance, *Journal of Econometrics* 139, 318-339.
Singh, Jasjit, and Lee Fleming, 2010, Lone Inventors as Sources of Breakthroughs: Myth or Reality?, *Management Science* 56, 41-56.
Smith, Ken G., Christopher J. Collins, and Kevin D. Clark, 2005, Existing Knowledge, Knowledge Creation Capability, and the Rate of New Product Introduction in High-technology Firms, *Academy of Management Journal* 48, 346-357.
Smith, Keith H., 2002, What is the 'Knowledge Economy'? Knowledge Intensity and Distributed Knowledge Bases (Online), *UNU/INTECH Discussion Paper Series*, United Nations University, Institute of New Technologies, Available at https://eprints.utas.edu.au/1235/1/2002-6.pdf (Accessed on 3rd December 2018)
Soh, Pek-Hooi, and Annapoornima M. Subramanian, 2014, When Do Firms Benefit from University–Industry R&D Collaborations? The Implications of Firm R&D Focus on Scientific Research and Technological Recombination, *Journal of Business Venturing* 29, 807-821.
Sørensen, Jesper B., and Toby E. Stuart, 2000, Aging, Obsolescence, and Organizational Innovation, *Administrative Science Quarterly* 45, 81-112.
Spender, J.-C., 1996, Making Knowledge the Basis of a Dynamic Theory of the Firm, *Strategic Management Journal* 17, 45-62.
Stern, Scott, 2004, Do Scientists Pay to be Scientists?, *Management Science* 50, 835-853.
Sternitzke, Christian, 2009, Patents and Publications as Sources of Novel and Inventive Knowledge, *Scientometrics* 79, 551-561.
Stewart, Thomas A., 1998. *Intellectual Capital: The New Wealth of Organizations* (Nicholas Brealey Publishing, London).
Stone, Alexandra, and Kenneth Flamm, 2009, Knowledge and Productivity in the Pharmaceutical Industry, Science and Innovation Policy, 2009 Atlanta Conference on Science and Innovation Policy (IEEE) (Online), Available at https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5367866 (Accessed on 1st November 2013).
Su, Hsin-Ning, Carey Ming-Li Chen, and Pei-Chun Lee, 2012, Patent Litigation Precaution Method: Analyzing Characteristics of US Litigated and Non-litigated Patents from 1976 to 2010, *Scientometrics* 92, 181-195.
Subramanian, Annapoornima M., Kwanghui Lim, and Pek-Hooi Soh, 2013, When Birds of a Feather Don't Flock Together: Different Scientists and the Roles they Play in Biotech R&D Alliances, *Research Policy* 42, 595-612.
Sveiby, Karl-Erik, 2001, A Knowledge-based Theory of the Firm to Guide in Strategy formulation, *Journal of Intellectual Capital* 2, 344-358.
Szulanski, Gabriel, 1996, Exploring Internal Stickiness: Impediments to the Transfer of Best Practice within The Firm, *Strategic Management Journal* 17, 27-43.
Szulanski, Gabriel, Robert J. Jensen, and Tanya Lee, 2003. Adaptation of Know-how for Cross-border Transfer, in *Governing Knowledge-Processes* (Gabler Verlag, Springer Book Archive).
Tabachnick, Barbara G., and Linda S. Fidell, 2001. *Using Multivariate Statistics* (Allyn & Bacon/Pearson Education, Boston, Mass.; London).
Tabachnick, Barbara G., and Linda S. Fidell, 2007. *Using Multivariate Statistics* (Allyn & Bacon/Pearson Education, Boston, Mass.; London).

Teece, David J., 1998, Capturing Value from Knowledge Assets: The New Economy, Markets for Know-how, and Intangible Assets, *California Management Review* 40, 55-79.
Teece, David J, Gary Pisano, and Amy Shuen, 1997, Dynamic Capabilities and Strategic Management, *Strategic Management Journal* 18, 509-533.
Tijssen, Robert J. W., 2000, Technological Relevance of Science: An Assessment of Citation Linkages Between Patents and Research Papers, *Scientometrics* 47, 389-412.
Tijssen, Robert J. W., 2001, Global and Domestic Utilization of Industrial Relevant Science: Patent Citation Analysis of Science-Technology Interactions and Knowledge Flows, *Research Policy* 30, 35-54.
Toivanen, Otto, Paul Stoneman, and Derek Leonard Bosworth, 2002, Innovation and the Market Value of UK Firms, 1989-1995, *Oxford Bulletin of Economics and Statistics* 64, 39-61.
Tong, Xuesong, and J. D. Frame, 1994, Measuring National Technological Performance with Patent Claims Data, *Research Policy* 23, 133-141.
Torabzadeh, Khalil M., Criss G. Woodruff, and Nilanjan Sen, 1998, FDA Decisions on New Drug Applications and the Market Value of Pharmaceutical Firms, *American Business Review* 16, 42.
Torjesen, Ingrid, 2015, Drug Development: The Journey of a Medicine from Lab to Shelf, *The Pharmaceutical Journal* March (Online), Available at https://www.pharmaceutical-journal.com/test-tomorrows-pharmacist/tomorrows-pharmacist/drug-development-the-journey-of-a-medicine-from-lab-to-shelf/20068196.article (Accessed on 18th October 2019).
Trajtenberg, Manuel, 1987, Patents, Citations and Innovations: Tracing the Links, NBER *Working Paper* 2457.
Trajtenberg, Manuel, 1990, A Penny for Your Quotes: Patent Citations and the Value of Innovations, *RAND Journal of Economics* 21, 1990.
Trajtenberg, Manuel, 2002, Government Support for Commercial R&D: Lessons from the Israeli Experience, *Innovation Policy and the Economy* 2, 79-134.
Trappey, Amy J. C., Charles V. Trappey, Chun-Yi Wu, and Chi-Wei Lin, 2012, A Patent Quality Analysis for Innovative Technology and Product Development, *Advanced Engineering Informatics* 26, 26-34.
Van Looy, Bart, Tom Magerman, and Koenraad Debackere, 2007, Developing Technology in the Vicinity of Science: An Examination of the Relationship Between Science Intensity (Of Patents) and Technological Productivity within the Field of Biotechnology, *Scientometrics* 70, 441-458.
Van Looy, Bart, Edwin Zimmermann, Reinhilde Veugelers, Arnold Verbeek, Johanna Mello, and Koenraad Debackere, 2003, Do Science-Technology Interactions Pay off When Developing Technology?, *Scientometrics* 57, 355-367.
Van Vianen, B. G., H. F. Moed, and A. F. J. Van Raan, 1990, An Exploration of the Science Base of Recent Technology, *Research Policy* 19, 61-81.
Vega-Jurado, Jaider, Antonio Gutiérrez-Gracia, and Ignacio Fernández-De-Lucio, 2008, Analyzing the Determinants of Firm's Absorptive Capacity: Beyond R&D, *R&D Management* 38, 392-405.
Venkatraman, Natarjan, and Vasudevan Ramanujam, 1986, Measurement of Business Performance in Strategy Research: A Comparison of Approaches, *Academy of Management Review* 11, 801-814.
Verbeek, Arnold, Koenraad Debackere, Marc Luwel, Petra Andries, Edwin Zimmermann, and Filip Deleus, 2002, Linking Science to Technology: Using Bibliographic References in Patents to Build Linkage Schemes, *Scientometrics* 54, 399-420.

Vincent-Lancrin, Stephan, 2006, What is Changing in Academic Research? Trends and Futures Scenarios, *European Journal of Education* 41, 169-202.
Volberda, Henk W., Nicolai J. Foss, and Marjorie A. Lyles, 2010, Absorbing the Concept of Absorptive Capacity: How to Realize its Potential in the Organization Field, *Organization Science* 21, 931-951.
Von Hippel, Eric, 1994, "Sticky Information" and the Locus of Problem Solving: Implications for Innovation, *Management Science* 40, 429-439.
Von Krogh, Georg, and Johan Roos, 1996. *Managing Knowledge: Perspectives on Cooperation and Competition* (Sage).
Wagner, Stefan, and Wakeman, Simon, 2014, What Do Patent-Based Measures Tell Us About Product Commercialization? Evidence from the Pharmaceutical Industry, ESMT Working Paper, No. 14-01, European School of Management and Technology (ESMT), Berlin, Available at http://Nbn-Resolving.De/Urn:Nbn:De:101:1-2014021113834 (Accessed on 10th March 2014)
Wallace, R. S. Olusegun, and Kamal Naser, 1995, Firm-specific Determinants of the Comprehensiveness of Mandatory Disclosure in the Corporate Annual Reports of Firms Listed on the Stock Exchange of Hong Kong, *Journal of Accounting and Public Policy* 14, 311-368.
Wang, Xiaoguang, 2013, Forming Mechanisms and Structures of a Knowledge Transfer Network: Theoretical and Simulation Research, *Journal of Knowledge Management* 17, 278-289.
Wang, B., J. Avorn, and A. S. Kesselheim, 2013, Clinical and Regulatory Features of Drugs Not Initially Approved by the FDA, *Clinical Pharmacology & Therapeutics* 94, 670-677.
Wenger, Etienne C., and William M. Snyder, 2000, Communities of Practice: The Organizational Frontier, *Harvard Business Review* 78, 139-146.
Wernerfelt, Birger, 1984, A Resource-based View of the Firm, *Strategic Management Journal* 5, 171-180.
Wiklund, Johan, and Dean Shepherd, 2003, Knowledge-based Resources, Entrepreneurial Orientation, and the Performance of Small and Medium-Sized Businesses, *Strategic Management Journal* 24, 1307-1314.
Wilcox King, Adelaide, and Carl P. Zeithaml, 2003, Measuring Organizational Knowledge: A Conceptual and Methodological Framework, *Strategic Management Journal* 24, 763-772.
Zack, Michael H., 2002, A Strategic Pretext for Knowledge Management, Proceedings of the Third European Conference on Organizational Knowledge, Learning and Capabilities (Online), Available at https://pdfs.semanticscholar.org/2d34/bfbc4c9ae0d86afd0766a7eac059521d223e.pdf (Accessed on 4th December 2018).
Zahra, Shaker A., and Gerard George, 2002, Absorptive Capacity: A Review, Reconceptualization, and Extension, *Academy of Management Review* 27, 185-203.
Zavgren, Christine V., 1985, Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis, *Journal of Business Finance & Accounting* 12, 19-45.
Zhou, Xiao-Hua, D. K. McClish, and N. A. Obuchowski, 2011. *Statistical Methods in Diagnostic Medicine* (Wiley-Blackwell, Hoboken, N. J).
Zucker, Lynne G., and Michael R. Darby, 1995, Virtuous Circles of Productivity: Star Bioscientists and the Institutional Transformation of Industry, NBER *Working Paper* 5342.
Zucker, Lynne G., Michael R. Darby, and Jeff S. Armstrong, 1994, Intellectual Capital and the Firm: The Technology of Geographically Localized Knowledge Spillovers, NBER *Working Paper* 4946.

Zucker, Lynne G., Michael R. Darby, and Jeff S. Armstrong, 2002, Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology, *Management Science* 48, 138-153.
Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer, 1998, Intellectual Human Capital and the Birth of US Biotechnology Enterprises, *The American Economic Review* 88, 290-306.