Analysis of Data on Spontaneous Reports of Adverse Events Associated with Drugs

by

Emmanuel Mensah Baah

A thesis submitted to the
College of Science and Engineering
at the University of Glasgow
for the degree of
Doctor of Philosophy

February 2014

Abstract

Some adverse drug reactions (ADRs) are not detected before marketing approval is given because clinical trials are not suited for their detection, for various reasons [5, 23]. Drug regulatory bodies therefore weigh the potential benefits of a drug against the harms and allow drugs to be marketed if felt that the potential benefits far outweigh the harms [26,48]. Associated adverse events are subsequently monitored through various means including reports submitted by health professionals and the general public in what is commonly referred to as spontaneous reporting system (SRS) [19, 23, 69]. The resulting database contains thousands of adverse event reports which must be assessed by expert panels to see if they are bona fide adverse drug reactions, but which are not easy to manage by virtue of the volume [6].

This thesis documents work aimed at developing a statistical model for assisting in the identification of bona fide drug side-effects using data from the United States of America's Food and Drugs Administration's (FDA) Spontaneous Reporting System (otherwise known as the Adverse Event Reporting System (AERS)) [28].

Four hierarchical models based on the Conway-Maxwell-Poisson (CMP) distribution [43,78] were explored and one of them was identified as the most suitable for modeling the data. It compares favourably with the Gamma Poisson Shrinker (GPS) of DuMouchel [19] but takes a dimmer view of drug and adverse event pairs with very small observed and expected count than the GPS.

Two results are presented in this thesis; the first one, from a preliminary analysis, presented in Chapter 2, shows that problems such as missing values for age and sex that militate against the optimal use of SRS data, enumerated in the literature, remain. The second results, presented in Chapter 5, concern the main focus of the research mentioned in the previous paragraph.

Acknowledgement

I am indebted to my supervisors: Prof. Stephen J. Senn, Prof. Adrian W. Bowman and Dr. Agostino Nobile for their guidance, criticisms and suggestions; I could not have come this far without your tutelage.

My appreciation to the staff and students of the School of Mathematics and Statistics, College of Science and Engineering and the University of Glasgow at large who in diverse ways have contributed to my studies in the University.

Takoradi Polytechnic deserve plaudits for funding the research work which is recorded in this thesis.

I am obliged to Thearch Daniel Arthur for the lessons in iots. I found your departure painful.

My deepest gratitude to my father, Egya Baah, whose abiding faith in God is a source of inspiration and to my mother, Maame Akosua, who unfortunately, did not live to see the fruits of her sacrifice and hard work.

To my siblings Alfred, Isaac, Grace and Daniel whose companionship, along with other things, have shaped my understanding of humanity, I say: I could not have had a better company!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Drug Safety and Related Issues

Many issues are involved in the difficult and often uncertain undertaking of drug development, amongst them finance, ethics, efficacy of product and safety of users. The process requires meticulous care right from the conception stage to well beyond the stage where approval has been given for marketing, primarily because of safety concerns.

Even though the foremost motivation in drug development is finding a treatment for an illness, safety is of utmost concern because drugs are basically chemicals [77]; they hold the potential to cause harm given the right (or rather wrong) circumstances. This places a huge responsibility on drug producing entities (sponsors) to not only ensure that their products are well formulated and safe, but also provide enough information on the best way to use them. Indeed safety issues are not and should not be the preserve of only sponsors but all, including regulatory bodies and consumers.

Regulatory bodies are there to ensure that only medicines that meet the necessary safety requirements are allowed to enter the market or effect the withdrawal of medicines that have been found unsafe from the market. Otherwise an unscrupulous sponsor could market an unsafe product [77], under the lure of pecuniary or commercial considerations; every facet of drug development is capital intensive and the sponsor is expected not only to have the wherewithal to carry through the venture, but be able to recoup the investment, keep body and soul of its facilitators, meet shareholder expectations and as a commercial entity expand by exploring other remedies. Additionally, it is not difficult to perceive the existence of the huge market for drugs given the proliferation of diseases, in spite of the impressive advances in the science of medicine.

1.1.1 Adverse Drug Reactions

The harm(s) a drug can cause are discussed in terms of the adverse reaction(s) associated with it. Put simply, an adverse drug reaction (ADR), otherwise known as side-effect, is any unwanted effect of a drug [18, 51]. Factors that impinge on the severity and nature of an ADR include the overall health status of the individual taking the drug, dose level of drug, gender, genetic make up, age, chemical composition of the drug, weight and disease condition [18, 51]. Attention usually focuses on undesirable or harmful effects of drugs at the required dose level when side-effects come up for discussion.

1.1.2 Nature and Types of ADRs

A number of factors influence the way an ADR is viewed, which include how it is caused, how serious it is and the way it manifests itself. Based on these influencing factors, Edwards and Aronson [22], drawing from the works of other authors [39, 42, 71, 73], present six classes of ADRs: "dose-related (Augmented), non-dose-related (Bizarre), dose-related and time-related (Chronic), time-related (Delayed), withdrawal (End of use), and failure of therapy (Failure)" [22] in their article "Adverse Drug Reactions: Definitions, Diagnosis, and Management". Another classification in the literature on ADRs puts them into two classes, namely Type A and Type B reactions. Type A reactions (also known as pharmacological reactions) are predictable because they relate to dose, and the chemical process by which they result are understood while Type B reactions (also known as idiosyncratic reactions) are unpredictable from knowledge of the composition of the drug; the process through which they result are not yet understood and are not dose-related. They occur in some people because they are allergic to, or their immune system does not respond favourably to, the medication as a result of their genetic makeup [51, 67, 71, 77].

Some adverse reactions are relatively common, often less serious and easy to manage than others [51]. Examples of common ADRs are "weakness, sweating, nausea and palpitations" [51]. At the risk of belabouring the point, some adverse reactions are rare; they tend to occur in a minority of people and are often more serious [51, 77]. "Skin rashes, jaundice, anaemia, a decrease in the white cells count, kidney damage, and nerve injury that may impair vision or hearing" [51].

1.1.3 Prevalence of ADRs

Over 80 percent of side-effects are Type A reactions [63, 67]. In the United States of America (US), the proportion of hospital admissions attributable to side effects is about 3 to 7 percent. Of those admitted to hospitals for reasons other than side-effects, between 10 to 20 percent manifest side effects during their stay "and about 10 to 20 percent of these are severe" [51]. The corresponding values for the United Kingdom (UK) are 5 percent and 10 to 20 percent respectively, with about 0.11 percent of side effects resulting in deaths [64]. The respective values of 5.2%, 14.7% and 0.15%, obtained in some fairly recent studies of three major hospitals in the UK are consistent with the above values [16,65]. These values are expected to be higher in countries where the literacy rate is low, prescription-only-medications are more or less treated like over-the-counter drugs because of weak regulatory systems and virtually non-existent systems of reporting ADRs.

The sixty-fifth edition of the British National Formulary [74] presents a classification of ADR on the basis of prevalence as shown in Table 1.1.

Table 1.1: A classification of ADRs based on prevalence.

Prevalence	Description
1 in 10	Very Common
1 in 100 to 1 in 10	Common
1 in 1000 to 1 in 100	Less common
1 in 10000 to 1 in 1000	Rare
Less than 1 in 10000	Very Rare

Source: British National Formulary, March 2013 [74].

1.1.4 Detecting ADRs

As mentioned above some adverse reactions are rare in occurrence because they occur in a small minority of people, for a given medication. They are therefore often not detected at the development stage where the number of subjects on whom the drug is tested is, for various reasons, considerably smaller than the number of patients that take the medication when it is marketed [23,64,77]. The number of patients that may have undergone trials with a drug by the time it is marketed is on average less than 3000 [77]. While this number may be enough to identify frequently occurring side-effects, it may not be enough to pinpoint

the attributes of all of them [64], let alone detect rare side-effects which occur at a rate of about one in 10000 or less [5, 74]. This is all the more palpable when one considers the fact that some ADRs result from drug-drug and drug-disease (other than the one being treated) interactions [51,64] which are often not the focus of clinical trials. Thus one needs many more subjects than studied in a clinical trial to fully capture and understand the attributes of a medication with respect to side-effects. This is only possible when a drug has been marketed, as apart from numbers the population of patients is more diverse than in pre-approval studies [5, 23, 70].

Some ADRs take time to manifest (referred to as long-latency) [23, 70] or result from continuous use of the drug over a long period [23,70] and the limited time of clinical trials, given the interplay of factors, might not permit their detection.

Indeed there are various reasons for an ADR escaping detection before approval is given for marketing, which could include failure of the sponsor to do due diligence in respect of all precautionary measures that must be taken before approval is sought; the onus lies on the sponsor to ensure that all the necessary safety measures are met as it is not the primary responsibility of regulatory bodies to conduct safety tests [26, 48, 77]. Also what might eventually happen with the use of a drug, as far as dealing with nature is concerned, might simply be beyond the recognition of man [77].

1.2 Pharmacovigilance

Drug regulatory bodies such as the Medicines and Healthcare Products Regulatory Agency (MHRA) and Commission on Human Medicines (CHM) of the UK and the Food and Drugs Administration (FDA) of the US therefore weigh the potential benefits of a drug against the harms, on the basis of the documentation submitted by the sponsor, and allow drugs to be marketed if it is felt that the potential benefits far outweigh the harms [26, 48], and then monitor the associated adverse events that arise in order to fully characterize the side effects of the drug and to take remedial action where necessary, including advising health care practitioners and the general public on what to do [27, 50]. The World Health Organization (WHO) also has a monitoring centre in Uppsala, Sweden, known as the WHO Collaborating Centre for International Drug Monitoring (Uppsala Monitoring Centre) that, among other things, serves as the unifying point for the drug monitoring activities of various drug regulatory agencies of member countries [86].

There is an ethical implication in approaching drug administration this way in the sense that one could be unwittingly toying with people's lives if the drug is a potential health hazard to, for instance, a minority sub-population who do not react favourably to the drug. However, regulatory bodies and sponsors can for the present hardly do otherwise; the complex process of pharmaceuticals development involves seeking a trade-off between a number of factors which present as 'obstacles' in the process. These impediments, paradoxically, include ethics; one needs, and rightly so, the educated consent of subjects who undergo trials and this could also limit the number of people who volunteer.

The system of tracking the use of drugs throughout their marketed life in order to find unknown harms or changes in adverse reactions associated with it, with the view to take remedial action if necessary, is known as Pharmacovigilance [23, 50].

It involves the detection of hitherto unknown adverse drug reactions including those resulting from drug-drug interactions, through uninterrupted safety surveillance of drugs that are in use, particularly newly approved ones and additional indications (diseases or circumstances of use of drug) [23, 70]; keying out sub-populations of users who are at risk along the lines of "dose, age, gender, underlying disease" [70], drug class, genetic make up and any other relevant variable; superintending proper administration of medications by health professionals and the general public in respect of prescription-only-medicines (POMs) and over-the-counter (OTC) drugs including off-label use [23, 70]; studying the adverse reaction characteristics of a medication relative to those of the same therapeutic class [70] and "providing information to healthcare professionals and patients to optimize safe and effective use of medicines" [50].

The action a regulatory body, in collaboration with a sponsor, may effect takes several forms depending on the enormity and urgency of the problems associated with a medical product, once the problems have been identified. They range from improving precautionary and warning messages on packages and information leaflets, labeling modification; limiting indications, mandatory monitoring of patients, dose modification; and limiting distribution and prescription of product, seeking informed consent of patients; to suspension of distribution and marketing, drug withdrawal from market, banning of product and revoking of licenses [17, 23, 50, 89].

1.2.1 Spontaneous Reporting System (SRS)

One of the main approaches used by regulatory bodies and sponsors to conduct drug safety surveillance is what is generally known as the Spontaneous Reporting System (SRS) [19, 23]. This is known as the Yellow Card Scheme (YCS) and Adverse Event Reporting System (AERS/MedWatch) in the UK and the US respectively. Under this system, health professionals and the general public report adverse events associated with a medication either directly to the regulatory bodies or to drug firms, who must by regulation pass on the information to the regulatory bodies [28, 49].

1.2.2 Problems of the Spontaneous Reporting System

Undoubtedly this system plays a leading role in drug surveillance and is very important in facilitating the identification of rare but bona fide harms associated with medications, especially those that have just been approved [17, 81], as in the case of the link between aplastic anaemia and remoxipride [64]. However it suffers from a number of problems: Adverse events are generally under-reported [4, 70] but the problem appears to be more serious with regard to some adverse events and drugs than with others. For instance in a study of the problem by Alvarez-Requejo et al. [4] in Spain, they found out that serious events tended to be reported more than non-serious events. They found reporting to be higher for newly marketed medications and unclassified events. They also pointed out that under-reporting in the case of psychiatric and gastrointestinal disorders were relatively more pronounced than others. Further, medications belonging to the anti-infective and cardiovascular class were more likely to be cited as being the causative agents of adverse events. They thus concluded that the problem of under-reporting is significant but not uniform across events and medications; it is more likely to involve common and non-serious events, and underscored that this phenomenon, in some way, augurs well for pharmacovigilance as rare or novel but serious adverse reactions are likely to show up in spontaneous reporting system as events with 'unusually' high frequencies, warranting further investigation [4]. However, it must be pointed out that the downside of this phenomenon is that new adverse reactions whose attributes are reminiscent of commonly occurring adverse reactions or diseases could be missed if care is not taken [23, 70], so are adverse reactions whose attributes bear close semblance to the disease under treatment [54]. For instance Moore et al [54] report on the inability of the spontaneous reporting system to either discover mortality increases due to use of flosequinan in congestive heart failure or detect that

cardiac arrest could arise from the use of flecainide and encainide.

There seems to be no clear trend in the reporting rate of adverse events. Promotional activities of pharmaceutical companies, it is thought, influence reporting at times. Media attention resulting from episodes of adverse events could also make the public extra sensitive psychologically, and thereby result in irregular periods of increased reports of adverse events some of which are not real. Additionally, regulatory policy could also tilt the reporting rate in a given direction; regulatory bodies request reporting institutions to be particular about serious and uncommon events, which could bias reporting in favour of these events [4, 6, 70, 81].

Reporting partial and erroneous information on adverse events are also problems that plague the spontaneous reporting system. Variables that are affected include suspect drug, dose, cotherapy, indication, age and gender. Others are the duration of treatment and medical and disease profile of patients [6, 81]. Non-uniformity of reporting practices and conventions from company to company and from country to country or even between regions of the same country and amongst health personnel also make the system less useful than it might be [6, 70, 81].

Multiple reports of adverse event episodes arising from use of multiple channels or inappropriate tracking of events leading to misrepresentation of old cases as new is reported to be common [6, 29, 30]. The number of people actually using a given medication at any point in time is unknown and the current information situation does not permit accurate estimation of it.

1.2.3 Effects of the Problems of Spontaneous Reporting System

The above inadequacies make it impossible to accurately determine incidence rate and prevalence of adverse reactions [4, 70, 81]. It is not easy to establish whether or not the relationship between a drug and an adverse event which occurred during the administration of the drug is causal on the basis of spontaneous reporting system data alone; as the event may have occurred accidentally or have been associated with the disease under treatment. Other factors that may be responsible for the adverse event include an unrecognized disease, other medications being used at the same time [23, 29], or drug-drug interaction [23, 29, 64], such as happen when isoniazid is administered concurrently with rifampicin [64].

1.2.4 Contribution of Spontaneous Reporting System to Pharmacovigilance

The spontaneous reporting system, nonetheless, has played and continues to play an important role in the identification of adverse drug reactions which otherwise would have taken more time or been difficult to identify [6, 17, 54]. The link between troglitazone, tramadol, felbamate, temafloxacin and the respective side-effects of liver damage, seizures and addiction, aplastic anaemia and blood disorders were all established with the aid of the spontaneous reporting system in the 1990s [54]. Other instances where side-effects have been identified through the spontaneous reporting system and have led to regulatory action of one form or the other include the occurrence of tendinitis and tendon rupture in the use of quinolone antibiotics; renal failure induced by the use of aristolochia; "serious cardiovascular reactions" in the use of cisapride (prepulsid, alimix); and "hyperglycaemia, diabetes and exacerbation of diabetes" [17] occasioned by the use of olanzapine (zyprexa).

Indeed, as alluded to above, the spontaneous reporting system is not the only means by which pharmacovigilance is conducted. Any means that has the capability of assisting in establishing that the relationship between a suspect drug and an adverse drug event (ADE) is causal or otherwise can be used. These include laboratory and tolerability data from trials, case-control and cohort studies using data from case registries, general practitioners and hospitals, and vital statistics and information from the coroner or pathologist [17, 23, 70, 77]. However the spontaneous reporting system is, arguably, the most valuable because of the vanguard role it plays in the detection of unknown adverse drug reactions, especially rare ones [81]; which is in line with the primary aim for which the system was set up. It is often the case that the other means are called upon to complement the spontaneous reporting system in causality assessments when the SRS has identified a plausible causal relationship between a drug and an adverse event [89].

1.3 Motivation for this Work

Good health plays a pivotal role in the development of individuals and larger society. This is the justification for governments and local authorities doing all they can to provide health services. These services could be broadly classified as curative or preventive [85]. Curative services involve the treatment of diseases and the preventive include public health administration, which among other things, involves pest control, maintenance of hygiene,

periodic immunization and ensuring that quarantine requirements are adhered to. It also calls for the education of the public on health issues, all with the view to getting the public to act in such a way as to inhibit the growth of health problems if not avert and eliminate them [85].

For a health delivery regime to be total and effective, stakeholders must also address themselves to the possibility of dealing with iatrogenic problems as part of the preventive approach to health delivery. The preventive approach to health delivery holds a number of benefits, viz:

- eliminates unnecessary deaths, pain and discomfort associated with diseases and iatrogenic problems,
- obviates the huge cost in terms of time and funds involved in dealing with diseases and iatrogenic problems, and
- savings in time and funds could be channeled into supporting other areas of the economy needing attention.

The recognition of this situation has spawned such health fields as epidemiology and pharmacovigilance. Pharmacovigilance in particular, and indeed the drug industry in general (including the development process) faces a challenge which when overcome will benefit all. The challenge, as alluded to above, has to do with the detection of unknown adverse drug reactions associated with marketed drugs as soon as practicable. Some adverse drug reactions are not easy to detect after marketing approval because of inaccurate reporting of adverse events and lack of information regarding the population of users of drugs.

Timely detection of new and unknown serious adverse drug reactions would ensure the aforementioned benefits via:

- “reduced morbidity, sick leave days and impaired days
- reduced potential disabilities
- reduced mortality
- less need for hospital capacities
- reduced number of hospital stays and outpatient care" [35].

To place the discussion in perspective and bring to the fore the burden presented by ADRs as a whole and unknown ADRs in particular, we look at some studies that have been conducted on ADRs in general and then concentrate on unknown ADRs:

Using a meta-analysis of 39 studies on adverse drug reactions in the US, spanning three decades, it was estimated that over 1.5 million people were hospitalized in the US in 1994 due to serious adverse drug reactions, over 700 000 experienced a serious adverse drug reaction while on admission for reasons other than ADRs and over 100 000 cases of adverse drug reactions resulted in death in the same year, "making [ADRs one of the] leading cause of death in the United States" [45, 54]. The overall prevalence of ADRs amongst hospital patients was said to be 6.7% (95% CI: 5.2% – 8.2%). Most of these adverse reactions occurred at the required doses or doses considered to be normal in human use [45].

A study involving two general hospitals and a population of 18820 inpatients over a six-month period in the UK, published in 2004, found out that 5.2 percent of the admissions resulted directly from ADRs, with the overall prevalence of ADRs for the study standing at 6.5%. Length of stay was found to have a median of 8 days ($Q_1 - Q_3$: 4 – 18 days), taking up 4% of bed capacity. The study estimated that ADRs could cost the National Health Service an estimated £466 million annually [65]. In a recent publication (2009) of another study with the same duration as above but focusing on ADRs occurring after hospitalization, involving one hospital and two of the authors of the above report (including others), the prevalence rate of ADRs and the associated direct annual cost to the National Health Service was estimated to be 14.7% and £637 million respectively; and 26.8% of the patients stayed longer than expected as a result of experiencing ADRs. They pointed out that the latest estimate of the direct cost, which was arrived at after careful consideration of the circumstances of the study, is consistent with figures from mainland Europe and the US [16].

Muehlberger et al. [55] and Goettler et al. [35] published a twin investigation that focused on frequency, cost and preventability of adverse drug reactions that lead to hospital admissions. The first investigation looked at 25 studies that took place over the previous 25 years and published in English or German. The study concluded, inter alia, that the proportion of hospital admissions attributable to ADRs has a median of 5.8% ($Q_1 - Q_3$: 4.2 – 6%) [55]. The second investigation was a meta-analysis of 13 studies involving several countries of similar health delivery sophistication and published between 1975 and 1996 in English, French or German [35]. This investigation estimated the median length of hospital

stay due to ADRs to be 8.7 days ($Q_1 - Q_3$: 8 – 12.3 days). Based on the information from the two investigations and an average inpatient cost per hospital day of 465 DM in 1995, the researchers estimated that Germany incurred a direct cost of 1.05 billion DM per year from hospitalization occasioned by ADRs around 1995 [35].

The two studies used the WHO definition of an ADR, namely "an adverse drug reaction is a reaction that is noxious and unintended, and occurs at doses used for prophylaxis, diagnosis, or therapy of disease, or for modification of physiological function" [88]. However as acknowledged by the second study [35], total adherence to the above definition, as expected of a review of several studies whose foci were not exactly the same, proved difficult. Thus the above cost most likely includes cost arising from improper use of drugs or some anomaly in their use. Indeed 30.7% of hospitalizations due to ADRs were considered to be preventable [35]. Another meta-analysis by Beijer and de Blaey [9] involving 12 studies put the figure at 28.9% (±0.02%). The scenario above, nonetheless, illustrates the enormity of the problem of serious but unknown ADRs, as it is not inconceivable that a reasonable portion of the cost of the remaining seventy or so percent of the hospitalizations due to ADRs could be attributable to ADRs of this category.

The last two decades have seen a number of studies that estimated frequency of hospital admissions due to ADRs and the associated length of hospital stay and direct cost, and frequency of ADRs amongst inpatients admitted for reasons other than ADRs and the associated extra length of stay and direct cost. The figures from these studies demonstrate that ADRs have been a constant cause of economic loss over the years [8,15,16,38,65]. Table 1.2 gives the estimated annual cost attributable to ADR hospitalizations or experience of ADR while on admission for some selected studies.

Table 1.2: Cost of ADR hospitalization estimated in selected ADR studies.

Setting	Cost per annum (millions)	Reference year for cost	Country to which cost applies	Reference
Meta-analysis of 13 studies	DM 1050	1995	Germany	Goettler et al. [35]
Meta-analysis of 69 studies	£380	1994-1995	England	Wiffen et al. [84]
Meta-analysis of 68 studies	£110 - £256	2000	Netherlands	Beijer and de Blaey [9]
National hospital data	> €226	2001	Spain	Carrasco-Garrido et al. [11]
A major teaching hospital	> £637	2005	England	Davies et al. [16]
National hospital data	> €272	2006	Spain	Carrasco-Garrido et al. [11]

These studies did not follow a commonly accepted approach and metric in identifying and quantifying the problem of ADRs, owing in part to differing circumstances, and so there are no universally accepted estimates and the exact magnitude of the problem is not known [1, 38]. However, the fact that ADRs constitute a huge medical and economic burden is a ubiquitous theme that runs through all of them. The picture becomes even more serious when one considers that the huge cost of ADRs estimated by these studies does not include indirect costs - injuries and intangible costs to patients, costs due to misconduct; liability, claims or litigation costs [1, 35].

The need to search for ways of detecting unknown but serious adverse effects of drugs is all the more exigent when one considers the fact that an appraisal of FDA's drug review process, over the period 1976-1985, released in 1990 reported that 51.5 percent of the drugs approved over the decade entered the market with unknown side effects [37]. In the UK, at least 12 drugs underwent some form of regulatory action between the period from 1992 to 2002 because of discovery of adverse reactions after they had been approved [66] whilst at least 24 drugs were withdrawn from marketing or distribution over the period 1978 - 2001 in the US owing to safety concerns that emerged after approval [89]. There is cause to believe, as will be seen in Chapter 2 and attested to by the continual safety information that comes from the FDA [31] and other regulatory bodies, that the situation would have been worse had it not been for the role drug regulatory bodies play in preventing potentially harmful drugs from getting onto the market, as drug sponsors evolve and explore new bio-chemical agents to help combat medical conditions that have hitherto proven intractable.

The problem of ADRs is multifaceted but can be dealt with from three fundamental angles: taking steps to forestall the occurrence of preventable ADRs, taking steps to ameliorate the effects of ADRs that cannot be prevented and hunting for unknown ADRs, particularly those associated with newly marketed drugs. While acknowledging the inevitability of side-effects, no effort must be spared in curtailing the burden ADRs present and achieving the health imperative of optimizing the risk-benefit ratio of drugs and making sure that risks are handled in an efficient and effective way.

It is in this vein that this work was undertaken. In particular, this work concerns the third of the approaches mentioned above, with a focus on developing a statistical model for analyzing SRS databases, so drug boards, swamped by drug safety data, could be assisted to effectively detect unknown side-effects of drugs as soon as possible.

1.4 Objective(s) of the Research

As pointed out in the foregoing paragraph, the main objective of this work was to explore and develop a statistical model that can assist in the detection of unknown side-effects associated with drug use via the identification of drug and adverse event pairs with higher-than-expected or disproportionate frequencies for further scrutiny. Pursuant to this objective, the research also engaged the following sub-objectives:

1. To identify patterns, if any, in relevant variables connected with the problem of adverse events in drug use and how they relate to it,

2. To identify covariates (of number of reports linking a given drug and adverse event) for screening purposes and

3. To explore the data with the view to unearthing any unsuspected characteristics.

1.5 Outline of the Rest of the Thesis

The rest of the thesis is organized as follows: Chapter 2 presents the results of a preliminary analysis of FDA SRS data with a view to elucidating some of the issues raised in Chapter 1. Chapter 3 provides a brief review of the theory of some key methods used in the main research work while Chapter 4 proposes new quantitative signal detection model(s) and describes their implementation. The chapter begins with an overview of some of the existing quantitative methods. Chapter 5 presents the results of an application of the proposed models to FDA SRS data, with some discussion. A further discussion of the results is presented in Chapter 6, where one of the models is identified as most suitable for the data. Some thoughts on the results, the highlights of the work and the potential for further research are the subject matter of the last chapter, Chapter 7.

Chapter 2

Preliminary Analysis

This chapter presents the results of a preliminary analysis carried out on one of two data sets used in the work reported in this thesis. It first describes the nature of the data set which is the focus of this chapter and gives hints of what is involved in restructuring it for subsequent use, and why the second data set is needed. A description of the second data set is given in Chapter 5. The analysis was carried out to throw more light on some of the issues raised in Chapter 1, pursue the sub-objectives presented in Section 1.4 and gain an appreciation of what it takes to process the raw data into forms suitable for further analysis. The results are first presented with minimal discussion which is then followed by a full discussion of the results and some comments.

2.1 Data: Nature and Treatment

Food and Drugs Adminstration Data, 2004 - 2010

The data which is the subject of the analysis whose results follow were obtained from the Food and Drug Administration's website [29], which makes available to the public data from its Adverse Events Reporting System (AERS) database. The data cover the seven-year period from 2004 to 2010 and were downloaded between April 2010 and April 2011. It is made up of seven anonymized and linked quarterly extracts from the AERS database. These seven quarterly files, which are in ASCII format, come along with four other files (made up of two MS Word documents and two text files) which describe and give further information on the seven data files. The seven data files: Demographic and Administrative, Drugs, Reaction, Outcome, Report Source, Therapy and Indication hold between them 44 variables (including key or link variables) [29], but only 15 of these variables which were

deemed germane to the objectives of this study were used in the analysis. The selected variables are presented in Table 2.1; an expanded version can be found in Appendix A.

The collection process of these data allows for the presence of duplicate records, which must, under some circumstances, be removed before any meaningful analysis could be carried out. One of the informational files (ASC_NTS.DOC) [29] accompanying the data states that:

> ... such "duplicates" ... (multiple reports of the same event) will normally have the same CASE number (but different ISR numbers). Users wishing to remove such duplicates can identify a "best ISR" for the case by use of the CASE and FDA_DT fields - for reports with the same CASE number, select the latest (most recent) FDA_DT. For those occasions when both the CASE and FDA_DT fields are the same, select the report with the higher ISR number. (This procedure would remove "duplicates" not only in the sense just discussed – paper and e-sub reports – but also "duplicates" in the sense of initial and follow-up reports) [29].

The above algorithm was followed to remove the duplicates observations from the data.

Reports of adverse events occurring outside the United States were excluded. So are reports of adverse events occurring in studies (of sponsors) or in the literature. This was to ensure that the remaining data was as homogeneous as possible. As pointed out by Moore et al. [53] in "Serious Adverse Drug Events Reported to the Food and Drug Administration, 1998-2005", reports of adverse events occurring in studies, literature or coming from outside the United States may bring in additional variation because they do not fit the description of 'spontaneous' or had to meet different criteria for reports not originating from within the United States.

The Drugs, Reaction and Outcome files contain variables that are of the 'multiple response' type – the values these variables can take are not mutually exclusive; the variables can hold more than one 'response' per subject. This has an effect on the way the analysis can be carried out. For this reason one is constrained to dichotomize patient outcomes as: death and all other outcomes or hospitalization and all other outcomes et cetera depending on what one is looking for.

Adverse events reports to the FDA are of three types, namely expedited, periodic and direct. Expedited reports are those that concern serious adverse events not described in the product information and are unexpected – "not been previously observed" [31] in the

Table 2.1: Selected variables and their description.

Variable	Description
ISR	Primary link variable identifying each individual report received, across all seven data files.
CASE	Unique number identifying a given subject (record).
FDA_DT	Date report was received by the FDA.
REPT_COD	Holds codes identifying type of report made.
AGE	The age of a subject.
AGE_COD	Holds codes for unit in which age was measured.
GNDR_COD	Holds codes identifying the sex of a subject.
OCCP_COD	Holds codes for the occupation of the original reporter.
E_SUB	Holds codes for mode of submission of report.
ROLE_COD	Holds codes identifying the reported role of drug in adverse event.
DRUGNAME	Holds names of drugs involved in the adverse event, either the "Valid Trade Name" or the "Verbatim" name as stated on report.
VAL_VBM	Holds codes showing whether name of drug is the Valid Trade Name or the Verbatim name.
PT	Holds the Preferred Terms (PT) for describing the observed reaction to drug(s) using the Medical Dictionary for Regulatory Activities (MedDRA).
OUTC_COD	Holds codes identifying the outcome of the adverse experience.
RPSR_COD	Holds codes identifying the initial report source.

Source: ASC_NTS.DOC, US Food and Drugs Administration [29].

context of use of the drug in question. Such adverse events must be reported to the FDA by sponsors within 15 days. Periodic reports concern serious adverse events that are described in the product information. This type of report is usually submitted on a quarterly basis for newly approved drugs. Direct reports refer to those submitted to the FDA without recourse to the sponsor [31].

Reporters of adverse events can express the age of the subject in hours, days, weeks, months, years or decades [29]. All values of age expressed in units other than years were converted to years. Age was then recoded into three groups, viz: 0-17, 18-44, 45-64, and 65 and over. The rationale behind this categorization was to see how the active population

compares with the non-active population.

A reporter of an adverse event is required to specify the sex of a subject as M for male and F for female. Two additional codes (UNK and NS) are provided to cater for situations where sex is not known (can not be determined, such as that of a fetus) or not specified respectively [29].

An attempt was made to cross-classify the data set described above into an $n \times m$ table of drugs against adverse events, stratified on the basis of sex, age and time for use in developing the model for generating hypotheses about the relationship between drugs and adverse events, but it became clear that the data required careful clean-up. The clean-up include correcting misspellings of names of drugs and adverse events, finding all descriptions in the data that point to the same adverse event and recoding them into the Medical Dictionary for Regulatory Activities (MedDRA) Preferred Term (PT) [83] and finding all the different names in the data that indicate the same drug (active ingredients) and recoding them into a commonly accepted drug name. As noted by DuMouchel [20], it is an exercise that is time-consuming even with expert opinion. We therefore turned to a second data set, which was ready for use. This second data set is described in Chapter 5.

The analysis of the data described above was carried out with the aid of SAS software [76] and R software [68]. The SAS software was key in dealing with database and data processing issues.

The data described above are secondary; as such we cannot understand the data to the same extent as those who compiled it. Indeed it has to be pointed out (as alluded to elsewhere) that the data comes with a couple of challenges, notably missing values; a considerable number of the cases have missing values for some of the variables. The extent of the problem, in some cases, may call into question the validity of the analysis in respect of these variables. However, we proceeded to assess the situation based on the available data as there is hardly any other way of obtaining data on these variables. At least they shed some light on the nature of the problem of adverse events associated with use of drugs and their reporting.

2.2 Results of Preliminary Analysis

2.2.1 Overall Number of Reports and Trend Over Time

After removing reports of adverse events coming from foreign sources, studies or occuring in literature, a total of 1,919,848 adverse events reports remained for analyses for the seven-year period under consideration. Annual volume of adverse events reports to the FDA more than tripled from 165,229 in 2004 to 501,778 in 2010 (Table 2.2). This translates to an average annual increase of 20.3%, and in the last year (2010) alone the number of reports rose by 55.9%, relative to that of 2009.

Table 2.2: Annual and overall values for death, other outcomes and all reported adverse events.

Year	Total (all events)	Number of deaths	Deaths %	Other outcomes %	Total (excluding cases with missing values)	Deaths %	Other outcomes %
2004	165,229	16,008	9.7	90.3	103,597	15.5	84.5
2005	190,147	18,924	10.0	90.0	134,298	14.1	85.9
2006	208,693	17,928	8.6	91.4	146,441	12.2	87.8
2007	243,219	18,622	7.7	92.3	159,438	11.7	88.3
2008	288,845	26,776	9.3	90.7	183,292	14.6	85.4
2009	321,937	34,216	10.6	89.4	219,226	15.6	84.4
2010	501,778	54,327	10.8	89.2	313,477	17.3	82.7
Total	1,919,848	186,801	9.7	90.5	1,259,769	14.8	85.2

The trend in the number of reports submitted per 10,000 people in the US population over the years is as shown Figure 2.1. The trend in the figure indicates that the number of adverse event reports received per year is growing at a faster rate than the population size. The growth is particulary high for 2010. The increasing trend could be due to increasing awareness of the need to report adverse events or increasing incidence of adverse events or both.

2.2.2 Patient Outcomes

For the seven-year period under consideration, 660,079 (34.4%) out of the total of 1,919,848 cases (Table 2.3a) had missing patient outcome values.

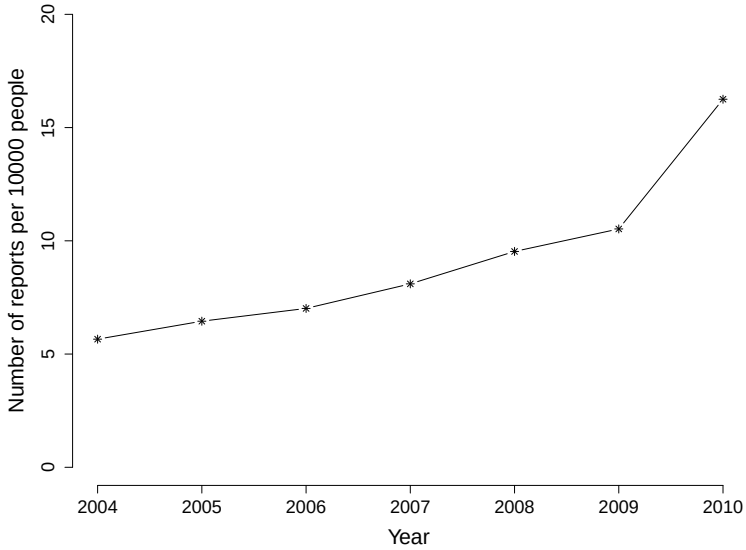


Figure 2.1: Number of reports per 10,000 people against time, 2004 – 2010.

Table 2.3a: Patient Outcomes, 2004-2010.

Cases					
Valid		Missing		Total	
Count	%	Count	%	Count	%
1,259,769	65.6	660,079	34.3	1,919,848	100

Of the remaining 1,259,769 cases, 186,801 (14.8%) had an outcome of death (Table 2.3b). As pointed out in Section 2.1, reporters of adverse events are permitted to describe the event by more than one of the outcome categories. This explains why the sum of the figures in Table 2.3b exceed 100%. The overall trend in the number of deaths suggest an increase over time, but the actual number of deaths decreased in 2006 (compared to that of 2005 in spite of increase in the number of reports) and then picked up from 2007 onwards. Figures 2.2a and 2.2b present a comparison of the trends in the annual number of deaths, other outcomes, all cases and all non-missing cases; and Figure 2.3 shows the annual percentage deaths for all the cases, and for all non-missing cases. Figures 2.2b and 2.3 were produced as a way of assessing the effect, if any, of the problem of missing values on the trend in the number of deaths over time.

Over the period, the proportion of reported cases of adverse events which had an

Table 2.3b: Patient Outcomes, 2004-2010.

Outcome	Cases	%
Death (DE)	186,801	14.8
Life-threatening (LT)	79,207	6.3
Hospitalization - initial or prolonged stay (HO)	514,807	40.9
Disability (DS)	61,155	4.9
Congenital anomaly (CA)	9,490	0.8
Required intervention to prevent permanent impairment or damage (RI)	38,614	3.1
Other (OT)	670,991	53.3

outcome of death in the US assumed a low of 11.7% in 2007 and a high of 17.3% in 2010. The figures come to a low of 7.7% in 2007 and a high of 10.8% in 2010 if the denominator is changed to number of all cases, assuming none of the missing outcomes is death. Of course the estimates would be higher if we assume that all the missing outcomes values are deaths – either of these assumptions is hardly possible, though. But this approach allows us to see what conservative estimates are likely to be.

The proportion of reported cases of adverse events in the US in which the subject had a hospital admission (initial or resulting in prolongation of hospital stay) over the period under consideration stands at 40.9% (excluding missing cases). This high figure just shows that spontaneous reports are more likely to involve serious adverse events. The minimum for the period is 39.7% and the maximum is 43.4%. The corresponding values when the denominator is changed to all cases, as was done for death, are 24.8% and 30.5% respectively. Values for other outcomes are presented in Appendix B (Tables B.1 and B.2).

The trend in the bars plotted with percentages determined from the number of all annual adverse events (lightblue colour) compares with that plotted with percentages determined from the number of annual non-missing cases (mistyrose colour) except that of 2005 (Figure 2.3). Thus the missing values appear not to have had any serious effect on the proportion of deaths reported.

2.2.3 Occupation of Reporters

FDA requires the identification of the occupation of the original reporter of a case of adverse event, submitted directly or not. The occupation of the original reporters of 290,833 (15.1%) of the cases were not accounted for (missing), for the period under consideration

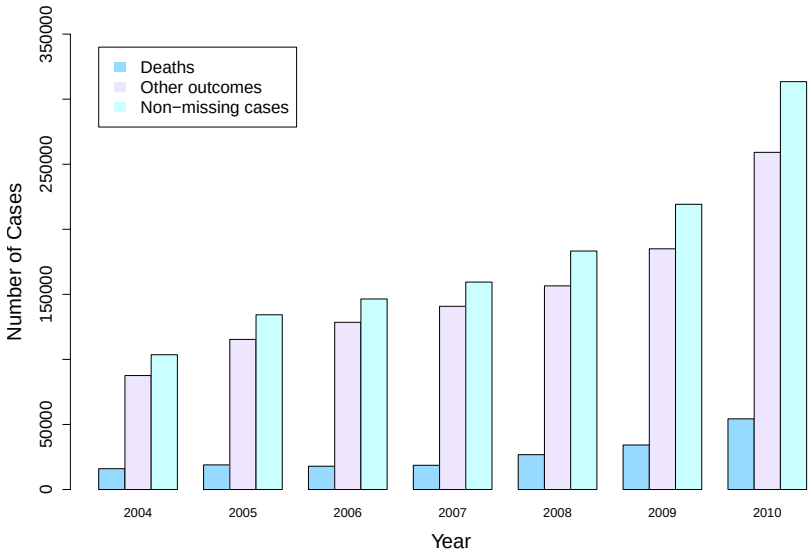


Figure 2.2a: Chart showing the trends in the number of deaths, other outcomes and all non-missing cases.

(Table 2.4a).

Table 2.4a: Occupation of original reporters, 2004-2010.

Cases					
Valid		Missing		Total	
Count	%	Count	%	Count	%
1,629,015	84.9	290,833	15.1	1,919,848	100

Health professionals were the original reporters of more than half (869,507, 53.4%) of the remaining 1,629,015 cases (Table 2.4b). The annual values follow a similar structure except 2007 when the non-professional source (consumers or their legal representatives) was greater than the professional one by 3.6% (Figure B.1).

2.2.4 Types of Report

Of the total 1,919,848 reports submitted in the period under discussion, 186,636 (9.7%) were direct reports, 949,291 (49.5%) were expedited reports and 783,921 (40.8%) were periodic reports (Table 2.5).

Expedited reports have been accounting for the highest number of reports since 2005

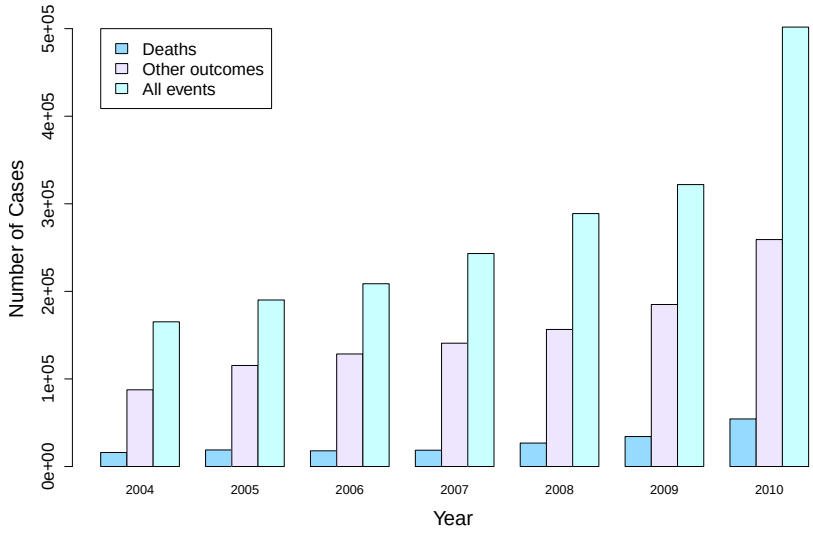


Figure 2.2b: Chart showing the trends in the number of deaths, other outcomes and all events.

Table 2.4b: Occupation of original reporters, 2004-2010.

Occupation	Cases	%
Physician (MD)	461,287	28.3
Pharmacist (PH)	120,930	7.4
Other Health-Professional (OT)	287,290	17.6
Lawyer (LW)	66,087	4.1
Consumer (CN)	693,421	42.6

when it overtook periodic reports, with direct reports accounting for the least number of reports throughout the period under discussion. The reporting structure has remained pretty much the same since 2005 although the 2009 percentage for expedited report increased by 6.1 percentage points over that of 2008 and then dropped by 4.2 points in 2010 and that of periodic dropped by 5.2 points in 2009 then increased by 9.0 points in 2010 (Figure B.2, Appendix B).

2.2.5 Mode of Submission of Reports

Overall, 68.6% (Table 2.6b) of the reports were submitted via the internet, with the remaining 31.4% in hard copy. The trend in the annual figures suggest that more reports are increasingly being submitted via the internet; while the number of reports submitted by

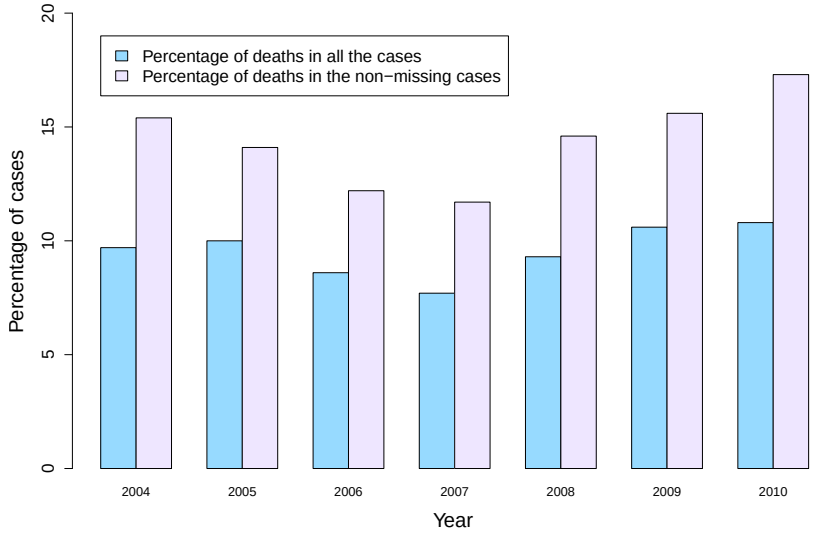


Figure 2.3: Chart showing the trends in the percentage of death in all the reports and in the non-missing cases.

Table 2.5: Report types, 2004-2010.

Type of report	Cases	%
Direct	186,636	9.7
Expedited	949,291	49.5
Periodic	783,921	40.8

this mode accounted for 33.1% in 2004, it had grown by 53.8 percentage points to 86.9% by 2010.

2.2.6 Sex of Subjects

As much as 160,054 (8.3%) of the reports submitted over the period had the sex of the subject to be indeterminate – the sex of the subject is not known, not specified or missing (Table 2.7a).

Of the 1,759,794 (91.7%) remaining reports, 61.3% were female and 38.7% were male (Table 2.7b). The dominance of reports on female subjects, which is consistent with the observation of Wysowski and Swartz [89], was a constant feature over the seven-year period, contributing about three-fifth of the total number of reports annually (Figure B.3, Appendix B).

Table 2.6a: Report format, 2004-2010.

Cases					
Valid		Missing		Total	
Count	%	Count	%	Count	%
1,919,846	100.0	2	0.0	1,919,848	100

Table 2.6b: Report format, 2004-2010.

Electronic submission	Cases	%
Yes	1,316,720	68.6
No	603,126	31.4

2.2.7 Age of Subjects

As can be seen from Table 2.8a, the ages of a large number 833,188 (43.4%) of the cases were unaccounted for (missing). Of the remaining 1,086,660 cases (Table 2.8b), 65,916 (6.1%) were in the group 0-17, 268,273 (24.7%) were in the group 18-44 and the groups 45-64 and 65 and over accounted for 37.9% (411,978) and 31.3% (340,493) respectively. A study of the annual data shows an increasing trend in the number and percentage of missing values up to 2009. The percentage missing dropped from 55.4 in 2009 to 40.6 in 2010. However the age structure of the annual non-missing cases has not only been fairly consistent from year to year since 2005 but also with the overall age structure for the period under consideration (Figure B.4, Appendix B).

2.2.8 Age and Sex Load of Adverse Events

Table 2.9 presents a cross classification of the number of reports on the basis of age and sex. It also presents percentage values for the size of the age groups in the overall US population, adjusted percentages (expected) for likelihood of drug use and the 'proportion' (p) of each of these age groups in the overall number of reports relative to the size of these age groups in the overall US population. The 'proportion' p for each age group was found in two stages: the number of non-missing cases for each age group for a particular year was divided by the number of people in that age group in the US population for that year, and then multiplied by 10,000. This gives the 'proportion' for that age group for that year.

Table 2.7a: Sex of subjects, 2004-2010.

Cases					
Valid		Missing		Total	
Count	%	Count	%	Count	%
1,759,794	91.7	160,054	8.3	1,919,848	100

Table 2.7b: Sex of subjects, 2004-2010.

Sex	Cases	%
Female	1,079,434	61.3
Male	680,360	38.7

The value of p for an age group for the period under consideration is then given by the geometric mean of the annual 'proportions' for that age group.

Table 2.9 shows that even though there are generally more reports on females than on males, the age group 0-17 is an exception; it has more reports on males than on females. It is also clear that the percentages of people 17 years or younger and between 18 and 44 years inclusive reported to be involved in adverse events are less than that in the overall US population, even when it has been adjusted for potential drug use. The reverse is true for the percentages of people between 45 and 64 years inclusive and 65 years or older; they are greater than the percentage in the overall US population and that of the adjusted values. The percentage for the 'active' age group 18-64 (that is combining the age groups 18-44 and 45-64 together) is almost on par with both the adjusted and overall US population percentages (Figure B.5, B). This result is consistent with that observed by Moore et al [53]. It is also worthy of note that the value of the 'proportion', p, increases down the table. A graphical rendition of the pattern in the values of the 'proportion' over the period under consideration is presented in Figure B.6 of Appendix B.

Table 2.8a: Age of subjects, 2004-2010.

Cases					
Valid		Missing		Total	
Count	%	Count	%	Count	%
1,086,660	56.6	833,188	43.4	1,919,848	100

Table 2.8b: Age of subjects, 2004-2010.

Age range	Cases	%
≤ 17	65,916	6.1
18 - 44	268,273	24.7
45 - 64	411,978	37.9
≥ 65	340,493	31.3

Table 2.9: Age and sex load of adverse events, 2004-2010.

Age range	Female		Male		Total		2010 US Pop'n $^{\zeta}$ %	Exp'd cases$^{\xi}$ %	Prop'n p
	Cases	%	Cases	%	Cases	%			
≤ 17	27,681	2.6	34,978	3.3	62,659	5.9	24.0	12.5	1.2
18 - 44	177,882	16.6	86,571	8.1	264,453	24.7	36.5	28.3	3.2
45 - 64	2,425,424	22.9	161,538	15.1	406,962	38.0	26.4	35.1	5.5
≥ 65	192,462	18.0	143,340	13.4	335,802	31.4	13.1	24.1	12.3
Total	643,449	60.1	426,427	39.9	1,069,876	100.0	100.0	100.0	

$^{\zeta}$ *From 2010 US population census [13].* $^{\xi}$ *Population adjusted for potential drug use based on the 2005-2008 data on prescription drug use [57].*

2.3 Discussion and Comments

We saw in Section 2.2.1 that the number of adverse events reported is growing at an annual rate of 20.3% and it more than trebled over the period under consideration, with a 55.9% increase in 2010 alone, over the 2009 figure. The rate of growth in the number of reports is relatively fast for a population which is growing at an approximate rate of 0.93% per annum (based on the 2000 and 2010 US population census) [13]. This could be explained in two ways: either there is growing awareness of the need to contribute to the pharmacovigilance process and hence the increase in the number of reports, when viewed against the well know phenomenon of under reporting [4, 70] or the number of adverse events is growing at an increasing rate or both.

The trend in the plot of the number of reports submitted per 10,000 people as captured in Figure 2.1 suggests the number of reported cases is likely to continue to rise after 2010. However the increase in the number of reported cases after 2010 is not likely to be as much as that of 2010 or will be similar to the pattern before 2010 given the exceptional nature of the increase in 2010.

Figures 2.2a and 2.2b appear to suggest an increasing trend in the number of cases with an outcome of death reported over time even though there was a decrease in the number of reported deaths from 2005 to 2006, in spite of an increase in number of reported cases of adverse events from 2005 to 2006. Figure 2.3 shows that the percentage of reported cases of adverse events that resulted in death in the US decreased from 2005 to 2007, and then began to pick up in 2008.

From Section 2.2.2, one could infer that of the reported cases of adverse events in the US in the period 2004 to 2010, at least 7.7% had an outcome of death and 24.8% had an outcome of hospital admission (including prolongation of hospital stay).

The percentage of reports coming from consumers and lawyers is reasonably high compared with that of health professionals (Section 2.2.3) and appears to suggest that there is an appreciable level of awareness of the need to contribute to the pharmacovigilance process amongst consumers and lawyers (compared to health professionals). This is in spite of the fact that there is under-reporting generally.

Expedited reports have been in the majority since 2005 (Section 2.2.4). The preponderance of expedited reports may be pointing to the lingering problem of uncommon but serious adverse drug reactions, and underscores the urgency of finding a means to curtail if not eliminate it.

The increasing trend in the number of reports submitted electronically reflects the growing importance of this unprecedented mode of information exchange in general, and in pharmacovigilance in particular. Effort at improving the monitoring system should not only focus on taking advantage of the speed of this mode of reporting but also on sensitizing the stake holders, especially the general public, on the need to do accurate reporting as improper description of drugs and adverse events is a problem militating against the optimal and timely use of this unique source of information.

The picture presented by the values in Table 2.7b (Section 2.2.6) appears to raise a number of questions: Are females more susceptible to adverse events than males, generally speaking? This question arises because 2010 population census of the US puts the respective percentages of males and females at approximately 49.2 and 50.8; a ratio of almost one to one. So why should the ratio of male cases to female cases be roughly 2:3 for cases whose sex are known? Could this be attributed to the phenomenon of missing cases? Could there be a reason for this phenomenon affecting males more than females?

We observed in Section 2.2.8 that the values of the percentages of the age groups 0-17 years and 65 years or older involved in adverse events do not compare with those in the overall US population, even when adjusted for drug use. More specifically the values are comparatively smaller for the group 0-17 and comparatively greater for the group 65 years or over. This state of affairs is buttressed by the increasing trend in the value of p down the table (Table 2.9), which together appear to demonstrates that the older segment of the population are more susceptible to adverse drug reactions than the younger segment.

The situation of not being able to account for the ages of a large number of the cases as seen from Table 2.8a does not augur well for pharmacovigilance, as the foregoing observations appear to lend credence to the fact that age and sex are potentially serious candidates when it comes to keying out factors that are linked to occurrence of adverse events [29], and hence their use as covariates of the number of reports in modeling and inference regarding signaling drug and adverse event pairs that have higher than expected frequencies.

Chapter 3

Review of Background Theory

This chapter presents a brief review of the theory of some key methods used in the rest of the work.

3.1 Bayesian Inference

The Bayesian approach to inference allows one to factor into an analysis prior knowledge or at least one's prior beliefs about the parameter of interest. The parameter is thought of as a random variable as opposed to the frequentist view of having a fixed but unknown value. Interest lies in the posterior distribution $\pi(\theta|\mathbf{x})$ of the parameter $\theta = (\theta_1, \theta_2, \cdots, \theta_p)$ given the observed data $\mathbf{x} = \{x_1, x_2, ..., x_n\}$. The posterior distribution represents the most current state of knowledge which results from combining the prior distribution with the observed data $\mathbf{x}$ in the form of the likelihood [33, 44].

3.1.1 Bayes' Theorem

Given the prior distribution $h(\theta)$ and the likelihood $f(\mathbf{x}|\theta)$, Bayes' theorem leads to the following:

$$\begin{aligned} \pi(\theta \mid \mathbf{x}) &= \frac{f(\mathbf{x}|\theta)h(\theta)}{\int f(\mathbf{x}|\theta)h(\theta)\,d\theta} \\ &\propto f(\mathbf{x}|\theta)h(\theta) \end{aligned} \tag{3.1}$$

Expression (3.1) forms the basis of Bayesian inference [10, 44].

3.1.2 Prior Specification

A key consideration in the specification of a prior distribution is that it should sufficiently represent the prior knowledge (before the 'observation' of data) about the parameter of interest. Another thing that impinges on the choice of prior distributions is the potential for simplifying the derivation of posterior distributions and the associated computations [10]. In this regard the prior is often chosen so that the posterior belongs to the same class of distributions as the prior distribution, with the posterior differing from the prior in terms of only the parameters. Such a prior is said to be conjugate to the likelihood [10, 47]. For instance independently and identically distributed data $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ from an exponential distribution $\texttt{Exp}(\theta)$ with $\theta \sim \texttt{Ga}(\alpha, \beta)$ a priori, leads to the posterior $\texttt{Ga}(\alpha + n, \beta + n\bar{\mathbf{x}})$. Thus $\texttt{Ga}(\alpha, \beta)$ is conjugate to $\texttt{Exp}(\theta)$.

A prior chosen to reflect the paucity of knowledge about the parameter of interest is referred to as a noninformative prior. For example, if we do not have information about a discrete parameter, θ, except that $\theta = \{\theta_1, \theta_2, \ldots \theta_n\}$, then we could assign the discrete Uniform distribution $p(\theta_i) = 1/n$, $i = 1, 2, \ldots n$, on θ. Such a prior distribution is regarded as noninformative as it assigns the same probability value to all the possible values θ, yielding no further information about θ [10]. One could also specify a hierarchy of priors by placing a secondary prior (hyperprior) on the prior parameters (hyperparameters) in lieu of noninformative priors as a way of diffusing the effect of any prior assumptions on the posterior [44, 47]. Given independently and identically distributed data $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ from the Poisson distribution $\texttt{Pois}(\lambda)$, the prior specification $\lambda \sim \texttt{Ga}(a, b)$ with $a \sim \texttt{Exp}(\theta_a)$ and $b \sim \texttt{Inv-Exp}(\theta_b)$ is an instance of a hierarchical prior.

3.1.3 Prior Sensitivity

The question of how the posterior distribution is affected by different prior specifications often comes up in Bayesian analysis. As prior elicitation is not easy, particularly in cases where prior knowledge about the parameters of interest is scant, it is useful to assess how robust the posterior is to different prior specifications. This assessment is referred to as prior sensitivity analysis [10, 44]. For example, how does the posterior resulting from an exponential distribution prior compare with that resulting from an inverse exponential distribution prior.

3.1.4 Hierarchical Models

As alluded to above, a Bayesian hierarchical model is one in which the prior is made up of multiple levels of probability distributions. In other words, at the least, the prior parameters are themselves regarded as random variables, which are assumed to come from some given probability distribution. This approach is useful for modeling in a relatively simple and effective way data with complicated structure, such as data on a universe of interest made up of interrelated subgroups or units with unique characteristics [10, 33, 60]. The mean effect ψ_j of group j, $j = 1, \ldots, J$ are estimated with the data of all the other groups under the assumption that the mean effects ψ_j are drawn from a common distribution G. That is $\psi_j|\varphi \sim G(\varphi)$, where φ is the mean effect of the common distribution G [60]. The pooling of the data from the groups results in a pulling of the ψ_j's towards the mean effect of the common distribution from which the groups obtain. This phenomenon is referred to as the shrinkage effect. Thus this modeling technique takes into consideration the interdependence between the cognate parameters ψ_j to produce results that are thought to be more reliable [33, 60].

3.1.5 Posterior Inference

Point estimates such as mean, median and mode of posterior parameters are commonly used to describe the location of the posterior distribution while the spread is estimated by the standard deviation (or the variance). Also used to describe the spread is the central $100(1-\alpha)\%$ credible interval – a Bayesian analogue of the classical confidence interval [33, 44]. The choice of which summary statistics to use is often dictated by the problem on hand (or the data) and the objectives of the analysis.

3.2 Stochastic Simulation

As with all probability distributions, a knowledge of the moments and quantiles of the posterior distribution (or associated conditional and marginal distributions) may be required to understand it. Estimating these quantities from samples drawn from the posterior is crucial in this regard where it is not possible to do so analytically, especially in high-dimensional problems [10,33]. The technique of using simulation to obtain summary estimates of a given distribution is referred to as the Monte Carlo method. The Law of Large Numbers ensures that the estimates so obtained tend to the true values of the estimands as the sample size

grows large [10]. There are routines that facilitate the generation of independent identically distributed random samples from standard distributions. However nonstandard posterior distributions arise that cannot be sampled using these routines (they play an intermediate role at best) and another approach must be used to sample these distributions [33]. The methods of Markov Chain Monte Carlo (MCMC), based on Markov chains provide a convenient way of sampling from posterior distributions under such circumstances, especially in high-dimensional problems [10, 33].

3.2.1 Markov Chains

A Markov chain is a sequence of correlated random variables $\{\theta^{(t)}\}_{t=0}^{\infty}$ such that

$$P\left(\theta^{(t+1)}|\theta^{(0)},\theta^{(1)},\theta^{(2)},\ldots,\theta^{(t)}\right) = P\left(\theta^{(t+1)}|\theta^{(t)}\right) \tag{3.2}$$

Thus the conditional distribution of the next state $\theta^{(t+1)}$, given all the past states, depends only on the current state $\theta^{(t)}$ [44, 72]. The right hand side of the conditional probability (3.2), written as $\mathcal{K}\left(\theta^{(t)},\theta^{(t+1)}\right)$, is referred to as the transitional kernel. The $\theta^{(t)}$s take values from a set Ω called the state space [72].

For a Markov chain to be useful in the Bayesian setting, it must have a stationary distribution or converge to a target probability distribution, which distribution is the posterior distribution $\pi(\theta|\mathbf{x})$ (equivalently written as $\pi(\theta)$) from which samples are desired. Stationarity implies that the Markov chain is irreducible and aperiodic (a condition that is satisfied if $\mathcal{K}(\theta,\phi) > 0,\ \forall\,\theta,\phi \in \Omega$, including the possibility of the chain remaining in the current state). A Markov chain with the above attributes will converge to the limiting distribution $\pi(\theta)$ irrespective of the initial state [44, 72]. It follows from the Law of Large Numbers that given the realizations $\theta^{(1)},\theta^{(2)},\cdots,\theta^{(n)}$ from the steady state of a Markov chain and any integrable function g [44, 72]:

$$\bar{g}_n = \frac{1}{n}\sum_{i=1}^{n} g(\theta^{(i)}) \rightarrow \mathbb{E}_\pi[g(\theta)] \quad \text{as} \quad n \rightarrow \infty \tag{3.3}$$

Relation (3.3) expresses what is called the Ergodic Theorem [44, 72]. It guarantees that one can estimate features of the distribution $\pi(\theta)$ such as $\mathbb{E}_\pi[g(\theta)]$ using time averages.

3.2.2 Metropolis-Hastings (MH) Algorithm

In the Metropolis-Hastings (MH) algorithm, we construct the transition kernel $\mathcal{K}(\theta,\phi)$ by finding a candidate generating distribution $q(\phi|\theta)$, which generates the candidate value ϕ

of the chain dependent only on the current value θ. If the reversibility condition (3.4) is satisfied by the Markov chain so created, then the chain will converge to $\pi(\theta)$ [14, 44, 72].

$$\pi(\theta)\mathcal{K}(\theta,\phi) = \pi(\phi)\mathcal{K}(\phi,\theta) \tag{3.4}$$

To achieve this, a candidate value ϕ, generated according to the proposal distribution $q(\phi|\theta)$, is accepted with probability $\alpha(\theta,\phi)$ [14, 44] given by

$$\alpha(\theta,\phi) = \min\left\{1, \frac{\pi(\phi)q(\theta|\phi)}{\pi(\theta)q(\phi|\theta)}\right\} \tag{3.5}$$

The MH algorithm is thus given by

1. Given the current state $\theta^{(t)}$, generate a candidate value $\phi \sim q(\phi|\theta^{(t)})$ and $u \sim U(0,1)$.
2. Set $\theta^{(t+1)} = \phi$ if $u \leq \alpha(\theta,\phi)$, else set $\theta^{(t+1)} = \theta^{(t)}$
3. Return to Step 1.

The MH algorithm can be used to update a single parameter or a block of parameters at a time (often used to reduce correlation between parameters). It is to be noted that the algorithm can be used even if the normalizing constant of the target distribution is not known since it occurs in the numerator and denominator of the ratio of the target in $\alpha(\theta,\phi)$. However, the proposal density $q(\phi|\theta)$ has influence on the performance of the algorithm and its selection typically requires some experimentation. In this regard the proposal density should be such that the resulting Markov chain can explore the whole support of the target distribution in good time [14, 44, 72].

The MH algorithm takes several forms which are determined by the proposal density used. Two such forms, Random Walk Metropolis and Gibbs Sampling, which are relevant to this work, are described below.

Random Walk Metropolis Algorithm

In the random walk Metropolis algorithm, the candidate value is given by $\phi = \theta^{(t)} + \delta$ where $\delta \sim g$ is a random displacement and g is some distribution [14, 44, 72]. The proposal density is then $q(\phi|\theta) = g(\phi - \theta)$. Symmetric distributions such as the normal, uniform and the t distribution are commonly used for g as they simplify the acceptance probability $\alpha(\theta,\phi)$ if they are centered on zero [14, 44, 72]. In such a case $q(\phi|\theta) = q(\theta|\phi)$ and the acceptance probability is given by

$$\alpha(\theta,\phi) = \min\left\{1, \frac{\pi(\phi)}{\pi(\theta)}\right\} \tag{3.6}$$

Gibbs Sampling

Given the target distribution $\pi(\theta)$ where $\theta = (\theta_1, \theta_2, \cdots, \theta_p)$, the distribution $\pi(\theta_j|\theta_{-j})$ of θ_j, conditional on the rest of the parameters $\theta_{-j} = (\theta_1, \theta_2, \cdots, \theta_{j-1}, \theta_{j+1}, \cdots, \theta_p)$, is called the full conditional distribution of θ_j. Full conditional distributions are the main ingredients of a special case of the MH algorithm called Gibbs sampling [32, 44, 72]. The transition from $\theta^{(t)}$ to $\theta^{(t+1)}$ in Gibbs sampling is as outlined below [32, 44, 72]:

$$
\begin{aligned}
&\text{Generate } \theta_1^{(t+1)} \text{ from } \pi\left(\theta_1|\theta_{-1}^{(t)}\right)\\
&\text{Generate } \theta_2^{(t+1)} \text{ from } \pi\left(\theta_2|\theta_1^{(t+1)}, \theta_{3,}^{(t)}, \cdots, \theta_p^{(t)}\right)\\
&\text{Generate } \theta_3^{(t+1)} \text{ from } \pi\left(\theta_3|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_{4,}^{(t)}, \cdots, \theta_p^{(t)}\right)\\
&\qquad\vdots \qquad\qquad \vdots\\
&\text{Generate } \theta_j^{(t+1)} \text{ from } \pi\left(\theta_j|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \cdots, \theta_{j-1}^{(t+1)}, \theta_{j+1}^{(t)}, \cdots, \theta_p^{(t)}\right)\\
&\qquad\vdots \qquad\qquad \vdots\\
&\text{Generate } \theta_p^{(t+1)} \text{ from } \pi\left(\theta_p|\theta_{-p}^{(t+1)}\right)
\end{aligned}
$$

Gibbs sampling is relatively easy to use if all the full conditionals are of standard form. However computational problems may arise if some of the full conditionals are not of standard form and are not easy to sample from using existing routines. In which case the (general) MH algorithm may be used to sample the nonstandard conditionals. Gibbs sampling is therefore often used in conjunction with the MH algorithm. On the other hand the MH algorithm requires some calibration of the proposal density, which is not required in Gibbs sampling [44, 72].

3.2.3 Convergence and Related Issues

The essence of MCMC rest on the following: Given a Markov chain $\{\theta^{(t)}\}_{t=0}^{\infty}$ with the attributes outlined in Section 3.2.1, there is some value k such that for $t > k$, $\theta^{(t)}$ is thought to come from the target distribution $\pi(\theta|\mathrm{x})$. The chain is then said to have converged and realizations from the steady state, which are seen to be coming from the posterior distribution, can be used for inference [33, 44, 72].

The algorithm then must be run long enough to reach k and to ensure that the starting values have no influence on the draws. The first k iterates are then discarded as they are

assumed not to come from the posterior distribution. This initial part of the run which is discarded is referred to as the burn-in [33, 44, 72]. The question then is how large should be the burn-in? Linked to the forgoing question is the performance of the Markov chain. These and other related issues are explored in the following sections:

Trace Plots

A basic approach in deciding the length of the burn-in is to examine the trace plots of the parameters and posterior density. The practice is to run multiple chains from dispersed starting points and then display the parameter values from different runs on the same plot. As the stationary distribution is unique, the trajectory of the trace plots of the different runs should gravitate towards a common range of values with increasing number of iterations. The point at which these trace plots merge should give an idea of the length of the burn-in [33, 44, 60]. Running multiple chains also provides some guarantee that all the modes of the posterior distribution have been visited, if the chains assume a common range of values after some reasonable number of iterations. This is critical in a multi-modal posterior as a mode may be missed by a single chain leading to the wrong impression that the stationary state has been reached [33, 44].

Gelman-Rubin Diagnostic

One could also use the Gelman-Rubin (GR) criterion to assess convergence. It revolves around the idea that the draws from the multiple chains are coming from the same distribution at convergence and hence the between-chain variance and the within-chain variance should be roughly equal [33, 44]. The criterion comes in different forms. One form in Gelman et al. [33] suggest running m chains and drawing a sample of n iterates apiece. The between-chain B and within-chain W variances are then computed as follows:

$$\bar{\psi}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\psi_{ij} \qquad\qquad \bar{\psi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\bar{\psi}_{.j}$$

where $\psi_{ij}, i = 1, 2\ldots, n, j = 1, 2\ldots, m$ are the draws for parameter ψ.

$$\mathrm{s}_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\psi_{ij} - \bar{\psi}_{.j}\right)^2$$

$$B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\bar{\psi}_{.j} - \bar{\psi}_{..}\right)^2 \qquad\qquad W = \frac{1}{m}\sum_{j=1}^{m}\mathrm{s}_j^2$$

A measure of agreement of the two variances is given by the GR statistic [33]:

$$\hat{R} = \sqrt{\frac{n-1}{n} + \frac{B}{nW}}$$

Their recommendation is to run the chains until the $\hat{R}$ for each of the parameter under consideration is close to one. The next nm iterates could then be used for inference. Values of $\hat{R}$ below 1.1 are said to be acceptable, with higher levels of accuracy required for sensitive applications [33].

Mixing, Autocorrelation and Acceptance rate

The parameter values drawn from a Markov chain are correlated [72]. Given sample iterates $\theta^{(1)}, \theta^{(2)}, ...\theta^{(n)}$ from a stationary Markov chain: the autocorrelation in the iterates, the mixing attributes of the chain, the acceptance rate and convergence of the chain are all intertwined. A gentle decline in the autocorrelation values, which can be ascertained with an autocorrelation function (acf) plot, is an indication of a slow mixing chain [14,44,60,72]. What it means is that either the proposal steps in the MH algorithm are small, leading to proposal values that are similar to the current value and are therefore accepted, with corresponding high acceptance rate, or the proposals are far out in the periphery of the parameter space leading to a rejection. This will lead to a low acceptance rate and a repetition of the same values over a number of iterations. Both situations will lead to a delayed convergence. On the other hand a sharp decline in the autocorrelation values is an indication of a good mixing chain, a reasonable acceptance rate and a fast convergence [14, 44, 60, 72].

The performance of a chain can be improved by adjusting the proposal variance. An increase in the proposal variance is associated with a decrease in the acceptance rate and vice versa [14, 44]. Mixing can also be improved by jointly updating (block update) parameters that are a posteriori correlated. Parameters that are correlated tend to constrain the range of acceptable values of each other if they are updated individually [44].

Beyond adjusting the proposal variance, high autocorrelation could be dealt with by changing the candidate generating density altogether or by thinning – keeping only every j^{th} iterate value from the sampled sequence [44, 60].

Chapter 4

Data Models

The need for timely release of information about side-effect of drugs by drug regulatory agencies, after marketing approval has been given, has led to the development of a number of statistical tools aimed at assisting in the generation of signals – "hypothesis about a possible drug safety problem" [87]; manual evaluation of the case reports by expert panels is prohibitive both in terms of the time and cost, owing to the volume of data in a typical pharmacovigilance database [6]. These statistical methods, aim at discriminating against occurrences in an SRS data set that are not attributable to the association between drug and adverse event [24], so that the potential nature of the association – whether causal or not – can be inferred tentatively. They are intended to give a measure of the degree to which the association between a drug and adverse event departs from what would be expected if the adverse event is independent of the drug [25, 46]; the methods are said to be 'disproportionality' [6] measures . We review some of the existing quantitative methods in this chapter and then propose new model(s) based on the Conway-Maxwell-Poisson distribution [43, 78] in a hierarchical framework. We begin by first describing a simplified version of an SRS database.

4.1 Simplified SRS Database

In terms of drug and adverse events, a Spontaneous Reporting System (SRS) database can be conceived as an I by J contingency table linking I drugs with J adverse events as shown in Table 4.1 [3, 19]. For any drug and adverse events pair, Table 4.1 could be reduced to Table 4.2 [3, 19].

Table 4.1: A cross-tabulation of drugs and adverse events.

	Adverse Event							
Drug		E_1	E_2	$\cdots$	E_j	$\cdots$	E_J	
	D_1	N_{11}	N_{12}	$\cdots$	N_{1j}	$\cdots$	N_{1J}	$N_{1.}$
	D_2	N_{21}	N_{22}	$\cdots$	N_{2j}	$\cdots$	N_{2J}	$N_{2.}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\ddots$	$\vdots$	$\vdots$
	D_i	N_{i1}	N_{i2}	$\ddots$	$\ddots$	$\ddots$	N_{iJ}	$N_{i.}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\ddots$	$\vdots$	$\vdots$
	D_I	N_{I1}	N_{I2}	$\cdots$	N_{Ij}	$\cdots$	N_{IJ}	$N_{I.}$
		$N_{.1}$	$N_{.2}$	$\cdots$	$N_{.j}$	$\cdots$	$N_{.J}$	$N_{..}$

Table 4.2: A cross-tabulation of drug i and adverse events j.

	Adverse Event			
Drug		E_j	$\sim E_j$	
	D_i	N_{ij}	$N_{i\bar{j}}$	$N_{i.}$
	$\sim D_i$	$N_{\bar{i}j}$	$N_{\bar{i}\bar{j}}$	$N_{\bar{i}.}$
		$N_{.j}$	$N_{.\bar{j}}$	$N_{..}$

D_i : The i^{th} drug

$\sim D_i$: Drugs other than the i^{th} one

E_j : The j^{th} adverse event

$\sim E_j$: Adverse events other than the j^{th} one

N_{ij} : Number of cases involving the i^{th} drug and the j^{th} adverse event

$N_{i\bar{j}}$: Number of cases involving i^{th}drug and adverse events other than the j^{th} one

$N_{\bar{i}j}$: Number of cases involving j^{th} adverse event and drugs other than the i^{th} one

$N_{\bar{i}\bar{j}}$: Number of cases not involving i^{th} drug nor the j^{th} adverse event

$N_{i.}$: Total number of cases involving the i^{th} drug

$N_{\bar{i}.}$: Total number of cases not involving the i^{th} drug

$N_{.j}$: Total number of cases involving the j^{th} adverse event

$N_{.\bar{j}}$: Total number of cases not involving the j^{th} adverse event

$N_{..}$: Total number of adverse event cases across all cells

4.2 Some Existing Methods

4.2.1 Relative Report Rate (RR)

Under the assumption of no association between drug i and adverse event j, the Relative Report Rate (RR) [19] given by

$$\begin{aligned} \mathrm{RR}_{ij} &= \frac{N_{ij}}{E_{ij}} \\ E_{ij} &= \frac{N_{i.}N_{.j}}{N_{..}} \end{aligned} \tag{4.1}$$

E_{ij} : Expected frequency of the ij^{th} cell

is expected to be roughly equal to 1 [19], so that values of RR > 1 could be thought of as interesting; requiring further scrutiny to determine whether indeed the relationship between the drug and the adverse event is causal.

In order to address problems of confounding due to covariates, N_{ij} and E_{ij} could be redefined as:

$$\begin{aligned} N_{ij} &= \sum_k N_{ijk} \\ E_{ij} &= \sum_k \frac{N_{i.k}N_{.jk}}{N_{..k}} \end{aligned}$$

where k is the number of strata resulting from potential confounding factors such as sex and age [19].

4.2.2 Proportional Reporting Ratio (PRR)

The Proportional Reporting Ratio (PRR) introduced by Evans et al. [25] relates the proportion of an adverse event in the database cited with a given drug to the proportion of the same event cited with all other drugs. It is defined as

$$\mathrm{PRR} = \frac{N_{ij}/N_{i.}}{N_{\bar{i}j}/N_{\bar{i}.}}.$$

Like RR, PRR is expected to have a value of 1 under the assumption of no association between drug and adverse event. However the authors use a different criterion to define a signal. A signal is deemed to be present when all the conditions: PRR ≥ 2, $N_{ij} \geq 3$ and $\chi^2 \geq 4$ are satisfied, where the χ^2 test of independence of drug and adverse event is based on Table 4.2. The condition $\chi^2 \geq 4$ is an approximation to the exact critical value ($\chi^2 = 3.841$) for a chi-square test of independence at 5% significance level with 1

degree of freedom; a PRR of at least 2 suggest the drug and adverse event combination is being reported at a rate that is at least twice as much as would be expected if the drug and adverse event are independent; and an observed count of at least three represents a trade-off between not permitting too many false positives due to the presence of drug and adverse event pairs with weak data support and not encouraging false negatives because the number of drug and adverse event pairs required for signal generation has been set too high.

4.2.3 Reporting Odds Ratio (ROR)

The Reporting Odds Ratio (ROR) is a method allied to PRR [24, 52]. It relates the odds of an adverse event cited with a given drug to the odds of the same event cited with all other drugs, which is defined by

$$\text{ROR} = \frac{N_{ij}/N_{i\bar{j}}}{N_{\bar{i}j}/N_{\bar{i}\bar{j}}}.$$

The above methods are prone to large numbers of false positive signals [19, 24]. This is the case because drug and adverse events with very small E_{ij} tend to have relatively higher $N_{\bar{i}\bar{j}}$ and $N_{\bar{i}.}$ leading to very large values of the above measures [24, 34], which do not correspond with the observed count.

To address the issue of uncertainty and reduce the level of false positives, two Bayesian methods, the Gamma Poisson Shrinker (GPS) [19] and the Bayesian Confidence Propagation Neural Network (BCPNN) [7], were developed. Their strength lies in the ability to shrink RR towards the null value of 1. The shrinkage tends to be more pronounced for drug and adverse event pairs for which the observed count or the expected count is very low [6], thereby reducing the numbers of false positives.

4.2.4 Gamma Poisson Shrinker (GPS)

The GPS, proposed by DuMouchel [19], assumes that each observed count N_{ij} is independently poisson distributed with parameter μ_{ij}. For drug i and adverse event j, the quantity of interest is λ_{ij}, where $\lambda_{ij} = \mu_{ij}/E_{ij}$. Thus

$$Pr(N_{ij} = n_{ij} \mid \lambda_{ij}, E_{ij}) = \frac{e^{-\lambda_{ij}E_{ij}}(\lambda_{ij}E_{ij})^{n_{ij}}}{n_{ij}!}.$$

Probability N_{ij} takes on value n_{ij}, after integrating out λ_{ij}, given $\lambda_{ij} \sim \text{Ga}(\alpha, \beta)$, is given by

$$Pr(N_{ij} = n_{ij} \mid \alpha, \beta, E_{ij}) = \left(1 + \frac{\beta}{E_{ij}}\right)^{-n_{ij}} \left(1 + \frac{E_{ij}}{\beta}\right)^{-\alpha} \frac{\Gamma(n_{ij} + \alpha)}{n_{ij}!\Gamma(\alpha)} \tag{4.2}$$

which is the negative binomial with parameter $1/\left(1 + \frac{\beta}{E_{ij}}\right)$.

The posterior distribution of λ_{ij}, when a value n_{ij} of N_{ij} has been observed, given a $\text{Ga}(\alpha, \beta)$ prior distribution on λ_{ij} is given by

$$\lambda_{ij} \mid \alpha, \beta, n_{ij} \sim \text{Ga}(\alpha + n_{ij}, \beta + E_{ij}) \tag{4.3}$$

If, however, instead of supposing a single gamma distribution on λ_{ij}, we consider a mixture of two gamma distributions [19], then

$$\lambda_{ij} \sim [\varrho \text{Ga}(\alpha_1, \beta_1) + (1 - \varrho)\text{Ga}(\alpha_2, \beta_2)]$$

where ϱ is the proportion of the λ_{ij}'s that come from $\text{Ga}(\alpha_1, \beta_1)$ and $1 - \varrho$ otherwise. It follows from (4.3) that

$$\lambda_{ij} \mid \alpha_1, \beta_1, \alpha_2, \beta_2, N_{ij} = n_{ij} \sim \omega_{ij}\text{Ga}(\alpha_1 + n_{ij}, \beta_1 + E_{ij}) + (1 - \omega_{ij})\text{Ga}(\alpha_2 + n_{ij}, \beta_2 + E_{ij})$$

where ω_{ij}, the posterior probability that λ_{ij} came from the first component of the mixture, is given by [19]

$$\omega_{ij} = \frac{\varrho Pr(N_{ij} = n_{ij} \mid \alpha_1, \beta_1, E_{ij})}{\varrho Pr(N_{ij} = n_{ij} \mid \alpha_1, \beta_1, E_{ij}) + (1 - \varrho)Pr(N_{ij} = n_{ij} \mid \alpha_2, \beta_2, E_{ij})}$$

and

$$Pr(N_{ij} = n_{ij} \mid \alpha_1, \beta_1, E_{ij})$$

is as given in (4.2).

The values of α_1, β_1, α_2, β_2 and ϱ are estimated by the method of maximum likelihood estimation, in which the product of the marginal densities given by (4.4) is maximized [19].

$$\begin{aligned} &Pr(N_{ij} = n_{ij} | \alpha_1, \beta_1, \alpha_2, \beta_2, \varrho, E_{ij}) = \\ &\varrho Pr(N_{ij} = n_{ij} \mid \alpha_1, \beta_1, E_{ij}) + (1 - \varrho)Pr(N_{ij} = n_{ij} \mid \alpha_2, \beta_2, E_{ij}) \end{aligned} \tag{4.4}$$

Drug and adverse event pairs are assessed on the basis of either the geometric mean of the posterior distribution of λ_{ij}, called the Empirical Bayes Geometric Mean (EBGM) [19], or on the basis of the 5% quantile of the posterior distribution of λ_{ij} (EB05) [21]. Szarfman et al. [82] proposed EB05 > 2 as a condition for the generation of a signal, which was arrived at through empirical studies.

4.2.5 Bayesian Confidence Propagation Neural Network (BCPNN)

The BCPNN model [7] uses what is referred to as the Information Component (IC) to generate signals. The IC for drug i and adverse event j is given by

$$\mathrm{IC}_{ij} = \log_2\left(\frac{\gamma_{ij}}{\eta_i \cdot \xi_j}\right).$$

The ICs are found via the Bayesian paradigm as follows [7]:

$$\begin{aligned} N_{ij} &\sim \mathrm{Bin}(N_{..}, \gamma_{ij}), \quad \text{where} \quad \gamma_{ij} \sim \mathrm{Beta}(\alpha_{ij}, \beta_{ij}) \\ N_{i.} &\sim \mathrm{Bin}(N_{..}, \eta_i), \qquad\qquad\quad \eta_i \sim \mathrm{Beta}(\alpha_{1i}, \beta_{1i}) \\ N_{.j} &\sim \mathrm{Bin}(N_{..}, \xi_j), \qquad\qquad\quad \xi_j \sim \mathrm{Beta}(\alpha_{2j}, \beta_{2j}) \end{aligned}$$

It follows that

$$\begin{aligned} \gamma_{ij}|N_{ij} &\sim \mathrm{Beta}(\alpha_{ij} + N_{ij}, \beta_{ij} + N_{..} - N_{ij}) \\ \eta_i|N_{i.} &\sim \mathrm{Beta}(\alpha_{1i} + N_{i.}, \beta_{1i} + N_{..} - N_{i.}) \\ \xi_j|N_{.j} &\sim \mathrm{Beta}(\alpha_{2j} + N_{.j}, \beta_{2j} + N_{..} - N_{.j}) \end{aligned}$$

Independence is assumed throughout.

The prior parameter values are chosen as follows [7]:

$$\alpha_{ij} = \alpha_{1i} = \beta_{1i} = \alpha_{2j} = \beta_{2j} = 1$$

which amount to specifying $U[0,1]$ prior on η_i and ξ_j, assuming equal probability owing to the lack of further information. The value of β_{ij} is specified in one of two ways; either

i. $\beta_{ij} = \frac{1}{\mathrm{E}[\eta_i|N_{i.}]\cdot\mathrm{E}[\xi_j|N_{.j}]}$

   thereby introducing information from the data through the prior of γ_{ij}. This choice takes cognisance of the assumption that if N_{ij} is zero, then drug and adverse event are independent and so IC_{ij} should tend to 0 as N_{ij} tends to 0 [7, 61],

   or

ii. $\beta_{ij} = \frac{1}{\mathrm{E}[\eta_i]\cdot\mathrm{E}[\xi_j]} - 1$

   which, given $\alpha_{1i} = \beta_{1i} = \alpha_{2j} = \beta_{2j} = 1$, is equal to 3 [7, 36].

Alternative prior specifications, taking into account the possible dependencies in γ_{ij}, η_i and ξ_j have been proposed by Norén et al. [58] as part of an extension of the model by Bate et al. [7].

A signal is deemed to be present for a drug and adverse event pair if the lower limit of the 95% confidence interval of the associated IC is greater than 0. The distribution of IC_{ij} is approximated by the normal distribution whose mean and variance is given by $\mathrm{E}[\mathrm{IC}_{ij}]$ and $\mathrm{Var}[\mathrm{IC}_{ij}]$ respectively. The values of $\mathrm{E}[\mathrm{IC}_{ij}]$ and $\mathrm{Var}[\mathrm{IC}_{ij}]$ are estimated by the delta method [7] or by a more accurate moments method when the sample size is small [36], using the posterior distributions of γ_{ij}, η_i and ξ_j.

4.2.6 Simple shrinkage Method

Norén et al. [59] have recently developed a 'shrinkage transformation' [59] of the observed-to-expected ratio (N_{ij}/E_{ij}) given by

$$\frac{N_{ij}+\alpha_1}{E_{ij}+\alpha_2} \tag{4.5}$$

where the presence of α_1 and α_2 causes shrinkage in the direction of α_1/α_2. To achieve shrinkage towards a given value k, one sets $\alpha_1 = k\alpha_2$. Thus choosing $k = 1$ causes shrinkage in the direction of the null situation of no association between drug and adverse event. The extent of shrinkage achieved, which decreases as the observed and expected counts increases, is determined by the choice of α_1 and α_2. The higher the values of α_1 and α_2, the more reliable the safeguard they offer against false signaling [59]. The ratio

$$\frac{N_{ij}+\alpha_1}{E_{ij}+\alpha_2}$$

is said to be analogous to the posterior mean of a parameter μ_{ij} where $N_{ij} \sim \mathrm{Pois}(\mu_{ij} \cdot E_{ij})$, $\mu_{ij} \sim \mathtt{Ga}(\alpha_1, \alpha_2)$ and $\mu_{ij}|\alpha_1, \alpha_2, N_{ij} \sim \mathtt{Ga}(N_{ij}+\alpha_1, E_{ij}+\alpha_2)$ [59].

The lower and upper bound of an approximate 95% 'credibility interval' estimate for

$$\log_2\left(\frac{N_{ij}+\alpha_1}{E_{ij}+\alpha_2}\right) \tag{4.6}$$

is respectively given by

$$\log_2\left(\frac{N_{ij}+\alpha_1}{E_{ij}+\alpha_2}\right) - 3.3\cdot(N_{ij}+\alpha_1)^{-1/2} - 2\cdot(N_{ij}+\alpha_1)^{-3/2}$$

and

$$\log_2\left(\frac{N_{ij}+\alpha_1}{E_{ij}+\alpha_2}\right) + 2.4\cdot(N_{ij}+\alpha_1)^{-1/2} - 0.5\cdot(N_{ij}+\alpha_1)^{-3/2}.$$

Norén et al. [59] report that (4.6) provides a very good approximation to the IC measure of the BCNPP model when $\alpha_1 = \alpha_2 = 1/2$ and the number of records is greater than or equal to 1000.

4.2.7 Confounding and Other Methods

The 'innocent bystander effect' and 'masking' [46] are some of the other sources of confounding that beset the use of SRS data for signal detection, apart from age, sex and time. The innocent bystander refers to a relatively innocuous drug for which a false signal is generated because it is often co-prescribed with the drug which is the cause of an ADR – its association with the culprit drug makes the innocent drug come under the suspicion of being the cause of the ADR, by virtue of its value on a disproportionality measure [6,19,46]. Masking is a phenomenon that results from the interrelationship between combinations of drugs and adverse events. The value of a disproportionality measure for a given drug and adverse event pair X is not independent of that of other pairs having the same adverse event or drug in combination X [6]. A large increase in the observed count (N_{ij}) for combination X will cause an increase in the expected counts of other drug and adverse event pairs having the same adverse event or drug in combination X. This may cause the value of a disproportionality measure for these other combinations to decrease below the threshold for signal generation. A sudden increase in the rate of reporting of a drug and adverse event pair resulting from regulatory action or media attention could create such a situation [6, 46]. It may be necessary to remove combination X at some stage of the analysis to ensure that no signals are obscured [6]. An adverse event which has a high observed count (N_{ij}) with several drugs will have a relatively large marginal total count ($N_{\cdot j}$) and the expected counts of all drugs that are reported with it will be large. This will make the value of a disproportionality measure for a new drug which is reported with the adverse event relatively small because of its relatively low observed count, which may result in the hiding of disproportionate reporting of the new drug and the adverse event [6].

A slightly different situation arises when the observed counts (N_{ij}) of the drugs reported with an adverse event are not significantly different but the marginal total counts ($N_{i\cdot}$) of the drugs differ substantially, then the drugs with lower marginal total counts may easily generate a signal than the ones with higher marginal total counts because of their lower expected counts [6].

Other approaches aimed at addressing some of the problems confronting 'quantitative signal detection' [6], beyond stratifying along the lines of sex, age and time, and hence the problem of false signaling have recently been explored. They include a method which employs the idea behind PRR and ROR by exploiting information from additional sources (which are not always available) supplementary to SRS data [34], an extension of the ex-

isting Bayesian methods to incorporate the concepts of false discovery rate (FDR) and false negative rate (FNR) [2], and the use of logistic regression [12, 46]. The latter method, which can uncover masking, involves regressing each adverse event on all the drugs, an exercise which involves estimating several thousand regression coefficients for each adverse event. Madigan et al. [46] report that the conventional logistic regression could be hampered by "lack of convergence, large estimated coefficient variances and poor predictive accuracy" [46], and so they opt for Bayesian logistic regression to address these issues using model (4.7) [12, 46]

$$Pr(y_i = +1|\beta, \mathrm{x}) = \psi(\beta^T \mathrm{x}) \tag{4.7}$$

where the value y_i indicates the inclusion or otherwise of an adverse event in the i^{th} report, x is a vector of binary variables including a leading 1 to accord with the intercept β_0 and whose values indicate the presence or otherwise of a given drug, and ψ is the logistic link function. The values of β are thought to be analogous to those of the logarithm of ROR [12]. Madigan et al. [46] present two possible prior specifications for β. Either

i. $\beta_j \sim \mathtt{N}(0, \tau), \qquad \tau > 0$

   or the hierarchical prior distribution given by

ii. $\beta_j \sim \mathtt{N}(0, \tau_j) \qquad$ where $\qquad \tau_j|\gamma \sim \mathrm{Exp}\left(\frac{\gamma}{2}\right), \qquad \gamma > 0$

   which amount to a double exponential or Laplace prior when τ_j has been integrated out [46].

Estimating the maximum a posteriori (MAP) value of β is likened to ridge logistic regression in the case of (i) or Lasso logistic regression in the case of (ii) [12, 46]. Efficient implementation of the method is said to be computationally involving and requires heavy parallel computing [12, 46]. While this method holds promise for dealing with problems that come with confounding, its regular use could be hampered by peculiar computational demands [12] and there are drug safety problems that it is unable to detect as quickly as the other methods. However, it can also detect other drug safety problems quicker than the other methods [46].

Indeed the various quantitative methods have their limitations. The limitations are partly due to the very character of SRS data [6]. The need to develop the signal detection process to the highest level dictates that we continue to study the attributes of the existing quantitative signal detection methods with the view to having a deeper understanding of

their abilities and inabilities [6], and to see how and when they could be used in complementary roles. By the same token we must continue to explore other quantitative signaling methods. We, therefore, propose another Bayesian shrinkage model as described in the next section.

4.3 Proposed Model(s)

4.3.1 Background of Model(s)

The COM-Poisson (CMP) distribution is a generalization of the Poisson distribution developed by Conway and Maxwell [43,78]. We have the special case of the Poisson distribution when $\nu = 1$. The probability mass function of the CMP distribution is defined by

$$\begin{aligned} Pr(N = n) &= \frac{\tau^n}{(n!)^\nu} \cdot \frac{1}{z(\tau, \nu)} \\ z(\tau, \nu) &= \sum_{s=0}^{\infty} \frac{\tau^s}{(s!)^\nu} \end{aligned}$$

where the parameters are τ and ν (dispersion parameter), with $\nu > 1$ indicating under-dispersion and $\nu < 1$ over-dispersion [41,78].

Approximate expressions for the first and second moments, as given in Shmueli et al. [78] and Guikema and Goffelt [41], are respectively

$$\begin{aligned} \mathrm{E}[N] &\approx \tau^{1/\nu} + \frac{1}{2\nu} - \frac{1}{2} \\ \mathrm{Var}[N] &\approx \frac{1}{\nu}\tau^{1/\nu} . \end{aligned}$$

A reparameterization of the above density, by Guikema and Goffelt [41], using $\rho = \tau^{1/\nu}$ yields

$$Pr(N = n) = \left(\frac{\rho^n}{n!}\right)^\nu \cdot \frac{1}{z(\rho^\nu, \nu)} \tag{4.8}$$

with the corresponding moments being

$$\mathrm{E}[N] \approx \rho + \frac{1}{2\nu} - \frac{1}{2} \tag{4.9}$$

$$\mathrm{Var}[N] \approx \frac{\rho}{\nu} . \tag{4.10}$$

The parameter $\rho = \tau^{1/\nu}$ is the mode of the CMP distribution [41]. Parameterization (4.8) is preferred because it provides a clear centering parameter in ρ. It will be used in the rest of the thesis and will be denoted by $\mathtt{CMP}(\rho, \nu)$.

4.3.2 Models C-G and P-G

The modeling proceeds as follows: each cell count N_{ij} is assumed to be distributed as

$$N_{ij} \mid \phi_{ij}, \nu \sim \texttt{CMP}(\phi_{ij} E_{ij}, \nu),\ i = 1, 2, \ldots, I;\ j = 1, 2, \ldots, J \ldots \text{ independently.} \quad (4.11)$$

where ϕ_{ij} is a scale factor, the ratio between the mode of N_{ij} and the expected count E_{ij} as given in (4.1). Values of ϕ_{ij} larger than one correspond to a distribution of N_{ij} with more mass assigned to values larger than E_{ij}.

The ϕ_{ij}'s are assumed to come from a common gamma distribution with parameters α and β which are, in turn, assumed to be exponentially distributed with respective parameters θ_α and θ_β. The parameter ν which is also assumed to be exponentially distributed with parameter θ_ν is thought of as a general dispersion parameter. Thus we have

$$\begin{aligned} \phi_{ij} \mid \alpha, \beta &\sim \texttt{Ga}(\alpha, \beta)\ \ i = 1, 2, \ldots, I;\ \ j = 1, 2, \ldots, J\ \ \ldots \text{ i.i.d.} \quad (4.12) \\ \nu \mid \theta_\nu &\sim \texttt{Exp}(\theta_\nu) \\ \alpha \mid \theta_\alpha &\sim \texttt{Exp}(\theta_\alpha) \\ \beta \mid \theta_\beta &\sim \texttt{Exp}(\theta_\beta) \end{aligned}$$

The likelihood is therefore given by

$$f(N \mid \phi, \nu, E) = \prod_{ij} \frac{(\phi_{ij} E_{ij})^{\nu N_{ij}}}{(N_{ij}!)^\nu} \cdot \frac{1}{z\left[(\phi_{ij} E_{ij})^\nu, \nu\right]} \quad (4.13)$$

where

$$\begin{aligned} N &= (N_{ij},\ \ i = 1, 2, \ldots, I,\ \ j = 1, 2, \ldots, J) \\ E &= (E_{ij},\ \ i = 1, 2, \ldots, I,\ \ j = 1, 2, \ldots, J) \\ \phi &= (\phi_{ij},\ \ i = 1, 2, \ldots, I,\ \ j = 1, 2, \ldots, J) \end{aligned}$$

and the prior is given by

$$h(\phi, \nu, \alpha, \beta) = \prod_{ij} \frac{\beta^\alpha}{\Gamma(\alpha)} \phi_{ij}^{\alpha-1} e^{-\beta\phi_{ij}} \times \theta_\nu e^{-\theta_\nu \nu} \times \theta_\alpha e^{-\theta_\alpha \alpha} \times \theta_\beta e^{-\theta_\beta \beta} \quad (4.14)$$

Thus the joint density of all the random quantities in the model, conditional on the E_{ij}s, which are assumed constant, is

$$\begin{aligned} \pi(\phi, \nu, \alpha, \beta, N \mid E) &= \prod_{ij} \frac{(\phi_{ij} E_{ij})^{\nu N_{ij}}}{(N_{ij}!)^\nu} \cdot \frac{1}{z\left[(\phi_{ij} E_{ij})^\nu, \nu\right]} \times \prod_{ij} \frac{\beta^\alpha}{\Gamma(\alpha)} \phi_{ij}^{\alpha-1} e^{-\beta\phi_{ij}} \\ &\quad \times \theta_\nu e^{-\theta_\nu \nu} \times \theta_\alpha e^{-\theta_\alpha \alpha} \times \theta_\beta e^{-\theta_\beta \beta} \quad (4.15) \\ &\propto \frac{\beta^{IJ\alpha} e^{-\beta(\sum_{ij} \phi_{ij} + \theta_\beta) - \nu(\sum_{ij} \log N_{ij}! + \theta_\nu) - \theta_\alpha \alpha}}{\left[\Gamma(\alpha)\right]^{IJ}} \prod_{ij} \frac{E_{ij}^{\nu N_{ij}} \phi_{ij}^{\nu N_{ij} + \alpha - 1}}{z\left[(\phi_{ij} E_{ij})^\nu, \nu\right]} \end{aligned}$$

It follows that the full conditional of β is given by

$$\pi(\beta|\cdot) \quad \propto \quad \beta^{IJ\alpha} e^{-\beta(\sum_{ij}\phi_{ij}+\theta_\beta)} \tag{4.16}$$

where $|\cdot$ means given all other variables in the model. Thus

$$\beta \mid \cdot \sim \texttt{Ga}\left([IJ\alpha+1], \left[\sum_{ij}\phi_{ij}+\theta_\beta\right]\right)$$

When $\nu = 1$, the likelihood 4.13 reduces to the Poisson likelihood, given by

$$f(N \mid \phi, E) = \prod_{ij} \frac{e^{-\phi_{ij}E_{ij}}(\phi_{ij}E_{ij})^{N_{ij}}}{N_{ij}!} \tag{4.17}$$

with the corresponding joint density of the random quantities being

$$\pi(\phi, \alpha, \beta, N \mid E) \propto \frac{\beta^{IJ\alpha} e^{-\beta(\sum_{ij}\phi_{ij}+\theta_\beta)-\sum_{ij}\log N_{ij}!-\theta_\alpha\alpha-\sum_{ij}\phi_{ij}E_{ij}}}{[\Gamma(\alpha)]^{IJ}} \prod_{ij} E_{ij}^{N_{ij}} \phi_{ij}^{N_{ij}+\alpha-1} \tag{4.18}$$

The full conditional for β remains as given in (4.16).

We designate the model based on (4.11) and (4.12) as C-G and that based on the Poisson distribution and (4.12) as P-G . In both cases the full conditionals of the parameters, apart from β, are not of standard form, so values of α, β, ν, and ϕ may be generated from a Gibbs and Metropolis-Hastings scheme as outlined below.

For C-G, using (4.15) and (4.16) we have:

1. Set initial values: α_0, β_0, ν_0 and ϕ_0.

2. For $(t \; in \; 1 : niter)$ {
   Generate $\beta^{(t)}$ from $\texttt{Ga}\left([IJ\alpha+1], [\sum_{ij}\phi_{ij}+\theta_\beta]\right)$
   Update $\alpha^{(t)}$ with a Metropolis-Hastings move.
   Update $\nu^{(t)}$ with a Metropolis-Hastings move.
   For $(i \; in \; 1 : I)$ and $(j \; in \; 1 : J)$ {
   Update $\phi_{ij}^{(t)}$ with a Metropolis-Hastings move.
   }
   }

Model P-G does not involve ν so the parameter updates, using (4.16) and (4.18), could be done as follows:

1. Set initial values: α_0, β_0 and ϕ_0.

2. For $(t\ in\ 1:niter)$ {
   Generate $\beta^{(t)}$ from $\texttt{Ga}\left([IJ\alpha+1],[\sum_{ij}\phi_{ij}+\theta_\beta]\right)$
   Update $\alpha^{(t)}$ with a Metropolis-Hastings move.
   For $(i\ in\ 1:I)$ and $(j\ in\ 1:J)$ {
   Update $\phi_{ij}^{(t)}$ with a Metropolis-Hastings move.
   }
   }

Updating the Parameters

This section describes how the parameter values were simulated.

The parameter β was updated using a Gibbs move, that is, the next value $\beta^{(t+1)}$ of β was generated from the Gamma distribution using the current values $\alpha^{(t)}$ and $\phi^{(t)}$ of α and ϕ respectively, as outlined above.

Values of α were generated by a random walk Metropolis Hastings (MH) move [14, 33, 44]. The candidate value $\tilde{\alpha}$ of α was proposed by

$$\tilde{\alpha} = \alpha^* + \delta, \qquad \delta \sim g$$

where α^* is the current state of α and g is a symmetrical distribution made up of three uniform distributions which have been superimposed such that they share a common mean of zero, as shown in the Figure 4.1. Distribution g is defined by

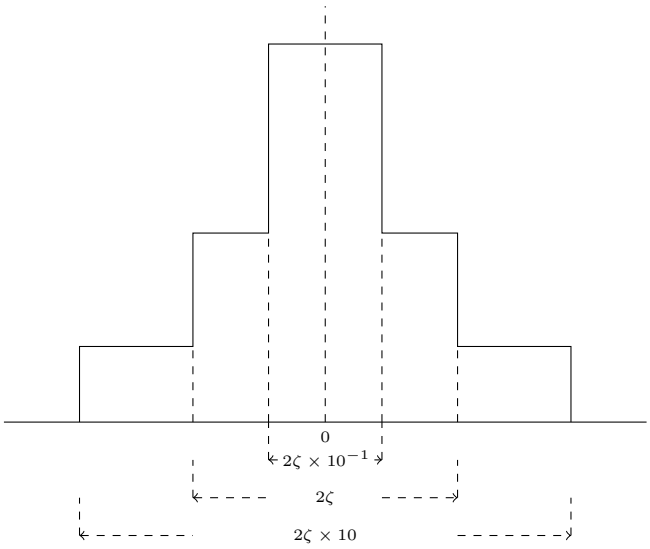


Figure 4.1: Density of the proposal distribution for α and ν (not to scale).

$$\delta = \begin{cases} \text{v} \times 10^{-1} & \text{if } \text{w} < \frac{1}{3} \\ \text{v} \times 10 & \text{if } \text{w} > \frac{2}{3} \\ \text{v} & \text{Otherwise} \end{cases}$$

where

$$\text{v} \sim U[-\zeta, \zeta] \quad \text{and} \quad \text{w} \sim U[0,1]$$

The need to ensure that the proposal steps are of the right magnitude (not always too small or too large) so that there is good mixing is the motivation for using this approach.

It follows from (3.6) that the acceptance probability is

$$a(\alpha^*, \tilde{\alpha}) = \min\left\{1, \frac{\pi(\tilde{\alpha})}{\pi(\alpha^*)}\right\}$$

and the updating scheme [14, 33, 44] is given by:

1. Given the current value $\alpha^{(t)} = \alpha^*$, generate the candidate value $\tilde{\alpha}$ as outlined above and $u \sim U[0,1]$.
2. Set $\alpha^{(t+1)} = \tilde{\alpha}$ if $u \leq a(\alpha^*, \tilde{\alpha})$, else set $\alpha^{(t+1)} = \alpha^{(t)}$

Values of ν were generated in the same way as in the case of α.

The technique used to generate values of ϕ is different from that of α and ν; the candidate value $\tilde{\phi}_{ij}$ of ϕ_{ij} was proposed by

$$\begin{aligned} \tilde{\phi}_{ij} &= \phi_{ij} \cdot e^u, \quad u \sim U[-\epsilon, \epsilon] \\ \implies u &= \log \tilde{\phi}_{ij} - \log \phi_{ij} \end{aligned}$$

It follows that

$$\frac{du}{d\tilde{\phi}_{ij}} = \frac{1}{\tilde{\phi}_{ij}}$$

and

$$\begin{aligned} q(\tilde{\phi}_{ij}|\phi_{ij}) &= f(u) \cdot \left|\frac{du}{d\tilde{\phi}_{ij}}\right| \\ &= \frac{1}{2\epsilon} \cdot \frac{1}{\tilde{\phi}_{ij}} \end{aligned}$$

Similarly

$$q(\phi_{ij}|\tilde{\phi}_{ij}) = \frac{1}{2\epsilon} \cdot \frac{1}{\phi_{ij}}$$

and

$$\begin{aligned}\frac{q(\phi_{ij}|\tilde{\phi}_{ij})}{q(\tilde{\phi}_{ij}|\phi_{ij})} &= \frac{\tilde{\phi}_{ij}}{\phi_{ij}} \\ &= e^u\end{aligned}$$

It follows from (3.5) that, given the current value ϕ_{ij}^*, the acceptance probability is given by

$$a({\phi_{ij}}^*, \tilde{\phi}_{ij}) = \min\left\{1, \frac{\pi(\tilde{\phi}_{ij})}{\pi({\phi_{ij}}^*)} \cdot e^u\right\}$$

The updating scheme follows the same steps as in the case of α. This updating technique was used to curtail autocorrelation in the iterates.

The MCMC samplers described above were implemented in Fortran and the samples dumped to files. The R statistical software was then used to make plots and compute posterior summaries.

4.3.3 Models C-IG and P-IG

One is expected to investigate how other prior specifications impact ϕ and ν a posteriori. For instance, the Inverse Gamma distribution is known to have longer tails than the Gamma distribution and could accommodate values of ϕ that are large much better than the Gamma distribution. We could, therefore, look at the alternative situation where ϕs are rather drawn from a common Inverse Gamma (instead of Gamma) distribution with parameters φ and ψ, where the φ and ψ are in turn exponentially distributed with respective parameters θ_φ and θ_ψ. The parameter ν is also exponentially distributed with parameter θ_ν. The role assigned to the Exponential distribution is not exclusive, the Inverse Exponential distribution could be used. Indeed the potential of the Inverse Exponential distribution was explored and preliminary results showed it has similar effect as the Exponential distribution. Owing primarily to the constraint of time the models described below were restricted to the use of the Exponential distribution. Thus, we look at the situation where

$$\begin{aligned}\phi_{ij} \mid \varphi, \psi &\sim \texttt{Inv-Ga}(\varphi, \psi) \;\; i = 1, 2, \ldots, I; \;\; j = 1, 2, \ldots, J \;\; \ldots \; i.i.d. \qquad (4.19)\\ \nu \mid \theta_\nu &\sim \texttt{Exp}(\theta_\nu) \\ \varphi \mid \theta_\varphi &\sim \texttt{Exp}(\theta_\varphi) \\ \psi \mid \theta_\psi &\sim \texttt{Exp}(\theta_\psi)\end{aligned}$$

In which case joint densities (4.15) and (4.18) become respectively:

$$\pi(\phi,\nu,\varphi,\psi,N \mid E) \propto \frac{\psi^{IJ\varphi}e^{-\psi(\sum_{ij}1/\phi_{ij}+\theta_\psi)-\nu(\sum_{ij}\log N_{ij}!+\theta_\nu)-\theta_\varphi\varphi}}{[\Gamma(\varphi)]^{IJ}}\prod_{ij}\frac{E_{ij}^{\nu N_{ij}}\phi_{ij}^{\nu N_{ij}-\varphi-1}}{z\left[(\phi_{ij}E_{ij})^\nu,\nu\right]} \tag{4.20}$$

and

$$\pi(\phi,\varphi,\psi,N \mid E) \propto \frac{\psi^{IJ\varphi}e^{-\psi(\sum_{ij}1/\phi_{ij}+\theta_\psi)-\sum_{ij}\log N_{ij}!-\theta_\varphi\varphi-\sum_{ij}\phi_{ij}E_{ij}}}{[\Gamma(\varphi)]^{IJ}}\prod_{ij}E_{ij}^{N_{ij}}\phi_{ij}^{N_{ij}-\varphi-1} \tag{4.21}$$

The models represented by joint densities (4.20) and (4.21) will be referred to as C-IG and P-IG respectively. In each case, the full conditional of ψ is given by:

$$\pi(\psi|\cdot) \quad \propto \quad \psi^{IJ\varphi}e^{-\psi(\sum_{ij}1/\phi_{ij}+\theta_\psi)}$$

and

$$\psi \mid \cdot \sim \texttt{Ga}\left([IJ\varphi+1],\left[\sum_{ij}1/\phi_{ij}+\theta_\psi\right]\right)$$

The parameters in C-IG and P-IG were simulated by the same techniques as used for parameters in C-G and P-G described in Section 4.3.2.

Chapter 5

Application of Proposed Model(s) to FDA SRS Data

The results of analysis illustrating the proposed models outlined in Section 4.3 with data from the Food and Drugs Administration's SRS database are presented in this chapter. We reiterate that the data set (referred to as the second data set in Chapter 2) used in the analysis presented in this chapter and described in Section 5.1 is different from the data set used in the analysis presented in Chapter 2. Ideally the data set used in the analysis presented in Chapter 2 should have been used throughout the investigation in order to present a holistic view of the data concerning adverse events associated with the use of medications. However it became necessary to use a different data set for the analysis presented in this chapter because the data used in the analysis presented in Chapter 2 requires careful clean-up. As noted in Chapter 2, the clean-up involves correcting misspellings of names of drugs and adverse events, finding all descriptions in the data that point to the same adverse event and recoding them into the Medical Dictionary for Regulatory Activities (MedDRA) Preferred Term (PT) [83] and finding all the different names in the data that indicate the same drug (active ingredients) and recoding them into a commonly accepted drug name. The clean-up is an expensive work both in terms of the time and the expertise required to do it [20], which the current research endeavour can not afford. A description of the original data set along with that of two other derivative data sets used in this chapter is given in the next section.

5.1 Data

The original data, which cover the United States and span the thirty-year period from 1969 to 1998, has 1398 drugs and 952 events. The observed count of each drug and adverse event pair was collated from 18 strata arising from six five-year periods and three sex categories – male, female and unspecified sex. The data is the same as the one used in the analysis reported in the article "Bayesian data mining in large frequency tables, with an application to FDA spontaneous reporting system" [19], except it does not include the expected count of drug and adverse event pairs with zero observed count. The data can be found at this web address: ftp://ftp.research.att.com/dist/gps/. The data has 385734 drug and adverse event pairs (cells); the number of pairs would have been 1330896 had the data included drug and adverse event pairs with zero observed count. We designate this data set as Data 2. Two other data sets, Data 1 and Data 3, were obtained from Data 2. Data 1 was derived from Data 2 by imputing the expected count of drug and adverse event pairs whose observed count is zero with values computed with (4.1), the equation for calculating the expected count (E_{ij}) for the unstratified situation. There are 1330896 drug and adverse event pairs (cells) in this data. Data 3 includes only pairs (132037 of them) whose observed counts are greater than or equal to five. It is worthy of note that the overall and marginal totals that determine the expected counts in Data 2 are the same as those that determine the expected counts in Data 3. In other words, Data 3 was derived from Data 2 by simply removing drug and adverse event pairs (cells) with observed counts less than five from Data 2 – a drug and adverse event pair present in Data 3 has the expected count to be the same as it has in Data 2. For the sake of clarity we restate the three data sets as follows:

Data 1: Data including all counts.

Data 2: Data including only counts $N \geq 1$.

Data 3: Data including only counts $N \geq 5$.

The rationale behind the data segmentation was to see how the models perform with these different data sets. However, it must be emphasized that Data 2 and Data 3 are essentially truncated versions of Data 1 but the models fitted to them are not for truncated data, and this could result in biased estimates in the case of Data 2 and Data 3. Also one is not oblivious of the fact that the ϕs are not independent a posteriori, and their estimates

could be affected by the exclusion of cells with zero counts or counts less than five.

Of course there is the issue of what should inform the choice of a number as the minimum threshold for the inclusion of a drug and adverse event pair in a quantitative analysis [19]. One could use an arbitrary choice, as has been done here, as a starting point and hope that the analysis will throw up something that could be used as a basis for setting the minimum threshold in a formal way.

5.2 Results of Analysis

This section presents the results of the analysis carried out with the models described in Section 4.3 with some discussion. We do a further discussion of the results in Chapter 6. The information that follows covers eight model-data combinations. Again, for purposes of clarity we restate the various models considered in the analysis:

C-G : CMP/Gamma model

P-G : Poisson/Gamma model

C-IG : CMP/Inverse Gamma model

P-IG : Poisson/Inverse Gamma model

5.2.1 Performance of algorithm

Three chains were run for each model-data combination and a sample of 1000 iterates were drawn per chain at a thinning interval of 40, after a burn-in of 50,000 in the case of Data 1 and 20,000 in the case of Data 2 and Data 3. Thus a total of 90,000 iterations were carried out for runs involving Data 1 and 60,000 for runs involving Data 2 and Data 3. A shorter burn-in was used for Data 2 and Data 3 as examination of the ACF and trace plots suggests convergence occurs faster compared to Data 1. The values of ζ and ϵ (see Section 4.3.2) used in the M-H moves for α, φ, ν and ϕ are shown in Table 5.1. From the results of C-G and C-IG (see Tables 5.4b and 5.4c), it makes no sense to set $\nu = 1$ for Data 2 and Data 3 and so P-G and P-IG were deemed unsuitable for these two data sets.

The ACF and trace plots of α, β, ν and the logarithm of the target distribution for C-G model and Data 1 are shown in Figures 5.1 and 5.2 respectively. The corresponding figures for the remaining seven model-data pairs, along with other diagnostic graphs are shown in Appendix A through Appendix D. The sharp fall in the ACF values of α, β,

Table 5.1: Values of ζ and ϵ used in the runs.

Model	Data 1			Data 2			Data 3		
	ζ		ϵ	ζ		ϵ	ζ		ϵ
	α, φ	ν	ϕ	α, φ	ν	ϕ	α, φ	ν	ϕ
C-G/C-IG	0.01	0.01	1.0	0.01	0.01	1.0	0.01	0.005	1.0
P-G/P-IG	0.01	0.01	1.0	–	–	–	–	–	–

ν and the logarithm of the target distribution is an indication of a good mixing Markov chain. There are three chains in each trace plot, coloured red, blue and black, which are overlapping quite well (with some oscillating pattern though, particularly, in the case of α, which is resulting from a relatively large lag 1 autocorrelation).

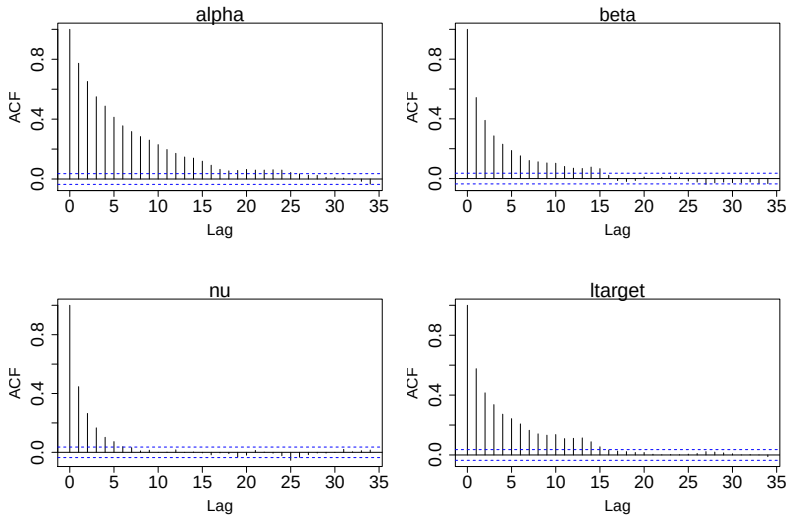


Figure 5.1: Acf plots of α, β, ν and the logarithm of the target distribution using C-G and Data 1.

The Gelman-Rubin $\hat{R}$ [33] values for α, β, φ, ψ, ν and the logarithm of the target distribution for various model-data combinations were computed. The largest of the $\hat{R}$ values across all model-data combinations considered is 1.005. The corresponding values for the ϕs of 100 randomly selected drug and adverse event pairs for each of the model-data pairs are also all approximately equal to 1. As pointed out by Gelman et al. [33] convergence of the algorithm can be assumed if $\hat{R}$ is close 1. Values of $\hat{R}$ below 1.1 are said to be acceptable.

Based on $\hat{R}$ values, the sharp fall in the ACF values and the stable chains of the trace

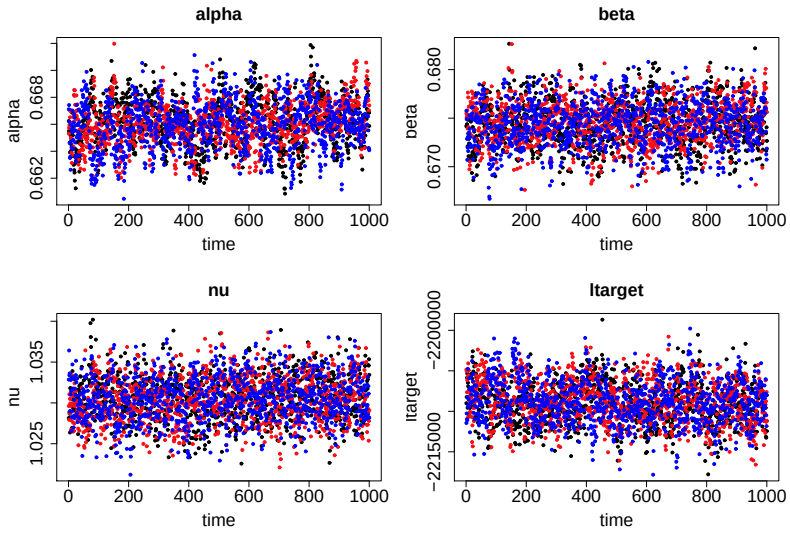


Figure 5.2: Trace plots of α, β and the logarithm of the target distribution for three chains using C-G and Data 1.

plots, the algorithm was assumed to have converged. In terms of convergence, the results of the other seven model-data pairs are similar to that of C-G and Data 1 described above.

Table 5.2 shows the acceptance rate of the candidate values of the parameters. The first of the two values shown against ϕ is the average acceptance rate for ϕ for the given model-data pair and the second, which is in square brackets, is the associated standard deviation. There appear to be no major issues with the values.

The times it takes to run the models with the three data sets for the specified total number of iterations and to generate replicate data are shown in Table 5.3. It may come as a surprise that it requires comparatively more time to run C-G and C-IG for 60,000 iterations (including burn-in) with Data 2 and Data 3, which are about a one-third and one-tenth respectively of Data 1, than to run them for 90,000 iterations with Data 1. The situation is explained by the fact that the time required to evaluate the function $z(\rho^\nu, \nu)$ in C-G and C-IG increases as the value of ν gets smaller and smaller below one. The value of ν for C-G and C-IG with Data 2 and Data 3 are less than 0.3 (see Tables 5.4b and 5.4c), hence the situation. Indeed the presence of $z(\rho^\nu, \nu)$ in C-G and C-IG accounts for the dramatic difference between the times required to run C-G/C-IG and P-G/P-IG, which is the principal motivation for running the MCMC samplers in Fortran. However, the time used in the runs is also a function of the magnitude of the proposal steps. For instance changing ζ from 0.005 to 0.01 (with all other things fixed) in the case of ν for

Table 5.2: Acceptance rate (%) of candidate values of the parameters.

Model	Parameter	Acceptance Rate		
		Data 1	Data 2	Data 3
C-G	α	23.7	37.2	47.2
	ν	36.9	22.0	20.9
	ϕ	78.5[14.7]	73.2[11.4]	72.0[8.7]
P-G	α	24.2	–	–
	ϕ	78.3[14.5]	–	–
C-IG	φ	31.5	42.3	51.9
	ν	37.2	25.3	27.2
	ϕ	70.5[12.0]	65.7[10.0]	64.1[8.3]
P-IG	φ	32.0	–	–
	ϕ	70.5[11.5]	–	–

Data 3 decreases the run time by seven days. The change in value of ζ from 0.005 to 0.01 increases the magnitude of the proposal step, δ (see Section 4.3.2) for ν, which results in the Markov chain traversing the parameter space faster and hence the decrease in time used. The acceptance rate for ν is lower at the bigger value of ζ than at the smaller value. For instance the acceptance rate for ν in the case of C-G and Data 3 are 11.6% and 20.9% for $\zeta = 0.01$ and $\zeta = 0.005$ respectively. However the posterior summaries corresponding to the two values of ζ are roughly the same.

Table 5.3: Running times of the models.

Model	Unit of time	Time Required		
		Data 1	Data 2	Data 3
C-G / C-IG	Days	16	14	24
P-G / P-IG	Hours	21	–	–

5.2.2 Parameter Estimates

Estimates for α, β, φ, ψ and ν for the various model-data combinations are shown in Tables 5.4a through 5.4c. A close scrutiny of the values in Table 5.4a shows that, for the common parameters, the mean values obtained via C-G and P-G agree to at least the first decimal place for Data 1. This observation also holds for values of the parameters common to C-IG and P-IG. The corresponding standard deviations are the same to the third decimal place.

The Poisson distribution is a special case of the CMP distribution, where $\nu = 1$, and the prior distributions for ϕ in C-G is the same as in P-G. Again the prior distributions for ϕ in C-IG is the same as in P-IG. Models P-G and P-IG are therefore special cases of C-G and C-IG respectively. The level of agreement between the values of the parameters common to C-G and P-G and also C-IG and P-IG, when Data 1 is used, is therefore not unlikely given that the posterior mean of ν is close to 1 (Table 5.4a) for this data.

Table 5.4a: Parameter estimates for the various models using Data 1.

Data 1					
Model	Parameter	Mean	Standard	Cred. Inter.	
		Value	Deviation	2.5%	97.5%
C-G	α	0.666	0.002	0.663	0.669
	β	0.674	0.002	0.670	0.679
	ν	1.031	0.003	1.025	1.036
P-G	α	0.663	0.002	0.660	0.667
	β	0.682	0.002	0.678	0.687
C-IG	φ	1.544	0.004	1.537	1.552
	ψ	0.627	0.002	0.623	0.632
	ν	1.087	0.003	1.081	1.092
P-IG	φ	1.569	0.004	1.562	1.576
	ψ	0.618	0.002	0.614	0.623

Table 5.4b: Parameter estimates for C-G and C-IG using Data 2.

Data 2					
Model	Parameter	Mean	Standard	Cred. Inter.	
		Value	Deviation	2.5%	97.5%
C-G	α	1.223	0.006	1.212	1.234
	β	1.386	0.007	1.373	1.400
	ν	0.205	0.001	0.204	0.207
C-IG	φ	2.165	0.009	2.148	2.183
	ψ	1.115	0.007	1.102	1.129
	ν	0.259	0.001	0.257	0.261

Table 5.4c: Parameter estimates for C-G and C-IG using Data 3.

Data 3					
Model	Parameter	Mean Value	Standard Deviation	Cred. Inter. 2.5%	Cred. Inter. 97.5%
C-G	α	1.518	0.013	1.493	1.544
	β	1.832	0.017	1.799	1.868
	ν	0.054	0.0004	0.053	0.055
C-IG	φ	2.442	0.017	2.407	2.475
	ψ	1.436	0.014	1.407	1.464
	ν	0.083	0.001	0.082	0.084

5.3 Diagnostics

5.3.1 Validation of the distribution of ϕ

For the C-G model, Figures 5.3 and C.11(b) and C.16(b) of Appendix C suggest a fit between the distribution of the ϕs and the Gamma distribution corresponding to the MCMC estimates of the posterior means of α and β, for all three data sets. Figure D.1 also suggests a fit between the distribution of the ϕs and the Gamma distribution corresponding to the MCMC estimates of the posterior means of α and β for P-G and Data 1. As in the case of C-G and P-G, the distribution of the ϕs generated via C-IG and P-IG fit the Inverse Gamma distribution corresponding to the MCMC estimates of the posterior means of φ and ψ generated from these models (see Figures E.1, E.11(b), E.16(b) and F.1). A misfit between the distribution of the ϕs and the Gamma distribution corresponding to the MCMC estimates of the posterior means of α and β would have meant that something is amiss with the model (C-G or P-G); for instance the model is not generating ϕs from the required Gamma distribution because the model is not suitable for the data or something has gone wrong with the implementation of the model. The same reasoning goes for the relationship between the distribution of the ϕs and the Inverse Gamma distribution corresponding to the MCMC estimates of the posterior means of φ and ψ, for C-IG and P-IG.

5.3.2 Posterior Predictive Check

As a posterior predictive check, Bayesian p-value scatter plots of the logarithm of X_2 (determined from replicate counts) against X_1 (determined from observed counts) using

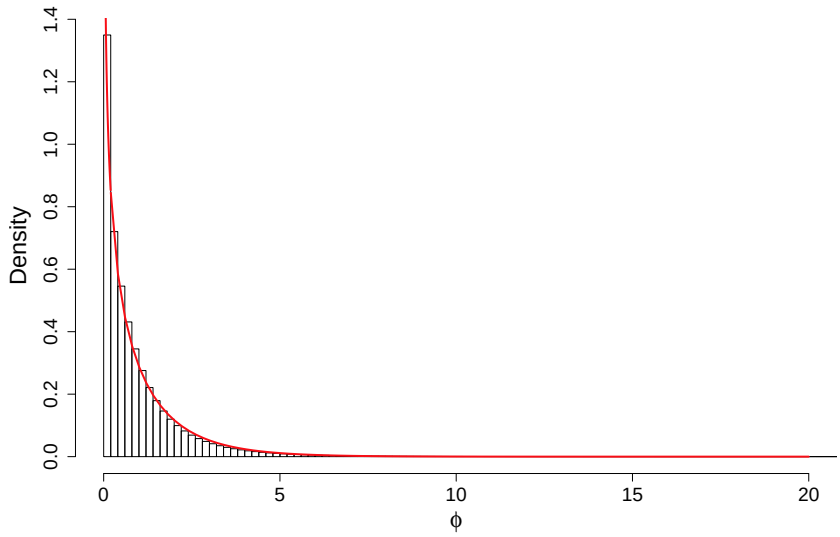


Figure 5.3: Histogram of ϕs for C-G and Data 1. The superimposed density curve (in red) represents the Gamma distribution corresponding to the MCMC estimates of the posterior means of α and β.

Equations 5.1 and 5.2 [33] were produced for all model-data combinations considered. The quantity $\theta^{(t)}$ is the t^{th} MCMC draw for θ, where θ is as defined in (5.3), and $N_{ij}^{(t)}$ is the t^{th} replicate count.

$$X_1^{(t)} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(N_{ij} - \mathbb{E}\left[N_{ij}^{(t)}|\theta^{(t)}\right]\right)^2}{\text{Var}\left[N_{ij}^{(t)}|\theta^{(t)}\right]} \tag{5.1}$$

$$X_2^{(t)} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(N_{ij}^{(t)} - \mathbb{E}\left[N_{ij}^{(t)}|\theta^{(t)}\right]\right)^2}{\text{Var}\left[N_{ij}^{(t)}|\theta^{(t)}\right]} \tag{5.2}$$

where

$$\theta = \begin{cases} (\phi, \nu) & \text{if model is C-G or C-IG} \\ \phi & \text{if model is P-G or P-IG} \end{cases} \tag{5.3}$$

The $N_{ij}^{(t)}$s were generated via the simulation method of inversion [33] using $\theta^{(t)}$. The p-value scatter plot for C-G and Data 1 based on X_1 and X_2 is shown in Figure 5.4 (a); those for other model-data pairs are shown in the appendices.

The models were also assessed on the basis of what is referred to as the deviance [33, p. 175], defined as

$$D(\mathbf{x}, \theta) = -2\ell(\mathbf{x}|\theta) \tag{5.4}$$

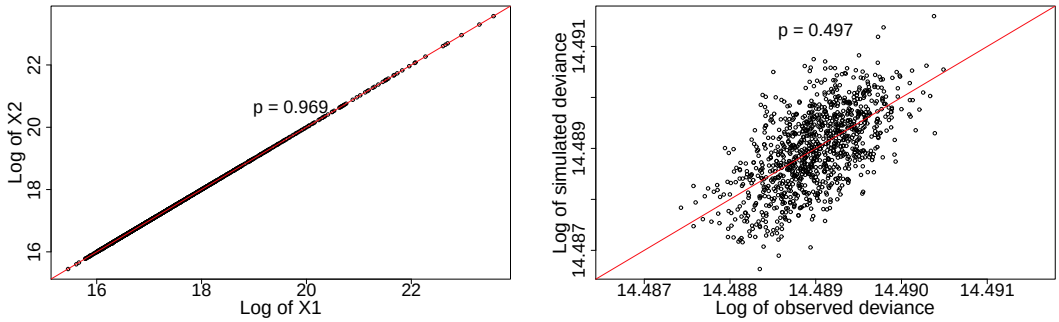


Figure 5.4: Bayesian p-value scatter plots for C-G and Data 1. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.

where $\ell(\mathbf{x}|\theta)$ is the logarithm of the likelihood, $\mathbf{x}$ is the data (in our case the observed data N or the simulated data $N^{(t)}$) and θ is the parameter. Thus, in our case, the observed deviance (OD) and the simulated deviance (SD) are respectively given by

$$OD^{(t)} = D(N|\theta^{(t)}, E) \tag{5.5}$$

$$SD^{(t)} = D(N^{(t)}|\theta^{(t)}, E) \tag{5.6}$$

where θ is as defined in (5.3).

The scatter plot for C-G model and Data 1 based on the deviance is shown in Figure 5.4 (b), and the Bayesian p-values for model-data pairs considered are shown in Table 5.5. The p-values are given by $P(X_1 \geq X_2)$ or $P(OD \geq SD)$ as the case may be [10,33]. Apart from C-G which has both p-values for Data 1 within the recommended 95% probability range of (0.025, 0.975) [10, 80], all other model-data pairs have both p-values outside the interval or one of the values is within the range and the other is outside.

Table 5.5: Bayesian p-values for all model-data pairs condidered.

Model	p-value					
	Data 1		Data 2		Data 3	
	(X_1, X_2)	(OD, SD)	(X_1, X_2)	(OD, SD)	(X_1, X_2)	(OD, SD)
C-G	0.969	0.497	0.000	0.492	0.000	0.515
P-G	0.435	0.000	–	–	–	–
C-IG	1.000	0.513	0.000	0.504	0.000	0.497
P-IG	0.000	0.000	–	–	–	–

5.4 Other Observations

5.4.1 Comparison of ϕ Values Generated from the Three Data Sets

Figures 5.5(a) and 5.5(b) were produced to assess how the values of ϕ generated via the C-G model (was done for C-IG too) compare across Data 1, Data 2 and Data 3. Figure 5.5(a) compares the values of ϕ generated from Data 1 with those generated from Data 2 and Figure 5.5(b) does so for Data 1 and Data 3. In each case, the plots were made with the top 10000 corresponding values from the two data sets under consideration. In both figures almost all the points are above the line of equality with a strip pattern. Figure 5.6 is a recast of Figure 5.5(a) using the observed count as the plot character. Each coloured strip in the figure represents an observed count of a given magnitude. We observe that strips of higher counts are nearer the line of equality than strips of lower counts (more noticeable in Figure 5.7). The comparison between the ϕs generated from the three data sets with C-IG show a similar pattern to that of C-G (see Figures 5.7, E.6(a) and E.6(b)) but the strips are more distinct than in the case of C-G. The figures show that ϕs generated from Data 1 tend to be larger than those from Data 2 and Data 3 with the disparity being larger for smaller observed counts. As acknowledged in Section 5.1, Data 2 and Data 3 are truncated versions of Data 1 but the models fitted to them are not for truncated data, and this could be the reason for which the ϕ's generated from Data 2 and Data 3 are smaller than those from Data 1.

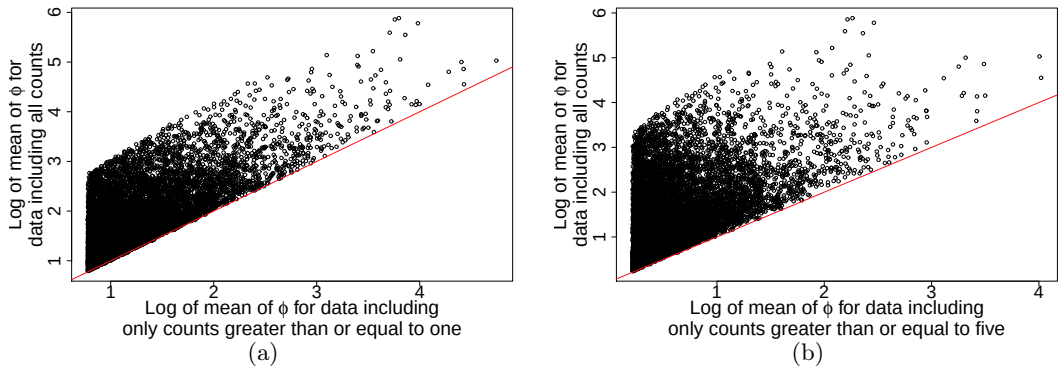


Figure 5.5: (a) Scatter plot of logarithm of posterior means of ϕ for Data 1 against corresponding values for Data 2, and (b) Scatter plot of logarithm of posterior means of ϕ for Data 1 against corresponding values for Data 3, using results from C-G. The plots were made with the top 10000 corresponding values from the two data sets under consideration.

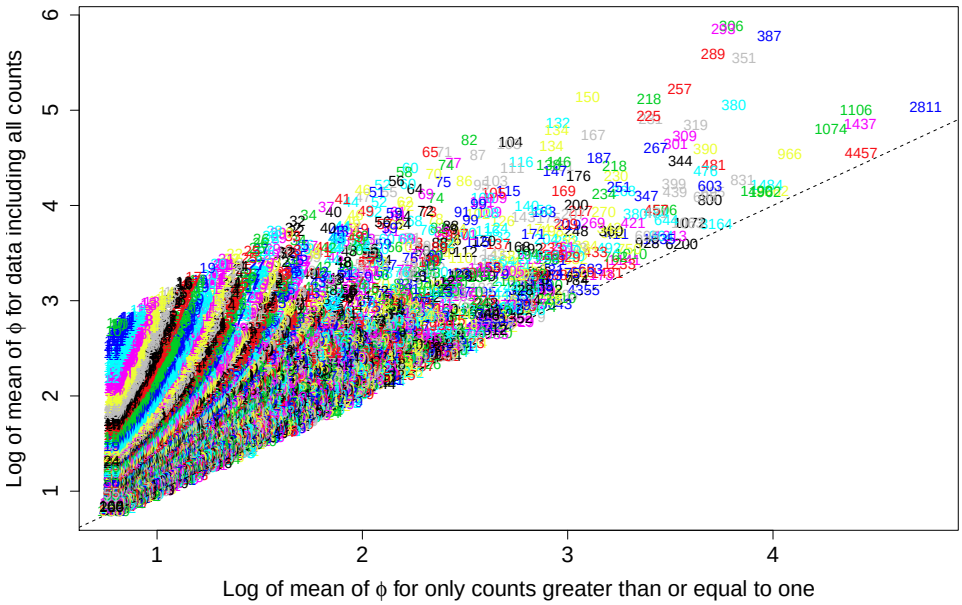


Figure 5.6: Scatter plot of logarithm of posterior means of ϕ for Data 1 against corresponding values for Data 2 using results from C-G. This is a recast of Figure 5.5(a) using the number of counts as the plot character. Each colour strip represent a given observed count.

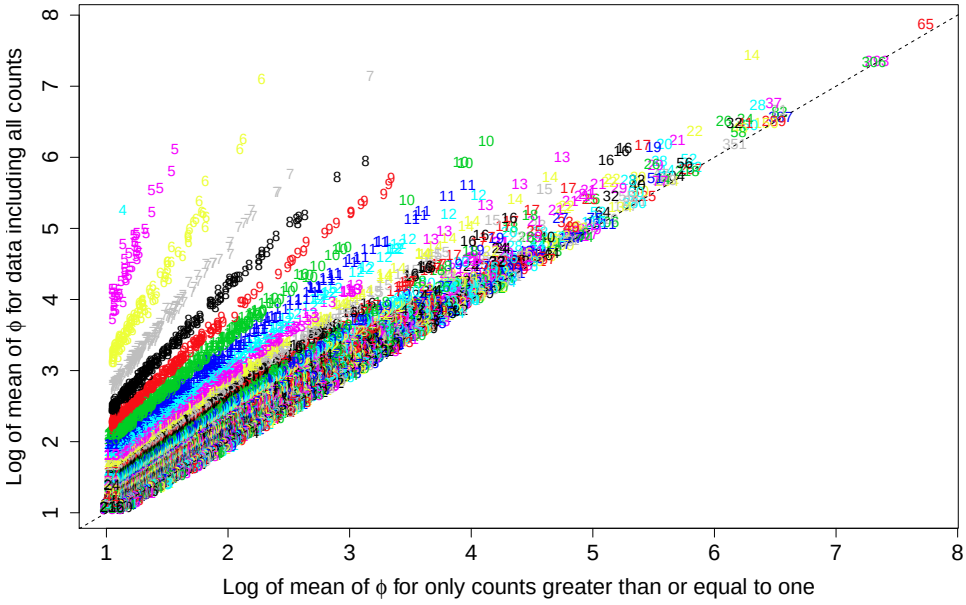


Figure 5.7: Scatter plot of logarithm of posterior means of ϕ for Data 1 against corresponding values for Data 2 using results from C-IG.

5.4.2 Comparison of Mean Replicate Count with Observed Count (N)

Tables C.1 and D.1 present the observed counts, mean replicate counts, expected counts, RR and ϕ values of 50 randomly sampled drug and adverse event pairs for C-G and P-G when Data 1 is used. The tables show that some of the mean replicate counts are substantially different from the observed count. The expected counts of these pairs are very small relative to the observed count, with the value of RR very high and substantially larger than the associated ϕ value. However these ϕ values are greater than one. The foregoing observation holds for Data 2 and Data 3 as well, for C-G, but some of the ϕ values of C-G are less than one in the case of Data 3. Unlike C-G and P-G, C-IG and P-IG tend to generate replicate counts that are closer to the observed count (than by C-G and P-G) when the expected count is small, for Data 1. For Data 2 and Data 3, C-IG's behaviour is not consistent as there are instances of generation of replicates that are relatively smaller than the observed count.

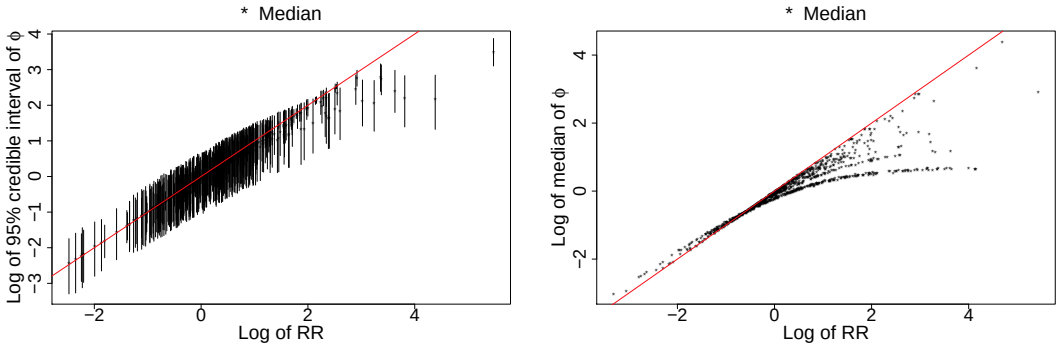
The size of the ϕs generated by the models have an affect on the replicate counts since the replicate counts are proportional to the ϕs. The ability of the Inverse Gamma distribution to accommodate ϕs that are larger than can be accommodated by the Gamma distribution explains why the replicate counts generated by C-IG and P-IG are closer to observed counts than by those generated by C-G and P-G for Data 1. Model C-IG's behaviour for Data 2 and Data 3 is not consistent with its behaviour for Data 1 because the ϕs generated from Data 2 and Data 3 tend to be smaller than those generated from Data 1 as seen in Section 5.4.1.

5.4.3 Credible Intervals of ϕ

Figure 5.8(a) shows the logarithm of the 95% credible intervals of ϕ plotted against the logarithm of the relative report rate (RR) for 500 randomly selected drug and adverse event pairs using results from the C-G model and Data 1.

For drug and adverse event pairs with small values of RR, there is general agreement between RR and the posterior median of the corresponding ϕ while for pairs with large RR values the associated posterior median of the ϕs tend to be much smaller. The credible intervals are also shorter for some drug and adverse event pairs than others.

To have a clearer picture of what is happening, Figure 5.8(b) was produced from the medians of the ϕs of 1000 drug and adverse event pairs randomly sampled from the results from C-G and Data 1. The figure shows a fanning out of the points to the right below the



(a) (b)

Figure 5.8: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for C-G. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 1, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for C-G. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 1.

line of equality (in red) for drug and adverse event pairs with comparatively higher RR, while the opposite occurs to a much smaller degree for pairs with comparatively lower RR. The phenomenon described above holds for all other model-data combinations considered but to a less or greater degree (see Figures C.10(b), C.15(b), D.5(b), E.5(b), E.10(b), E.15(b) and F.5(b)). Figure 5.8(b) shows an interesting pattern in which the points appear to separate into strips. Figure 5.9 is another rendition of Figure 5.8(b) but using the number of reports (observed count) of the drug and adverse event pairs as the plot character, which shows the points separate into strips according to the magnitude of the observed count, so that pairs with small counts are further away below the line of equality with increasing RR and further away above the line with decreasing RR. For each strip (given number of reports), the further to the right a point is, the smaller the associated expected count E_{ij}. The median (of the ϕs) increases with increasing RR albeit more slowly than RR and only up to certain values of RR, beyond which the median remains essentially constant. This pattern is unequivocal only for the smallest numbers of reports. The results described above show that large RR values associated with combinations with small observed and expected counts are shrunk considerably; the shrinkage is more pronounced where the expected count is very small.

We next take a look at how the lower bounds (RR_{025}) of the 95% confidence interval estimates for RR compare with the lower bounds (ϕ_{025}) of the 95% credible interval estimates for ϕ.

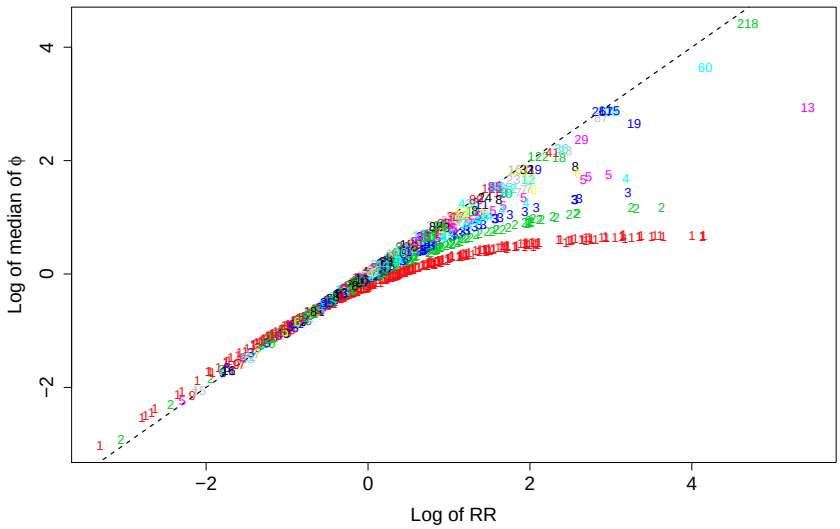


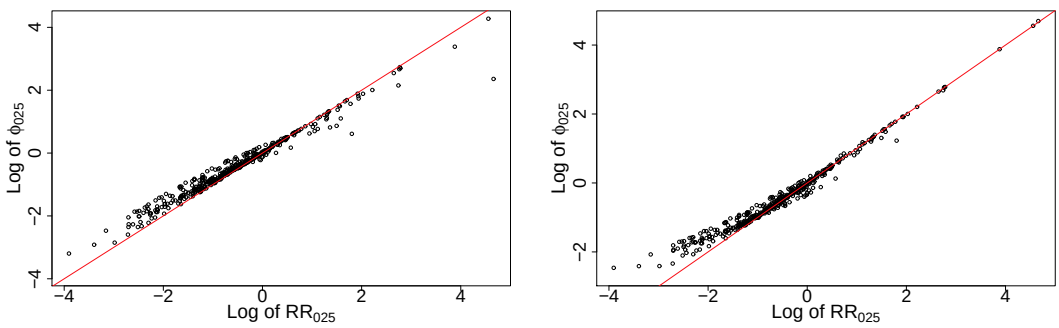
Figure 5.9: Logarithm of posterior medians of ϕ against the logarithm of the relative report rate, RR, for C-G. This figure is a recast of Figure 5.8(b) using the number of counts as the plot character. Each colour strip represents a given observed count.

The value of RR_{025} for each drug and adverse event pair was estimated by the inversion of the Poisson cumulative distribution function, assuming each observed count is Poisson distributed. We solve for the critical Poisson value, Q_{025} – the 2.5% quantile corresponding to the observed count N – and divide it by expected count E. That is:

$$\begin{aligned} Q_{025} &= \texttt{qpois}(0.025, N) \\ \text{RR}_{025} &= \frac{Q_{025}}{E} \end{aligned}$$

Figure 5.10(a) shows the scatter plot of the logarithm of ϕ_{025} against the logarithm of RR_{025} for C-G and Data 1 using the set of drug and adverse event pairs involved in Figure 5.9. The plots for the other model-data pairs are shown in the appendices. The strip pattern relating to small observed counts occurring to the right of the line of equality (broken line) in Figure 5.9 are not present in Figure 5.10(a). This is due to the fact that some drug and adverse event pairs with small observed counts have RR_{025} values that are considerably smaller than the corresponding RR values. This leads to instances where the RR value is greater than the associated ϕ but the RR_{025} value is slightly greater than or less than ϕ_{025} value, shifting the corresponding points (plot character) closer to or to the left of the line of equality in Figure 5.10(a). An example of such a pair is Propoxyphene-Asphyxia, which has observed and expected counts of 5 and 0.598 respectively. The RR

value for the pair is 8.36 which is greater than the ϕ value of 4.13, but the value of RR_{025} (1.67) is less than the value of ϕ_{025} (1.69). In particular drug and adverse event pairs with very small counts have the value of RR_{025} to be 0. The points (plot character) of such pairs do not appear in the figure when plotted on the logarithmic scale since the logarithm of 0 is undefined. On the ordinary scale such points are concentrated on the line $RR_{025} = 0$. It appears from Figure 5.10(a) that the bigger values of RR_{025} tend to be greater than their associated ϕ_{025} values while there is a large number of drug and adverse combinations for which the ϕ_{025} value is greater than the RR_{025} value. This is in contrast to what is observed in Figure 5.9, which suggest there is a large number of drug and adverse combinations for which the ϕ value is smaller than the associated RR value.



(a) (b)

Figure 5.10: Scatter plot of logarithm of ϕ_{025} against the logarithm of RR_{025} for (a) C-G and Data 1 and (b) C-IG and Data 1.

The behaviour of C-G with Data 2 and Data 3 appears to be different from that of Data 1 (see Figures E.11(a) and E.16(a)); the RR_{025} values appear be generally greater than the associated ϕ_{025} values. The situation for P-G and Data 1 is essentially like that of C-G and Data 1 as Figure D.6 shows. For C-IG and Data 1, Figure 5.10(b) suggests majority of the RR_{025} values are quite close to their associated ϕ_{025} values as most of the points lie close to or along the line of equality, except for the smaller values of RR_{025} where the associated ϕ_{025} values are larger and so the points occur to the left of the line of equality as in Figure 5.10(b) or do not appear in the figure since $RR_{025} = 0$, as explained in the case of C-G. Figures E.11(a) and E.16(a) show that in the case of C-IG with Data 2 and Data 3, majority of the RR_{025} values are greater than their corresponding ϕ_{025} values except, as in the case of Data 1, the smaller values of RR_{025} tend to be less than their associated ϕ_{025} values. It is unsurprising that for Data 2 and Data 3 the RR_{025} values are generally greater than the associated ϕ_{025} values, recall that ϕs tend to be smaller in the

case of Data 2 and Data 3, as pointed out in Section 5.4.1. Model P-IG's behaviour with Data 1 with respect to the relationship between ϕ_{025} and RR_{025} is similar to that of C-IG as depicted by Figure F.6.

Tables C.2a through C.2c of Appendix C show the top hundred drug and adverse event combinations selected by C-G based on the lower bound (ϕ_{025}) of the 95% credible interval estimate for ϕ, using Data 1. The tables for other model-data pairs are in the respective appendices. The drug and adverse event combinations have been ordered in descending magnitude of ϕ_{025}. Tables C.2a through C.2c also present the ranks assigned to the drug and adverse event combinations based on the values of RR, RR_{025}, ϕ and ϕ_{025}, which are designated RK1, RK11, RK3 and RK33 respectively.

The ranks presented in Tables C.2a through C.2c for C-G and Data 1 suggest that the agreement between the ranks based on ϕ (RK3) and ϕ_{025} (RK33) is higher than that between the ranks based on RR (RK1) and RR_{025} (RK11). For instance the (RK1, RK11) values for the pair Metformin-Acidosis Lactic are (437, 94) and the (RK3, RK33) values for the same pair are (20, 19), showing a greater disparity between the ranks based on RR and RR_{025} than that based on ϕ and ϕ_{025}. There are many such examples in the tables. The disparity between RK1 and RK11 and the close agreement between RK3 and RK33 appear to be general; the (RK1, RK11) correlation coefficient is 0.14, indicating there is virtually no correspondence between RK1 and RK11, while the (RK3, RK33) correlation coefficient is 0.79, indicating a strong correspondence between RK3 and RK33. Table 5.6 gives the (RK1, RK11) and (RK3, RK33) rank correlation coefficients for RR and all the proposed models for the three data sets.

Table 5.6: Rank correlation coefficients for models.

Model	Ranks	Correlation Coefficients		
		Data 1	Data 2	Data 3
RR	RK1, RK11	0.14	0.14	0.89
C-G	RK3, RK33	0.79	0.86	0.89
P-G	RK3, RK33	0.79	–	–
C-IG	RK3, RK33	0.82	0.86	0.88
P-IG	RK3, RK33	0.83	–	–

The correlation coefficients in the Table 5.6 show that the disparity between RK1 and RK11 holds for Data 2 as well, but not for Data 3. The difference in the way RR and RR_{025}

rank the drug and adverse event pairs is explained by the fact that RR gives preference to drug and adverse event pairs with low observed and expected counts, especially pairs with very low expected counts, while RR_{025} does not appear to have such preference. In the absence of low counts as is the case for Data 3, RR's behaviour, with respect to the ranking of the drug and adverse event pairs, is similar to that of RR_{025} and hence the correspondence between RK1 and RK11 for Data 3 as evidenced by the high correlation coefficient of 0.89. Neither ϕ nor ϕ_{025} appear have preference for drug and adverse event pairs with low observed and expected counts and so the presence or absence of such pairs does not influence the ranking by ϕ and ϕ_{025}. This is the reason for the high (RK3, RK33) correlation coefficient for all three data sets.

5.5 Selection of Drug and Adverse Event Pairs

Table 5.7a shows the number of drug and adverse event pairs common to the top 1000 combinations selected by the possible model pairs based on the posterior mean of ϕ for the three data sets and Table 5.7b shows the corresponding Spearman rank correlation coefficient for the ranks assigned by the models. The large number of drug and adverse event pairs common to the top 1000 selection of C-G and P-G and also C-IG and P-IG show how close these pairs of models are in their selection performance when Data 1 are used, more so when the high correlation coefficient between them is taken into consideration – not only do they have a large number in common, the ranks they assign to the drug and adverse event pairs are almost the same as far as the top 1000 pairs are concerned when Data 1 are used.

Table 5.7a: Number of drug and adverse event combinations common to the top 1000 combinations selected by all possible model pairs based on the point estimate of ϕ.

	Data 1				Data 2		Data 3	
Model	C-G	P-G	C-IG	P-IG	C-G	C-IG	C-G	C-IG
C-G	–	992	555	586	–	529	–	593
P-G	992	–	553	582	–	–	–	–
C-IG	555	553	–	961	529	–	593	–
P-IG	586	582	961	–	–	–	–	–

Values for model pairs whose ϕs do not have the same prior specification (such as C-G and C-IG) suggest a difference in the way they select the top 1000 drug and adverse event

Table 5.7b: Spearman rank correlation values for the possible model pairs using the ranks of the top 1000 selected combinations, based on the point estimate of ϕ.

	Data 1				Data 2		Data 3	
Model	C-G	P-G	C-IG	P-IG	C-G	C-IG	C-G	C-IG
C-G	–	0.999	0.359	0.383	–	0.258	–	0.406
P-G	0.999	–	0.337	0.355	–	–	–	–
C-IG	0.359	0.337	–	0.989	0.258	–	0.406	–
P-IG	0.383	0.355	0.989	–	–	–	–	–

pairs as they have a lower number of common pairs and lower correlation coefficient for all the data sets, in particular Data 2.

Table 5.8a: Number of drug and adverse event combinations common to the top 1000 combinations selected by all possible model pairs based on the estimate for ϕ_{025}.

	Data 1				Data 2		Data 3	
Model	C-G	P-G	C-IG	P-IG	C-G	C-IG	C-G	C-IG
C-G	–	989	690	705	–	634	–	684
P-G	989	–	685	700	–	–	–	–
C-IG	690	685	–	983	634	–	684	–
P-IG	705	700	983	–	–	–	–	–

The number of drug and adverse event combinations common to the top 1000 combinations selected by the possible model pairs based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ for the three data sets and the corresponding rank correlations are shown in Tables 5.8a and 5.8b respectively. Tables 5.7a through 5.8b show that both the number of drug and adverse event combinations selected in common and the corresponding correlation coefficients are higher when the selection is based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ than when the selection is based on the point estimate of ϕ, except for the model pair C-G and P-G, where the number of common combinations drop marginally from 992 to 989 and the correlation also drop from 0.999 to 0.998. Thus in general, the behaviour of the proposed models are more similar when compared on the basis of the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ than on the basis of the point estimate of ϕ.

Table 5.8b: Spearman rank correlation values for the possible model pairs using the ranks of the top 1000 selected combinations, based on the estimate for ϕ_{025}.

	Data 1				Data 2		Data 3	
Model	C-G	P-G	C-IG	P-IG	C-G	C-IG	C-G	C-IG
C-G	–	0.998	0.516	0.529	–	0.466	–	0.558
P-G	0.998	–	0.494	0.508	–	–	–	–
C-IG	0.516	0.494	–	0.993	0.466	–	0.558	–
P-IG	0.529	0.508	0.993	–	–	–	–	–

Table 5.9: Average difference between the point estimates and between the lower bounds of 95% confidence/credible interval estimates for all possible model pairs.

Data 1		
Model Pairs	Average of $\lvert\phi 1 - \phi 2\rvert$ (Δ_{ave})	Average of $\lvert\phi 1_{025} - \phi 2_{025}\rvert$ ($\Delta 025_{\text{ave}}$)
C-G, P-G	0.026	0.018
C-G, C-IG	0.462	0.207
C-G, P-IG	0.459	0.201
P-G, C-IG	0.459	0.210
P-G, P-IG	0.451	0.201
C-IG, P-IG	0.084	0.039
Data 2		
Model Pairs	Average of $\lvert\phi 1 - \phi 2\rvert$	Average of $\lvert\phi 1_{025} - \phi 2_{025}\rvert$
C-G, C-IG	0.251	0.197
Data 3		
Model Pairs	Average of $\lvert\phi 1 - \phi 2\rvert$	Average of $\lvert\phi 1_{025} - \phi 2_{025}\rvert$
C-G, C-IG	0.355	0.248

To see why the number of drug and adverse event combinations common to the top 1000 combinations selected by the possible model pairs and the corresponding correlation coefficients are higher when the selection is based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ than when the selection is based on the point estimate of ϕ, we denote the ϕ generated for a drug and adverse event combination from one of the models, say C-G, as $\phi 1$ and the ϕ generated for the same combination from another model, say P-G, as $\phi 2$, and their lower bounds (ϕ_{025}) of the 95% credible interval estimates of ϕ as $\phi 1_{025}$ and $\phi 2_{025}$ respectively. Table 5.9 shows the average of the difference $|\phi 1 - \phi 2|$ (Δ_{ave}) and the average of the difference $|\phi 1_{025} - \phi 2_{025}|$ ($\Delta 025_{\text{ave}}$) for all possible model pairs for the three data sets considered. From the table, every Δ_{ave} value is greater than the associated $\Delta 025_{\text{ave}}$ value. Thus for any model pair, the values of the models are closer when the lower bounds (ϕ_{025}) of the 95% credible interval estimates of ϕ are used than when the point estimates of ϕ are used. This appears to be the major reason for which the number of drug and adverse event combinations common to the top 1000 selected by the possible model pairs are higher when the values of ϕ_{025} are used. Also both Δ_{ave} and $\Delta 025_{\text{ave}}$ are smaller where the models involved have the same prior specification for ϕ (for example, C-G and P-G). This also appear to explain why such model pairs have a higher number of common drug and adverse event combinations in their top 1000 selected combinations than model pairs that do not have the same prior specification for ϕ.

5.6 Model Selection

The modeling process started with C-G model. The closeness of the posterior mean of ν to one prompted the consideration of P-G. The smallness of the magnitudes of the values of ϕ produced by these models (relative to the values produced by GPS) led to the consideration of C-IG, and that in turn led to P-IG for the same reason as C-G led to P-G. The question that arises is: which of these models is more plausible, given the data? Formally, this question is answered by one of two approaches: namely by considering (i) the Deviance Information Criterion (DIC) [79] values of the models or (ii) the posterior model probabilities via Reversible Jump Markov chain Monte Carlo (RJMCMC) [40].

5.6.1 DIC

Given the parameter θ (as defined in Equation 5.3), DIC and P_D are given by

$$DIC = 2\overline{D(N|\theta, E)} - D(N|\bar{\theta}, E) \tag{5.7}$$

$$P_D = \overline{D(N|\theta, E)} - D(N|\bar{\theta}, E) \tag{5.8}$$

where $\overline{D(N|\theta, E)}$ is the mean deviance across the sample MCMC iterates and $D(N|\bar{\theta}, E)$ is the deviance as computed from the MCMC estimate of the mean value $\bar{\theta}$ of θ. Suitable models are characterized by smaller DIC values [79], for a given data. Values of DIC and effective number of parameters (P_D) for all model-data combinations considered are presented in Table 5.10.

Table 5.10: Values of DIC and P_D for the various model-data combinations.

Model	Data 1		Data 2		Data 3	
	DIC	P_D	DIC	P_D	DIC	P_D
C-G	2280576	319531	1984702	105877	1040715	32753
P-G	2287577	313491	–	–	–	–
C-IG	2342507	345650	1937525	118740	1014623	44375
P-IG	2366524	329726	–	–	–	–

A comparison of the values in the table leads to the fact that for Data 1 C-G has the smallest DIC value of 2280576 followed by P-G at 2287577, while for Data 2 and Data 3 C-IG has the smallest DIC values of 1937525 and 1014623 respectively. In all cases P_D is far smaller than the total number of parameters under estimation, with P-G having the smallest P_D value in case of Data 1 (313491). Model C-G has the smallest P_D values in the case of Data 2 (105877) and Data 3 (32753).

5.6.2 RJMCMC

We tried to discriminate between P-G and C-G through Reversible Jump Markov chain Monte Carlo (RJMCMC). The technique of RJMCMC extends the MH algorithm such that two or more models can be implemented via a single Markov chain, thereby allowing the estimation of the posterior probabilities of the models, as the models themselves are viewed as parameters [40, 44]. The posterior model probability $\pi(m|\mathbf{x})$ of model m given the data $\mathbf{x}$ is estimated as the fraction of the time the Markov chain is in model m, after

the burn-in. In RJMCMC, the joint posterior density of the parameters and model is given by

$$\pi(\theta_m, m|\mathbf{x}) \propto f(\mathbf{x}|\theta_m, m)h(\theta_m|m)p(m) \tag{5.9}$$

where $f(\mathbf{x}|\theta_m, m)$ is the likelihood of model m given its set of parameters θ_m, $h(\theta_m|m)$ is the prior distribution of the parameters given model m, $p(m)$ is the prior probability of model m and $\mathbf{x}$ is the data. At each iteration, (i) we use the current model to update the parameters using the MH algorithm and then (ii) update the model conditional on the current set of parameters using RJMCMC algorithm [40, 44]. We describe the RJMCMC algorithm, in our context, as follows: we designate

$$\begin{aligned} \text{P-G as Model } m_1 : \theta_1 &= \{\alpha, \beta, \phi\} \\ \text{C-G as Model } m_2 : \theta_2 &= \{\alpha, \beta, \phi, \nu\} \end{aligned}$$

We propose to move from state $(\theta, m)_t$ at iteration t to state $(\theta^{'}, m^{'})$. Suppose the chain is in P-G at iteration t, we have $m = m_1$ and $\theta = \theta_1$, then we propose to move to C-G, in which case $m^{'} = m_2$ and $\theta^{'} = \theta_2$. We set

$$\begin{aligned} \alpha^{'} &= \alpha \\ \beta^{'} &= \beta \\ \phi^{'} &= \phi \\ \nu^{'} &= \text{v} \end{aligned}$$

where $\text{v} \sim q(\text{v})$, q is an arbitrary proposal distribution such that the set of possible values of ν, Λ, satisfies $\Lambda \subseteq \Omega$, where Ω is the support of q. We simulated v as $\text{v} \sim N(1.031, 0.01)$ taking a cue from the MCMC estimate of ν in an earlier run of C-G.

The move to state $(\theta^{'}, m^{'})$ is accepted with probability min $(1, \varpi)$ where ϖ is given by

$$\varpi = \frac{\pi(\theta^{'}, m^{'}|N)\mathbb{P}(m|m^{'})}{\pi(\theta, m|N)\mathbb{P}(m^{'}|m)q(\text{v})} \left| \frac{\partial(\theta^{'})}{\partial(\theta, \text{v})} \right| \tag{5.10}$$

where $\mathbb{P}(m|m^{'})$ is the probability of proposing to move to model m given the chain is currently in model $m^{'}$ and $\mathbb{P}(m^{'}|m)$ is the probability of proposing to move to model $m^{'}$ given the chain is currently in model m. Since there are only two models, we will always be proposing to move from one to the other and so $\mathbb{P}(m|m^{'}) = \mathbb{P}(m^{'}|m) = 1$. The term

$|\partial(\theta^{'})/\partial(\theta, \mathrm{v})|$ in (5.10) is the Jacobian of transformation and is given by

$$\begin{aligned}
\left|\frac{\partial(\theta^{'})}{\partial(\theta, \mathrm{v})}\right| &= \begin{vmatrix} \frac{\partial\alpha^{'}}{\partial\alpha} & \frac{\partial\beta^{'}}{\partial\alpha} & \frac{\partial\phi_{11}^{'}}{\partial\alpha} & \frac{\partial\phi_{12}^{'}}{\partial\alpha} & \cdots & \frac{\partial\phi_{IJ}^{'}}{\partial\alpha} & \frac{\partial\nu^{'}}{\partial\alpha} \\ \frac{\partial\alpha^{'}}{\partial\beta} & \frac{\partial\beta^{'}}{\partial\beta} & \frac{\partial\phi_{11}^{'}}{\partial\beta} & \frac{\partial\phi_{12}^{'}}{\partial\beta} & \cdots & \frac{\partial\phi_{IJ}^{'}}{\partial\beta} & \frac{\partial\nu^{'}}{\partial\beta} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial\alpha^{'}}{\partial\mathrm{v}} & \frac{\partial\beta^{'}}{\partial\mathrm{v}} & \frac{\partial\phi_{11}^{'}}{\partial\mathrm{v}} & \frac{\partial\phi_{12}^{'}}{\partial\mathrm{v}} & \cdots & \frac{\partial\phi_{IJ}^{'}}{\partial\mathrm{v}} & \frac{\partial\nu^{'}}{\partial\mathrm{v}} \end{vmatrix} \\
&= \begin{vmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{vmatrix} \\
&= 1
\end{aligned}$$

It follows that

$$\varpi = \frac{\pi(\theta^{'}, m^{'}|N)}{\pi(\theta, m|N)q(\mathrm{v})}$$

We set $(\theta, m)_{t+1} = (\theta^{'}, m^{'})$ if the move is accepted, else we set $(\theta, m)_{t+1} = (\theta, m)_t$. For a move from state $(\theta^{'}, m^{'})$ to (θ, m), we set

$$\begin{aligned}
\alpha &= \alpha^{'} \\
\beta &= \beta^{'} \\
\phi &= \phi^{'} \\
\mathrm{v} &= \nu^{\prime}
\end{aligned}$$

and the acceptance probability is given by $\min(1, \varpi^{-1})$.

Figure 5.11 shows the trace plots of the RJMCMC based on P-G and C-G when the models are thought to be equipropable a priori, and Figure 5.12 shows the trace when the prior probabilities have been re-weighted in favour of the P-G (0.999999). As the two figures (Figures 5.11 and 5.12) show, the Markov chain was for the most time in C-G (model 2) in both cases. From Table 5.4a, the posterior mean and standard deviation of ν for C-G are 1.031 and 0.003 respectively. Thus ν is more than 10 (10.33) standard deviations from 1, the value at which P-G is applicable. This explains why the Markov chain was for the most time in C-G, and hence C-G is more plausible than P-G.

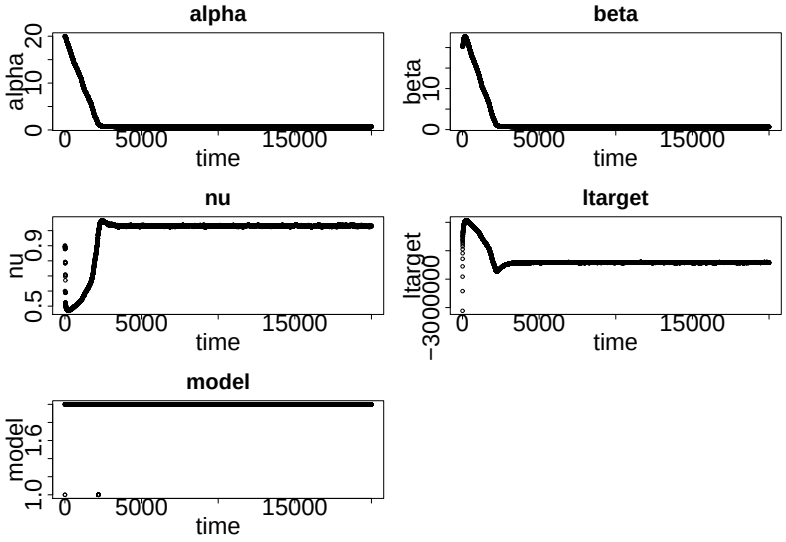


Figure 5.11: Trace plots of α, β, ν, log of the target distribution and model for the RJMCMC based on P-G and C-G when they are thought to be equiprobable a priori.

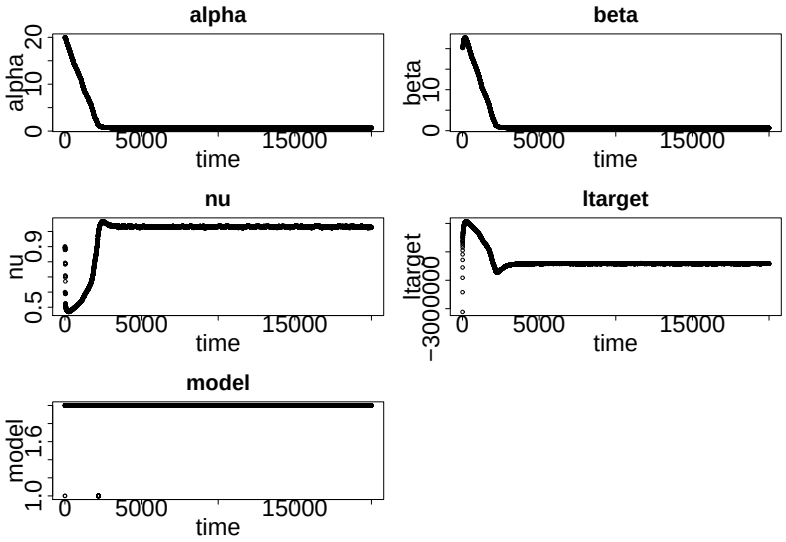


Figure 5.12: Trace plots of α, β, ν, log of the target distribution and model for the RJMCMC based on P-G and C-G when prior probability of P-G is set at 0.999999.

Chapter 6

Discussion of Results and Comments

We examine the findings of Chapter 5 and point out their implications in this chapter.

6.1 Suitable Model

The model(s) we are seeking, in the context of the problem on hand, is one that reasonably adjust the observed counts (in the form of the replicate counts it yields), and generates ϕs that reflect the nature of the association between drugs and adverse events. In a sense the posterior distribution of ϕ could be regarded as the Bayesian analogue of RR. A desirable model must therefore provide a vehicle for revising the values of RR in the form of the ϕ values that result, and the replicate counts must be reasonably close to the observed counts [33]. The ensuing discussion will focus on identifying the model(s) most suitable for the data and then move the commentary from the global to the particular traits of the chosen model(s).

6.2 Observations

Observation 1

It came up in Section 5.3.1 that there is a fit between the distribution of the ϕs and the Gamma distribution corresponding to the MCMC estimate of the means of α and β for the model-data pairs considered. This observation was also true for the distribution of the ϕs and the Inverse Gamma distribution corresponding to the MCMC estimate of the means of φ and ψ (see Figures 5.3, C.11(b), C.16(b), D.1, E.1, E.11(b), E.16(b) and F.1), so, on this basis, nothing seems to be amiss with the models, where they are applicable to a given

data.

Observation 2

In keeping with the idea of a suitable model outlined in Section 6.1, we sought an instrument that allows us to determine whether or not the replicate data compares favourably with the observed data at each iteration. That instrument was found in the deviance and variance measures presented in Section 5.3.2. The idea is that a suitable model will produce replicate data that is similar to the observed data [33], from the perspectives of the individual values and variability thereof. The Bayesian p-values for C-G and Data 1 and P-G and Data 1, using X_1 and X_2 (Section 5.3.2), were found to be 0.969 and 0.435 respectively, which are within the recommended range of (0.025, 0.975) [10, 80]. However, the p-values for the other six model-data pairs were found to be either zero or one. Thus C-G and P-G are the ones amongst the four models that give favourable results for Data 1 and none of the models yield positive results for Data 2 and Data 3 when assessed on the basis of X_1 and X_2. Using the deviance measures OD and SD, we found that the p-values for C-G and C-IG are around 0.500 for the all three data sets while those for P-G and P-IG when Data 1 are used are zero. Thus C-G is the only model that performs favourably on both the deviance and variance criteria when Data 1 are used and none of the models score favourably on both criteria for Data 2 and Data 3, where applicable.

As pointed out in Chapter 5, Data 2 and Data 3 are truncated versions of Data 1 but the models fitted to them are not for truncated data. This explains why the parameter estimates for Data 2 and Data 3 are biased as evidenced by the results presented in Chapter 5 (and recounted above), where the p-values obtained via the variance measures X_1 and X_2 for Data 2 and Data 3 are all zero (Section 5.3.2) and the ϕs generated from Data 2 and Data 3 are smaller than those from Data 1 (Section 5.4.1). Indeed the (marginal) posterior distributions of the ϕs are not independent; by dropping the cells with counts equal to zero or less than five, one is essentially doing away with the corresponding ϕ values needed to 'accurately' estimate the ϕs of the remaining cells. The case of the ϕs generated from Data 1 tending to be larger than those from Data 2 and Data 3 (Section 5.4.1) attest to the 'underestimation' of the ϕs when Data 2 and Data 3 are used. They (Data 2 and Data 3) were, however, included in the investigation to see how the models fared with them. The models are clearly not suitable for them.

Again, we saw in Section 5.4.2 that C-IG's behaviour with Data 2 and Data 3 was not consistent with its behaviour with Data 1 as, in some instances, the replicate counts

generated by C-IG from Data 2 and Data 3 were considerably smaller than the observed count, as opposed to the replicate counts generated from Data 1 being closer to the observed counts. The replicate counts are proportional to the ϕs, so the replicate counts would be small if the associated ϕ values are small, as it happens in the case of Data 2 and Data 3.

Observation 3

As observed in Section 5.6, we sought to discriminate between the models on the basis of DIC. Model C-G was found to have the smallest DIC value followed by P-G (Section 5.6). This reinforces the view that C-G is the most suitable for modeling Data 1. Also, an attempt was made to choose between C-G and P-G using RJMCMC initially, but apparently the posterior probability of Model C-G is so high that the Markov chain was for the most time in C-G. The reason for this, as pointed out in Section 5.6, is due to the fact that the mean value of ν (1.031) is in the units of its standard deviation (0.003) far removed from 1, the value at which P-G is valid.

6.3 Model of Choice

Given what has been noted under Observation 1 through Observation 3, one is inclined to conclude that C-G is the most suitable amongst the four models for modeling Data 1. The models considered are not suitable for modeling Data 2 and Data 3 as pointed out in Section 6.2.

Turning to C-G as the suitable model for modeling Data 1, we saw in Figure 5.9 (Section 5.4.3) that the medians separate into strips, with those representing higher RR (usually associated with low observed or expected count) being the farthest away below the line of equality with increasing RR and those of lower RR slightly away above the line with decreasing RR. Thus not only has there been a shrinkage effect in the ϕs, but the shrinkage tended to be more pronounced in ϕs corresponding to drug and adverse event pairs with low observed and expected count (especially those with very small expected counts) than those of high observed or expected count. This is expected in a hierarchical model [10, 33, 60] as the marginal mean effects (mean of each ϕ_{ij}) gravitate towards the mean effect of common distribution to which all the cells belong, owing to mutual sharing of information across the cells. This attribute of C-G could be key in its overall performance, as it holds the potential for reducing false positives.

Indeed Figure 5.10(a) also suggest there has been shrinkage; the bigger values of RR_{025}

tend to be greater than their associated ϕ_{025} values while the smaller values RR_{025} tend to be less than their associated ϕ_{025} values, even though there is a large number of drug and adverse combinations for which the ϕ_{025} value is greater than the RR_{025} value. However the evidence of shrinkage is not as dramatic as seen in Figure 5.9, which suggest there is a large number of drug and adverse event combinations for which the ϕ value is smaller than the associated RR value. Particularly so for combinations with very small observed count and even smaller expected count as observed above. Shrinkage appears to be subdued in Figure 5.10(a) because values of RR_{025} corresponding to drug and adverse combinations with very small observed and expected counts are relatively smaller.

The high (RK3, RK33) correlation coefficient of 0.79 for C-G and Data 1 is an indication that the degree of agreement between the rankings based on ϕ and ϕ_{025} is higher than that of the rankings based on RR and RR_{025}, which has a (RK1, RK11) correlation coefficient of 0.14. Thus C-G appears to be more consistent in the ranking of the drug and adverse event combinations than RR. As pointed out in Section 5.4.3, RR gives preference to drug and adverse event pairs with low observed and expected counts, especially pairs with very low expected counts, while RR_{025} does not have such preference, hence the lack of agreement between the rankings of RR and RR_{025} for Data 1 and Data 2, which include drug and adverse event combinations with small observed counts and often smaller expected counts.

6.3.1 Drugs Common and Uniquely Chosen by RR, C-G and GPS

We compare the three methods RR, GPS (DuMouchel's method [19]) and C-G on the basis of the point estimates of RR, λ and ϕ and also on the basis of the the lower bounds RR_{025}, λ_{025} and ϕ_{025} of the 95% confidence/credible interval estimates of RR, λ and ϕ respectively. The C-G (ϕ) values used are the geometric mean of the posterior distribution of ϕ_{ij}. This was done to put C-G values on the same scale as the GPS (EBGM) values, so that comparison between C-G and GPS is fair. The point estimate (EBGM) of λ and the lower bound (λ_{025}) of the 95% credible interval estimate of λ were obtained by fitting the GPS model to Data 1. To do so, one has to estimate the values of the hyperparameters α_1, β_1, α_2, β_2 and ϱ as described in Section 4.2.4; namely, by the method of maximum likelihood estimation, in which the product of the marginal densities given by (4.4) is maximized [19]. This was done by applying the R non-linear optimization function nlm to the logarithm of the product of the likelihoods given by (4.4). The maximization took 26 iterations in about 4 minutes, with the starting values ($\alpha_1 = 0.2$, $\beta_1 = 0.1$, $\alpha_2 = 2$,

$\beta_2 = 4$ and $\varrho = 1/3$) suggested by DuMouchel [19]. The estimated hyperparameter values, denoted Est_1, are shown in Table 6.1. The estimates of the hyperparameters obtained by DuMouchel [19] with the original data, denoted Est_2, are also show in the table. It can be seen from an examination of the values that the two estimates, Est_1 and Est_2, are quite close. Nevertheless the two data sets, Data 1 and the original data, are not exactly the same and it is more objective to compare the three models, C-G, RR and GPS, on the same data. Hence the λ (EBGM) and λ_{025} values used in the following comparisons are based on Est_1. The EBGM values are given by [19]

$$\text{EBGM}_{\text{ij}} = 2^{\text{EBlog2}_{\text{ij}}}$$

where

$$\begin{aligned} \text{EBlog2}_{\text{ij}} &= \text{E}\left[\log 2(\lambda_{ij})|N_{ij}\right] \\ &= \text{E}\left[\log(\lambda_{ij})|N_{ij}\right]/\log(2) \end{aligned}$$

and

$$\begin{aligned} \text{E}\left[\log(\lambda_{ij})|N_{ij} = n_{ij}\right] &= \omega_{ij}\left[\Psi(\alpha_1 + n_{ij}) - \log(\beta_1 + E_{ij})\right] \\ &\quad +(1-\omega_{ij})\left[\Psi(\alpha_2 + n_{ij}) - \log(\beta_2 + E_{ij})\right] \end{aligned}$$

where ω_{ij} is as given in Section 4.2.4 and the digamma function $\Psi(x) = \frac{d}{dx}\log[\Gamma(x))]$. Each λ_{025} value was determined as the mean of three 2.5% quantile values obtained from three samples of size one million apiece from the posterior distribution of the corresponding λ.

Table 6.1: Hyperparameter estimates for the GPS model.

Estimates	α_1	β_1	α_2	β_2	ϱ
Est_1 (Data 1)	0.1894	0.05764	1.4001	1.8299	0.1037
Est_2 (Original [19])	0.2041	0.05816	1.4150	1.8380	0.0969

Table 6.2 shows that 198 drug and adverse event pairs are common to the top 1000 pairs chosen by C-G, GPS and RR based on the point estimates of ϕ, λ (EBGM) and RR. The first 100 of the actual drug and adverse event pairs involved are shown in Tables G.1a through G.1c of Appendix G. The drug and adverse event pairs shown in Tables G.1a through G.1c have been sorted in descending magnitude of RR. The ranks assigned to the drug and event pairs by GPS based on the point estimate of λ (EBGM) and the lower

bound (λ_{025}) of the 95% credible interval estimate of λ have been designated RK2 and RK22 respectively. Recall that the ranks assigned by RR and RR_{025} were designated RK1 and RK11 respectively, and those assigned by ϕ and ϕ_{025} were designated RK3 and RK33 respectively.

Based on the lower bounds RR_{025}, λ_{025} and ϕ_{025} of the 95% confidence/credible interval estimates, 662 combinations were found to be common to the top 1000 drug and adverse event pairs selected by the three methods (Table 6.2). Thus the three methods are closer in the way they select the top 1000 combinations when compared on the basis of RR_{025}, λ_{025} and ϕ_{025} than when compared on the basis of the point estimates, as the number they have in common using the point estimates is less than a third of the number they select in common when the lower bounds of the 95% confidence/credible interval estimates are used. Tables G.2a through G.2c of Appendix G show the first 100 of the actual drug and adverse event combinations involved. The drug and adverse event combinations have been sorted in descending magnitude of RR_{025}.

Table 6.2: Number of drug and adverse event pairs common to the top 1000 chosen by combinations of the methods when Data 1 are used.

	Number common	
Model combination	Based on the point estimates of RR, ϕ and λ (EBGM)	Based on the estimates of RR_{025}, λ_{025} and ϕ_{025}
C-G/GPS/RR	198	662
C-G/GPS	616	722
C-G/RR	198	662
GPS/RR	373	926

Table 6.2 also shows the number of drug and adverse event pairs common to the top 1000 selected by combinations of two from the three methods, based on the point estimates and the lower bounds of the 95% confidence/credible interval estimates. The table shows that the number of pairs common to the top 1000 selected by C-G and GPS, based on the point estimates of ϕ and λ (EBGM) respectively, is more than three times the number in common when all three approaches are considered. This reflects the diversity in the way the three methods rank or select the drug and adverse event pairs and the fact that C-G is closer to GPS than to RR in the way it selects the top 1000 drug and adverse event pairs. However, the fact of GPS and RR having more drug and adverse event pairs (373)

in common in their top 1000s than it is between C-G and RR (198) shows GPS is more closely allied to RR in the ranking of the drug and adverse event pairs than to C-G.

The foregoing observation about the relationship between the methods parallels what is observed when the top 1000 drug and adverse event pairs selected by combinations of two from these three methods are compared on the basis of the lower bounds ϕ_{025}, λ_{025} and RR_{025} of the 95% confidence/credible interval estimates. Here, the number of pairs common to the top 1000 selected by C-G and GPS is 722, which is higher than when all three approaches are considered (662). The number of combinations common to GPS and RR (926) is more than that common to C-G and GPS (722), buttressing the above observation that GPS is closer to RR in the ranking of the drug and adverse event pairs than to C-G. Incidentally, in both the use of the point estimates and the lower bounds, the number of combinations selected in common by C-G and RR in their top 1000 combinations is not only the least when the methods are considered in pairs but is also the same as the number selected in common by all three methods. Thus it appears drug and adverse event combinations selected in common by all the three methods is determined by the combinations selected in common by C-G and RR. This situation underscores the above observation that C-G and RR are more different from each other than they are from GPS.

Table 6.3: Average difference between the point estimates and between the lower bounds of 95% confidence/credible interval estimates.

Point Estimates		Lower Bounds of 95% Confidence/ Credible Interval Estimates	
Item Pairs	Average Difference	Item Pairs	Average Difference
RR, ϕ	2.145	RR_{025}, ϕ_{025}	0.309
R, λ (EBGM)	2.015	RR_{025}, λ_{025}	0.285
ϕ, λ (EBGM)	0.381	ϕ_{025}, λ_{025}	0.219

The difference between an RR value and the corresponding ϕ value tends to be greater than the difference between the associated RR_{025} and ϕ_{025} values; on average the absolute difference between an RR value and the associated ϕ value is 2.145 while the absolute difference between the corresponding RR_{025} and ϕ_{025} values is 0.309. Let $\Delta_{(\mathrm{item1,item2})\mathrm{ave}}$ and $\Delta 025_{(\mathrm{item1,item2})\mathrm{ave}}$ denote the average of the difference |Item1−Item2|, where the items are ϕ, λ and RR, and ϕ_{025}, λ_{025} and RR_{025} respectively, so that $\Delta_{(\phi,\lambda)\mathrm{ave}}$ is the average of the difference $|\phi - \lambda|$ and $\Delta 025_{(\mathrm{RR},\lambda)\mathrm{ave}}$ is the average of the difference $|\mathrm{RR}_{025} - \lambda_{025}|$

and so on. Table 6.3 gives the average of the absolute difference between the possible item pairs for ϕ, λ and RR and also for ϕ_{025}, λ_{025} and RR_{025}. The values in the table show that $\Delta_{(\text{RR},\lambda)\text{ave}} > \Delta 025_{(\text{RR},\lambda)\text{ave}}$ and $\Delta_{(\phi,\lambda)\text{ave}} > \Delta 025_{(\phi,\lambda)\text{ave}}$. Thus the use of the lower bounds of the 95% credible intervals can be viewed as a way of moderating the values of RR and bridging the gap between the values of C-G, GPS and RR. This appears to explain why the models have more drug and adverse event combinations in common in their top 1000 selections when the lower bounds are used than when the point estimates are used. Further, the diversity between C-G and RR could be explained by the fact that $\Delta_{(\text{RR},\phi)\text{ave}}$ and $\Delta 025_{(\text{RR},\phi)\text{ave}}$ are the largest for the possible pairs of models for the point estimates and the lower bounds of 95% confidence/credible interval estimates respectively, as Table 6.3 shows.

The top ten drug and adverse event pairs uniquely chosen by RR, GPS, C-G and LogP in the top 1000 combinations selected by the point estimates of RR, λ (EBGM) and ϕ respectively, and the values of LogP are shown in Table G.3 of Appendix G. LogP is the negative of the logarithm of the Poisson probability of observing a count larger than or equal to that of a drug and adverse event pair. An approximate expression for LogP, given in Appendix G, was used whenever LogP grew large to avoid computational problems [19]. As the values in Table G.3 show, C-G places a premium on drug and adverse event pairs that have relatively higher observed and expected counts ($57 \leq N \leq 91$, $1.668 < E < 2.911$), when compared with those of GPS ($9 \leq N \leq 12$, $0.098 < E < 0.135$), which in turn are higher than those of the combinations uniquely selected by RR ($1 \leq N \leq 2$, $0.0002 < E < 0.0032$). Yet the observed and expected counts of the pairs favoured by C-G are comparatively lower than those favoured by LogP ($2429 \leq N \leq 7530$, $148.432 < E < 1003.651$). Incidentally, combinations uniquely chosen by the point estimate of RR have RR_{025} value of zero. Such pairs are not likely to generate a signal based on RR_{025} alone.

Table G.4 of Appendix G shows the top ten drug and adverse event pairs uniquely chosen by RR, GPS and C-G in the top 1000 combinations selected by the respective lower bounds RR_{025}, λ_{025} and ϕ_{025} of the 95% confidence/credible interval estimates. LogP was excluded from this comparison since it does not have such a thing as a lower confidence/credible bound. All the drug and adverse event combinations uniquely chosen by RR based on RR_{025} have an observed count of 4, with the expected counts lying between 0.0073 and 0.0202, which are larger than those of the combinations exclusively selected by RR based on the point estimate of RR. Again the combinations uniquely selected by

GPS have observed and expected counts ($7 \leq N \leq 22$, $0.100 < E < 0.624$) which are higher than those of the combinations uniquely selected by RR and smaller than those of the combinations uniquely selected by C-G ($134 \leq N \leq 5043$, $5.532 < E < 259.502$).

6.3.2 Other Characteristics of C-G

Model C-G, as in the case of GPS [19], appears not to be unduly swayed by very low observed and expected counts. For instance the pair Emla-Hypalgesia, presented in Table G.5 of Appendix G, has low observed and expected counts (1, 0.0009) with a high RR of 1114.14, but the corresponding point estimates of ϕ and λ (EBGM) are 1.80 and 2.37 respectively. While RR put the rank of the pair at 12, the rank by C-G is 63130 and that of GPS is 29826, indicating the ability of C-G and GPS to keep a 'proper' perspective on pairs with low observed and expected count values when the assessment is based on the point estimates. The behaviour of RR reverses when the three methods are assessed on the basis of RR_{025}, λ_{025} and ϕ_{025}, taking an objective view of pairs with weak data support; the rank (154785) it assigns to Emla-Hypalgesia is worse than when the point estimate of RR is used. The values of RR_{025}, λ_{025} and ϕ_{025} for the pair are 0.00, 0.94 and 0.23 and the ranks assigned by GPS and C-G are 55171 and 186757 respectively (see Table G.5). However, C-G appears to take a dimmer view of such pairs than GPS as it ranks the pair worse than GPS on both criteria – using the point estimate and the lower bound of the 95% credible interval estimate. The difference in behaviour between C-G and GPS is not unexpected as while they are both Bayesian, they differ fundamentally in their formulation; C-G being a fully Bayesian hierarchical model and GPS a Bayesian mixture model which appeals to the Empirical Bayes approach in the determination of the values of the prior parameters.

To further understand the behaviour of C-G relative to RR and GPS, we take a look at three other drug and adverse event pairs with combinations of the observed and expected counts different from the one examined above:

The drug and adverse event pair Copper-Uter Dis has a bigger observed count and a comparatively smaller expected count (4457, 46.398). The respective point estimates of RR, λ (EBGM) and ϕ are 96.06, 95.93, and 94.76, which are quite close. The corresponding ranks assigned by RR, GPS and C-G are 975, 155 and 30. Thus beyond being careful with pairs with very low observed and expected counts, the relative magnitudes of the observed and expected counts also matter more to GPS and C-G than to RR when it comes to

the ranking. In other words, even though ϕ and λ have lower point estimates for the pair relative to that of RR, these point estimates are higher relative to the values assigned by C-G and GPS respectively to other combinations and so Copper-Uter Dis is relatively better ranked by C-G and GPS than by RR when assessed on the basis of the point estimates, more so in the case of C-G. The relative importance, in terms of the ranks, attached to the pair by RR, GPS and C-G is maintained when the the methods are compared on the estimates for RR_{025}, λ_{025} and ϕ_{025}. The ranks are 151, 94 and 22 with the RR_{025}, λ_{025} and ϕ_{025} values being 93.26, 93.15 and 91.97 respectively.

The pair Vincristine-Purpura Thrombopen exemplify a situation where the expected count is bigger than the observed count (2, 11.479). The point estimates of RR, λ (EBGM) and ϕ are 0.17, 0.21 and 0.18 respectively. The ranks assigned by RR (373725), GPS (374503) and C-G (373694) are all large. The point estimate for ϕ and the corresponding rank assigned to Vincristine-Purpura Thrombopen by C-G, like those of RR and GPS, are 'reasonable' in the sense that if the observed count is considerably less than the expected count, under the assumption of independence of drug and adverse event, then there is probably no justification to suspect that the relationship between Purpura Thrombopen and Vincristine is causal. The rank (375525) assigned to Vincristine-Purpura Thrombopen based on the value of RR_{025} (0.00) is worse than that based on the point estimate of RR, and the ranks assigned by GPS (372101) and C-G (360887) based on the estimates for λ_{025} (0.07) and ϕ_{025} (0.04) are better than those based on the point estimates. However, it remains that the ranks assigned to the pair by the methods using both criteria are not meritorious and the relationship between Purpura Thrombopen and Vincristine is probably not causal.

We next consider a drug and adverse event pair for which both the observed and expected counts are large. The observed and expected counts of the pair Diltiazem-Rash are large (1697,1374.890). All three measures have the same point estimate of 1.23. This observation – the assignment of almost the same value by the three measures – is true for most pairs for which the observed count is large, more so when the observed and expected counts are both large. Some examples of such pairs are presented in Table G.5 of Appendix G. The observation suggest that three approaches tend to converge in behaviour when the observed and expected counts are large, in terms of the point estimates they assign to the drug and event pairs. As RR places a premium on pairs with low observed and expected counts, it ranks the pair 189850, worse than that of GPS (99124) and C-G (140633). The

values of RR_{025} (1.18), λ_{025} (1.18) and ϕ_{025} (1.18) are also the same, lending credence to the above observation that three approaches tend to converge in behaviour when the observed and expected counts are large. The associated ranks 33326, 41473 and 34513 respectively are better than in the case of the use of the point estimates.

6.3.3 Genuine Drug Problems Within the Top Fifty Drug and Adverse Event Combinations Selected by C-G, GPS and RR

Of the first fifty drug and adverse event combinations selected by C-G based on the posterior mean of ϕ, shown in Tables G.6a and G.6b, about 41 of the adverse events involved were found to be bona fide problems associated with the drug in the combination when checked with the online version of Physicians' Desk Reference (PDR) [62] and similar resources [18, 75]. The corresponding values for GPS is 40 and that of RR is 36 (the first fifty drug and adverse event combinations selected by GPS and RR are shown in Tables G.7a and G.7b; and G.9a and G.9b respectively). The number of drug and adverse event pairs, for which the adverse event was found to be a genuine problem associated with the drug, selected by C-G, GPS and RR within the top fifty combinations based on the lower bounds of the 95% confidence/credible interval estimates are 45, 43, and 40 respectively (see Tables C.2a and C.2a; G.8a and G.8b; and G.10a and G.10b respectively). The remaining drug and adverse event combinations could either not be verified for lack of information or the available information did not confirm the adverse event as a problem associated with the drug. The verification involved checking the adverse event in a drug and adverse event combination against the adverse events and risks cited for the drug in the aforementioned drug databases. Perhaps the higher number of combinations for which the adverse event is a bona fide problem associated with the drug selected by C-G, GPS and RR within the top fifty drug and adverse event combinations, when the lower bounds of the 95% confidence/credible interval estimates are used, is an indication that the methods are more able to give preference to genuine drug problems in the ranking of the drug and adverse event pairs when the ranking is done by the lower bounds of the 95% confidence/credible interval estimates than by the point estimates.

6.3.4 C-G values compared with that of GPS

Figure 6.1 shows plots of ϕ (C-G) against λ (EBGM, GPS) for various combinations of the observed (N) and expected (E) counts and Figure 6.2 shows the plots of ϕ_{025} against

λ_{025} for the same combinations of the observed (N) and expected (E) counts as in Figure 6.1. The figures show that the values of ϕ and ϕ_{025} are lower than or equal to those of λ (EBGM) and λ_{025} respectively for all combinations of the observed and expected counts (as all the points in the figures are on the line of equality or below it); ϕ and ϕ_{025} tend to be equal to λ (EBGM) and λ_{025} respectively at higher values of the observed and expected counts and tend to be lower than λ (EBGM) and λ_{025} respectively at lower values of the observed and expected counts.

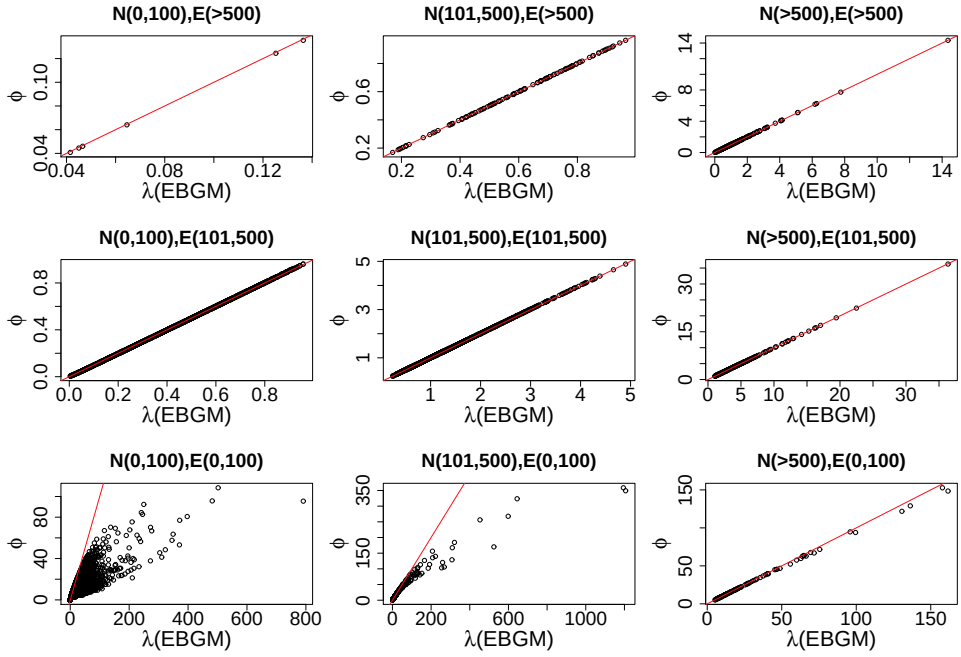


Figure 6.1: Plots of ϕ against λ (EBGM) for various combinations of the observed and expected counts. The red line is the line of equality.

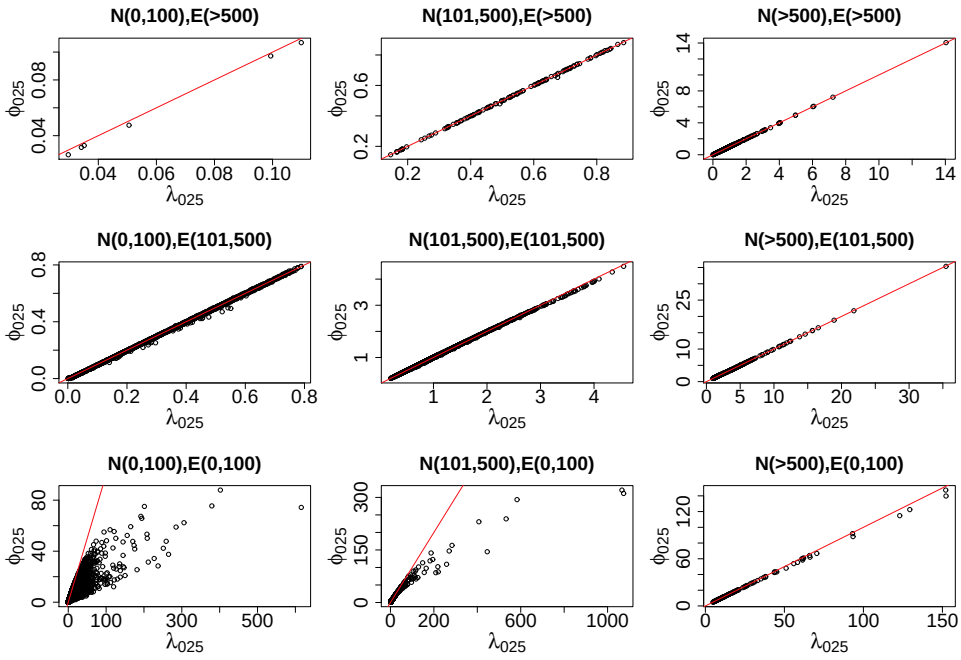


Figure 6.2: Plots of ϕ_{025} against λ_{025} for various combinations of the observed (N) and expected (E) counts. The red line is the line of equality.

Chapter 7

Conclusion

7.1 Concluding Remarks

The need to find unknown but bona fide side-effects of drugs as soon as possible after marketing approval can hardly be overemphasized given the health and economic burden they present, as clinical trials are unable to detect all side-effects of drugs before marketing for reasons outlined in Chapter 1. Spontaneous Reporting System data has been found to be useful in addressing this need [6, 54, 81].

It is imperative from the literature and the results presented in Chapter 2 to continue to contribute to the evolution of spontaneous reporting system databases to the point where extraneous factors that act against their optimal use are curtailed and to develop complementary tools to facilitate the realization of its full potential.

The increasing trend in the number of reports submitted per 10,000 people, as observed in Chapter 2, may be pointing to:

i. a growing awareness of the need to contribute to the pharmacovigilance process and hence the increase in the number of reports, when viewed against the well known phenomenon of under reporting [4, 70], or

ii. the number of adverse events is growing at an increasing rate, or

iii. both cases (i) and (ii) hold.

If the the first scenario holds, then there is the need to take action to sustain the trend to the point where almost all adverse events are reported. In the second scenario, more must be done to curtail avoidable adverse events and to sustain the hunt for unknown

side-effects so that they can be found as soon as practicable. This is all the more urgent when one considers the fact that expedited reports – which are related to serious adverse events – have been in the majority since 2005 (Section 2.2.4). Other areas of concern have been identified in Section 7.2 below.

It follows from the analysis presented in Chapter 5 that C-G is the model suitable for modeling Data 1 and none of the models is suitable for modeling Data 2 and Data 3 as stated in Section 6.1, where it was pointed out that Data 2 and Data 3 are essentially truncated versions of Data 1 while the models fitted to them are not for truncated data. Hence the misfit between the models and Data 2 and Data 3. In arriving at the above conclusions, we took into consideration the Bayesian p-values for the models (Section 5.3.2). Based on variance (X_1, X_2) and deviance (OD, SD) criteria (Section 5.3.2), model C-G was the only model whose Bayesian p-values are within the recommended range of (0.025, 0.975) [10, 80] for Data 1, the rest of the models have either both p-values lying outside the recommended interval or one of the values is within the range and the other is outside. For Data 2 and Data 3, in the case of C-G and C-IG, Bayesian p-values using (X_1, X_2) are all outside the recommended range, thereby ruling out C-G and C-IG as suitable models for Data 2 and Data 3.

The behaviour of C-G with respect to the ranking of the drug and adverse event combinations was found to be consistent than that of RR as the (RK3, RK33) correlation coefficient (0.79)(Section 5.4.3), which compares the ranking based on ϕ with that based on ϕ_{025}, is higher than the (RK1, RK11) correlation coefficient (0.14), which compares the ranking based on RR with that based on RR_{025}. However, GPS is more consistent in the ranking of the drug and adverse event combinations than C-G; its (RK2, RK22) correlation coefficient, which which compares the ranking based on λ (EBGM) with that based on λ_{025}, is 0.93, which is higher than the (RK3, RK33) correlation coefficient of 0.79.

The number of drug and adverse event combinations that is common to the top 1000 drug and adverse event combinations selected by the C-G, GPS and RR based on the lower bounds ϕ_{025}, λ_{025} and RR_{025} respectively of the 95% confidence/credible interval is 662 (Section 6.3.1). This figure is more than three times the number (198) common to the top 1000 combinations selected by the models when the point estimates of ϕ, λ (EBGM) and RR are used. Thus the models are more similar in the way they select the top 1000 combinations when the selection is based on the lower bounds ϕ_{025}, λ_{025} and RR_{025} of the 95% confidence/credible interval estimates than when the selection is

based on point estimates. Again, the results of Section 6.3.1 shows that as far as the selection of the top 1000 combinations are concerned C-G is closer to GPS than to RR. The above observations are supported by the fact that on average the difference between corresponding point estimates assigned any two of C-G, GPS and RR are bigger than the difference between their associated lower bounds of the 95% confidence/credible interval estimates. In particular, not only is $\Delta_{(\mathrm{RR},\phi)\mathrm{ave}} > \Delta 025_{(\mathrm{RR},\phi)\mathrm{ave}}$, both $\Delta_{(\mathrm{RR},\phi)\mathrm{ave}}$ and $\Delta 025_{(\mathrm{RR},\phi)\mathrm{ave}}$ are the largest as far as the average difference between the point estimates and the average difference between the lower bounds of the 95% confidence/credible interval estimates assigned by any two of C-G, GPS and RR respectively, are concerned.

The number (45) of drug and adverse event combinations in which the adverse event is a bona fide problem associated with the drug, selected by C-G within the top fifty combinations, based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ is higher than number (41) selected based on the posterior mean of ϕ (Section 6.3.3). This appear to suggest that C-G is more able to give preference to genuine drug problems in the ranking of the drug and adverse event pairs when the ranking is based on the lower bound (λ_{025}) of the 95% credible interval estimate of ϕ than when the ranking is based on the point estimate of ϕ. Indeed this observation is also true for GPS and RR; GPS selects 43 when the values of λ_{025} are used and 40 when the values of λ (EBGM) are used, while RR selects 40 when the values of RR_{025} are used and 36 when the values of RR are used.

As demonstrated in Chapter 6, model C-G could be helpful in sifting SRS databases for signal generation and prioritizing. The examples provided in Section 6.3.2, show how versatile C-G is;

(a) it keeps a 'proper' perspective on drug and adverse event combinations with very low observed and expected counts when ranking the combinations, and

(b) takes a balanced view of the drug and adverse event pairs by assigning:

   i. a higher ϕ value when the observed count is relatively bigger than the expected count,

   ii. a low ϕ value when the observed count is smaller than the expected count or when observed count is close the expected count.

We conclude that model C-G has desirable attributes the stands it in good stead to contribute positively to pharmacovigilance. Further work with a complete and more recent

data is required to establish the most useful cut-off point or what should constitute the minimum threshold for generating a signal. In the interim, the condition $\phi_{025} \geq 2$ is recommended as the minimum threshold for signal generation. This recommendation ensures that a signal is generated only when the rate of reporting of a drug and adverse event combination is at least twice as much as expected under the assumption of independence of drug and adverse event, so as to keep false positives to a minimum. As seen in Section 6.3.4, the values of ϕ and ϕ_{025} are lower than or equal to those of λ (EBGM) and λ_{025} respectively and for all combinations of the observed and expected counts and therefore C-G could help further reduce the level of false positives when well calibrated.

Model C-G has high computational requirements. It takes about 16 days to run C-G in its current Fortran implementation on an Intel Xeon E class processor workstation, running the CentOS 5.4 X86_64 operating system with 32Gb RAM and average speed of 3.0Ghz. Additional RAM would be needed for data with size larger than that of Data 1. The high computing time is due to the fact that the evaluation of the function $z(\rho^{\nu}, \nu)$ requires a lot of time. The amount of time required to evaluate $z(\rho^{\nu}, \nu)$ is appreciated when one considers the fact that P-G which does not involve $z(\rho^{\nu}, \nu)$ takes about 21 hours to run, which is approximately 5.46 percent of the time required to run C-G. As the computational time of any model impinges on its usefulness, research is needed to reduce the time required to evaluate $z(\rho^{\nu}, \nu)$ and hence the implementation time of C-G.

7.2 Highlights of the Research

The research presented an opportunity to gain at first hand an appreciation of:

- the degree of the problem of missing values (as seen in Chapter 2) and to contemplate the level of effort required to sensitize stake holders about the problem and the need to, as much as possible, supply/fill all the required information/fields when reporting an adverse event.

- the need for accurate reporting, especially the use of drug names and preferred terms. It was evident from the preliminary analysis that some of the preferred terms required to be used in describing adverse events were not properly used. For instance some of the terms used to describe the adverse events include: 'Chemotherapy', 'exposure' and 'drug level'. Clearly the descriptions of the adverse events in these instances were inaccurate; the required preferred terms were either not properly used or were

not applied at all.

- what it takes to transform raw data from an SRS database into forms suitable for analysis. In particular, what is involved in converting the raw data into one of drug and adverse event combinations for signal detection, either using only the primary suspect drug or including the secondary suspect and the other drugs.

- how collaborative and interdisciplinary the field of pharmacovigilance is. It requires a team of people of diverse but relevant backgrounds (statisticians, pharmacologists, computer savvy professionals et cetera) to be able to competently assess circumstances relating to a drug and adverse event combination and successfully carry out a pharmacovigilance exercise, at least from the perspective of using SRS and related data.

- the extent and diversity of computer resources required to carry out a pharmacovigilance exercise.

- the fact that small differences in the attributes of models can mean a lot in terms of the difference in their performance.

7.3 Future Work

A number of areas remain to be explored in furtherance of this work. They include:

- Exploring the possibility of speeding up the computation of $z(\rho^{\nu}, \nu)$ so that the model could be implemented within the shortest possible time.

- Exploring the possibility of using a more efficient updating scheme for the parameters. In particular a joint update for α and β or φ and ψ to further reduce the autocorrelation in the iterates.

- Extension of the C-G model into the multiple comparison context as presented in Müller et al. [56] and exemplified by Ahmed et al. [2], to further consolidate model C-G's potential for reducing false positive signals.

- Exploring the potential of C-G to signal 'higher order dependencies' [58] in an SRS database. A higher order dependency refers to a situation where other causal factor(s) beside a given drug are involved in the observed adverse event [58, 82]. An adverse

event resulting from drug-drug interaction is a situation of a higher order dependency ('multi-item association' [82]). Here more than two entities (three) are involved: the two drugs and the adverse event.

Appendix A

Selected Variables and their Description

Table A.1a: Selected variables and their description.

Variable	**Description**
ISR	Primary link variable identifying each individual report received, across all seven data files.
CASE	Unique number identifying a given subject (record).
FDA_DT	Date report was received by the FDA.
REPT_COD	Holds codes identifying type of report made. Code — Meaning EXP — Expedited(within 15 days) PER — Periodic DIR — Direct
AGE	The age of a subject.
AGE_COD	Holds codes for unit in which age was measured. Code — Meaning DEC — Decade YR — Year MON — Month WK — Week DY — Day HR — Hour

Source: ASC_NTS.DOC, [29].

Table A.1b: Selected variables and their description.

Variable	**Description**
GNDR_COD	Holds codes identifying the gender of a subject. Code — Meaning UNK — Unknown M — Male F — Female NS — Not Specified
OCCP_COD	Holds codes for the occupation of the original reporter. Code — Meaning MD — Physician PH — Pharmacist OT — Other Health Professional LW — Lawyer CN — Consumer
E_SUB	Holds codes for mode of submission of report. Yes for electronic submission and No otherwise.
ROLE_COD	Holds codes identifying the reported role of drug in event. Code — Meaning PS — Principal Suspect SS — Secondary Suspect C — Concomitant I — Interacting
DRUGNAME	Holds names of drugs involved in the event, either the "Valid Trade Name" or the "Verbatim" name as stated on report.
VAL_VBM	Holds codes showing whether name of drug is the Valid Trade Name or the Verbatim name. Code — Meaning 1 — Valid Trade Name 2 — Verbatim Name

Source: ASC_NTS.DOC, [29].

Table A.1c: Selected variables and their description.

Variable	**Description**	
PT	Holds the Preferred Terms (PT) for describing the observed reaction to drug(s) using the Medical Dictionary for Regulatory Activities (MedDRA).	
OUTC_COD	Holds codes identifying the outcome of the adverse experience.	
	Code	Meaning
	DE	Death
	LT	Life-Threatening
	HO	Hospitalization – Initial or Prolonged
	DS	Disability
	CA	Congenital Anomaly
	RI	Required Intervention to prevent Permanent Damage
	OT	Other
RPSR_COD	Holds codes identifying the initial report source.	
	Code	Meaning
	FGN	Foreign
	SDY	Study
	LIT	Literature
	CSM	Consumer
	HP	Health Professional
	UF	User Facility
	CR	Company Representative
	OTH	Other

Source: ASC_NTS.DOC, [29].

Appendix B

Some Selected Plots and Tables

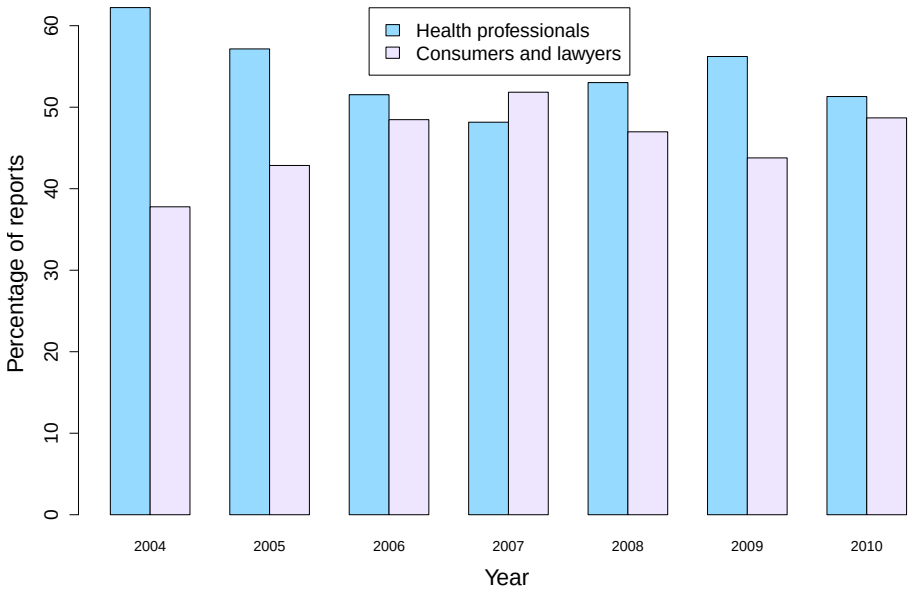


Figure B.1: Percentage of reports from health professionals and consumers and lawyers for 2004-2010.

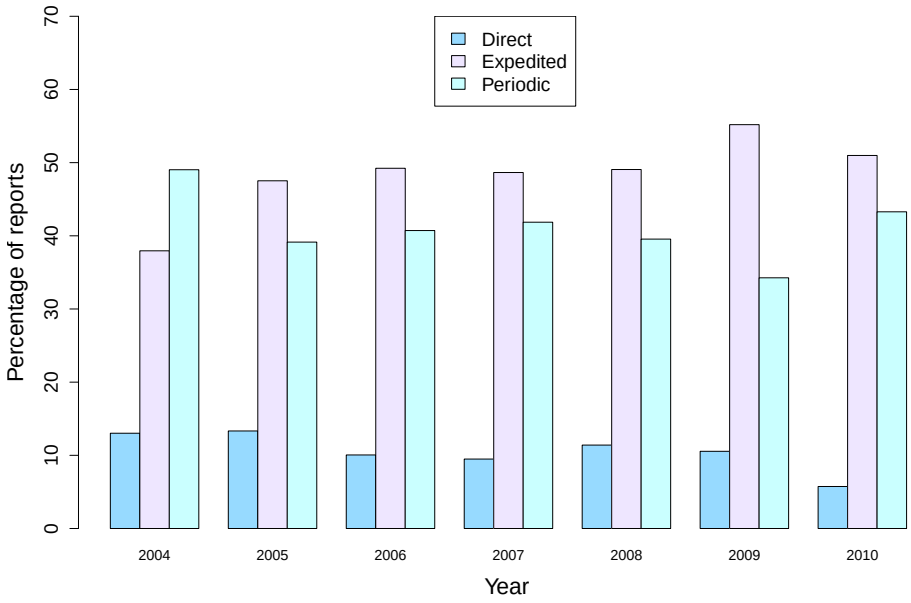


Figure B.2: Percentage of report types from 2004-2010.

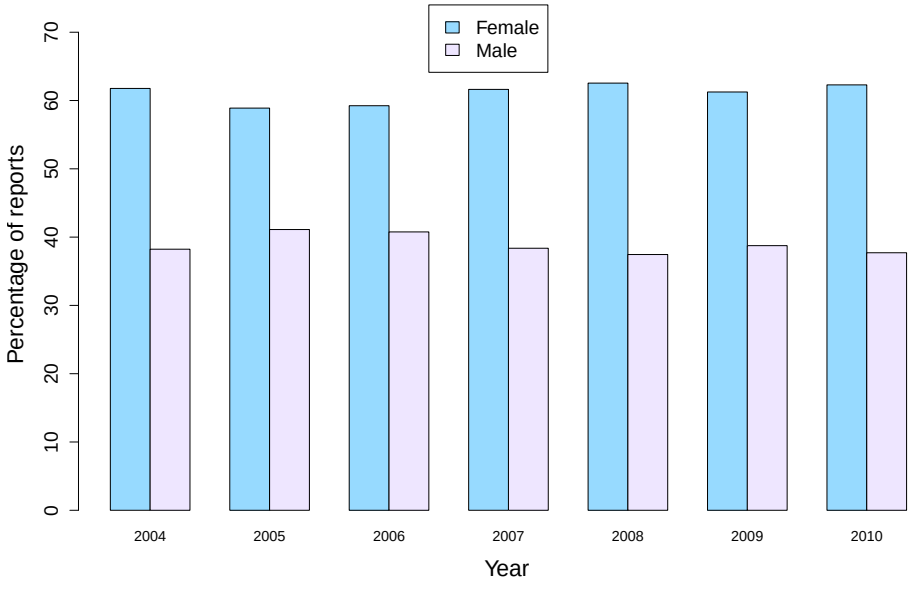


Figure B.3: Percentage of reports on male and female subjects from 2004-2010.

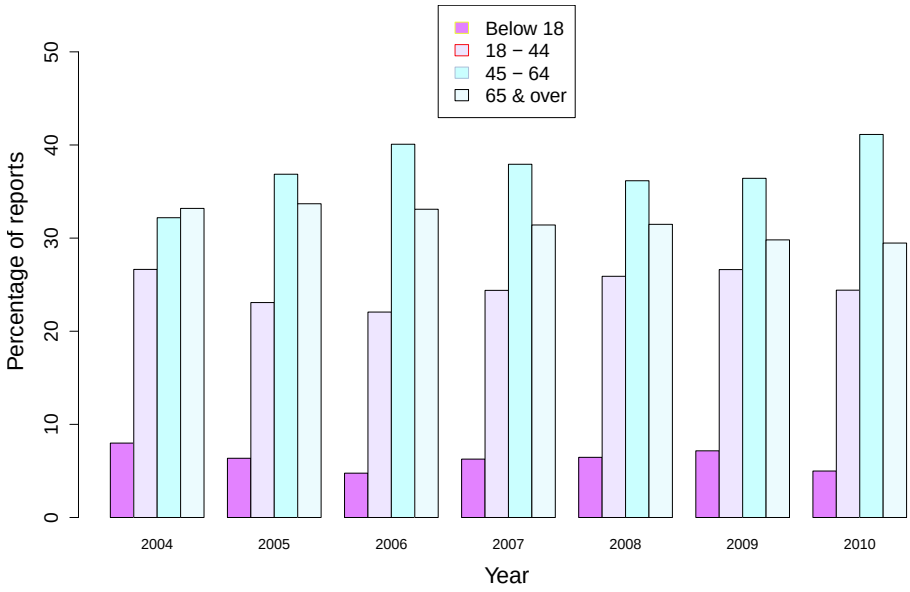


Figure B.4: Percentage of reports for the various age groups for the period 2004-2010.

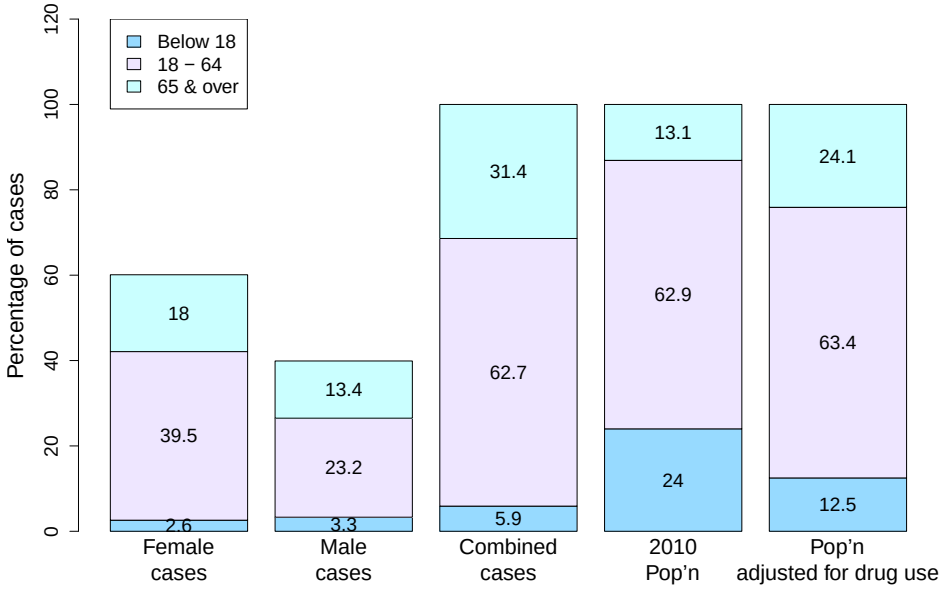


Figure B.5: Age and gender load of reported adverse events associated with drug use.

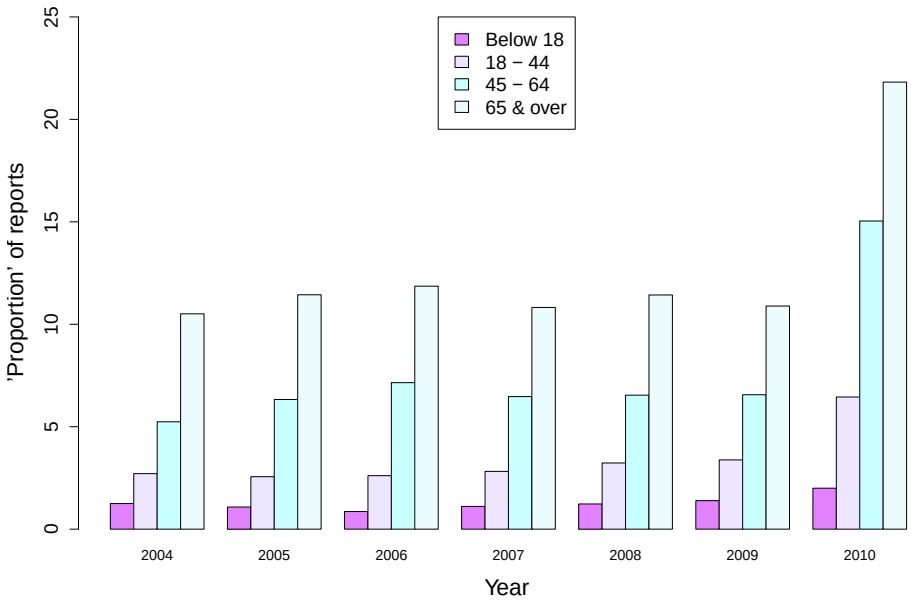


Figure B.6: 'Proportion' of the various age groups reported on for the period 2004 – 2010.

Table B.1: Percentages for Patient Outcomes calculated with number of all cases as denominator, 2004-2010.

Year	DE	LT	HO	DS	CA	RI	OT
2004	9.68	5.12	25.71	3.73	0.33	4.00	28.93
2005	9.95	4.97	28.57	4.73	0.38	3.24	36.43
2006	8.59	4.13	30.49	4.96	0.39	2.66	36.94
2007	7.66	4.20	26.95	3.17	0.44	1.58	37.21
2008	9.27	4.43	25.63	2.67	0.68	1.90	33.56
2009	10.63	4.27	28.02	2.76	0.53	1.91	36.17
2010	10.83	3.17	24.83	2.27	0.54	0.96	34.47
Composite	9.73	4.13	26.81	3.19	0.49	2.01	34.95

Table B.2: Percentages for Patient Outcomes calculated with number of non-missing cases as denominator, 2004-2010.

Year	DE	LT	HO	DS	CA	RI	OT
2004	15.45	8.16	41.00	5.95	0.53	6.37	46.14
2005	14.09	7.03	40.46	6.70	0.53	4.59	51.58
2006	12.24	5.89	43.44	7.07	0.55	3.79	52.64
2007	11.68	6.41	41.11	4.84	0.67	2.41	56.77
2008	14.61	6.98	40.40	4.20	1.07	2.99	52.89
2009	15.61	6.26	41.14	4.05	0.77	2.81	53.11
2010	17.33	5.08	39.74	3.63	0.86	1.54	55.17
Composite	14.83	6.28	40.87	4.85	0.75	3.07	53.26

Appendix C

C-G Model Results

C.1 Data 1

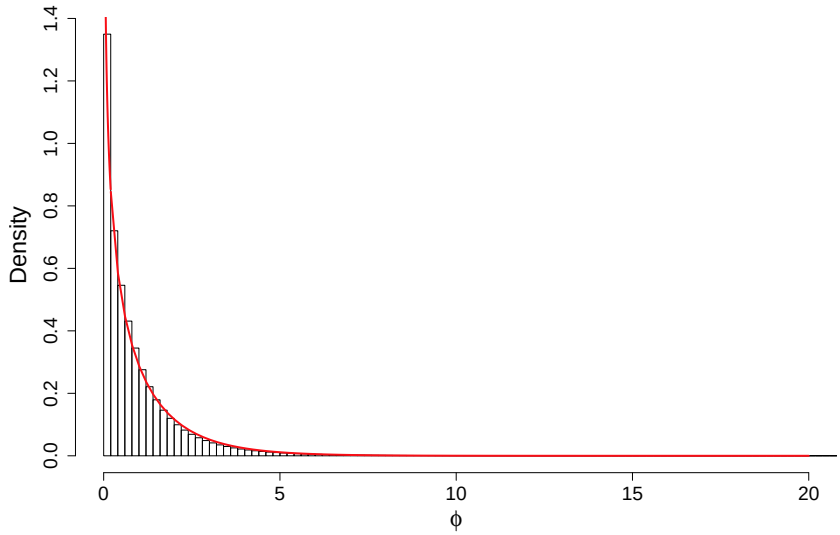


Figure C.1: Histogram of ϕs for C-G and Data 1. The superimposed density curve (in red) represents the Gamma distribution corresponding to the MCMC estimates of the posterior means of α and β.

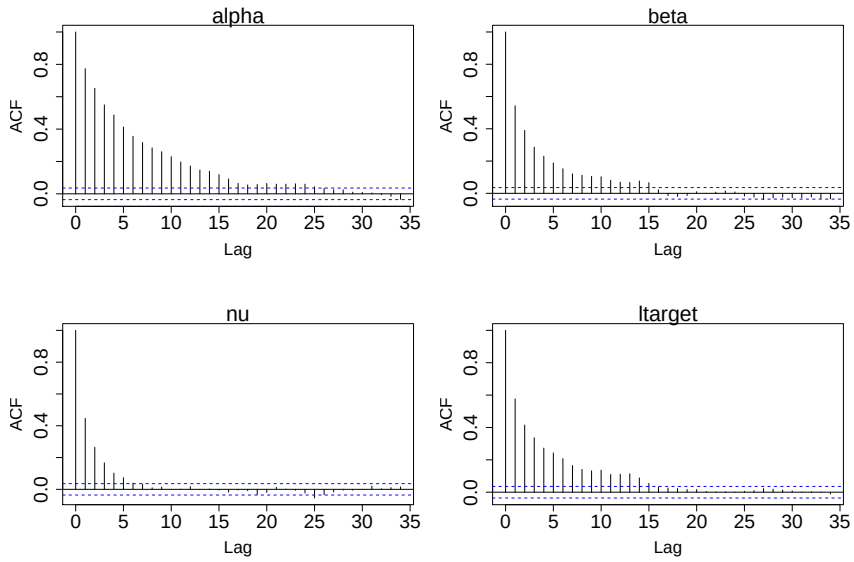


Figure C.2: Acf plots of α, β, ν and the logarithm of the target distribution for C-G and Data 1.

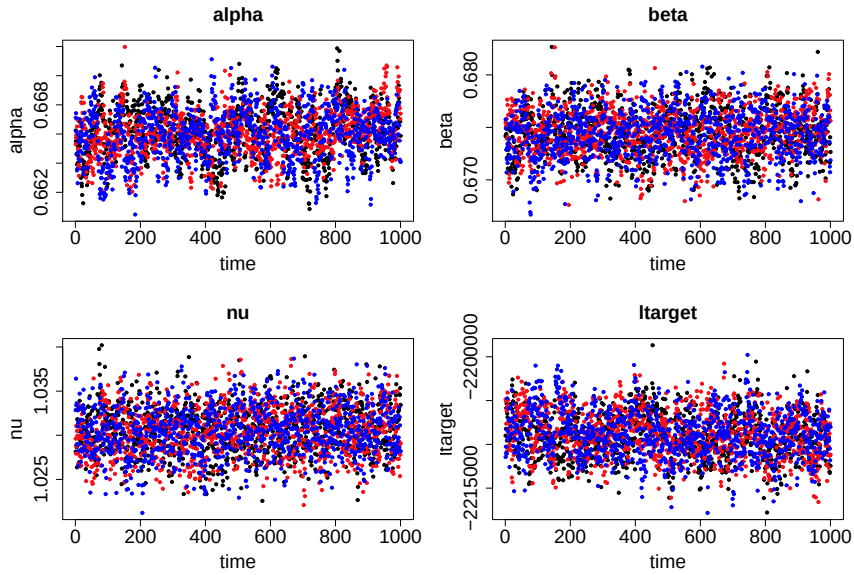
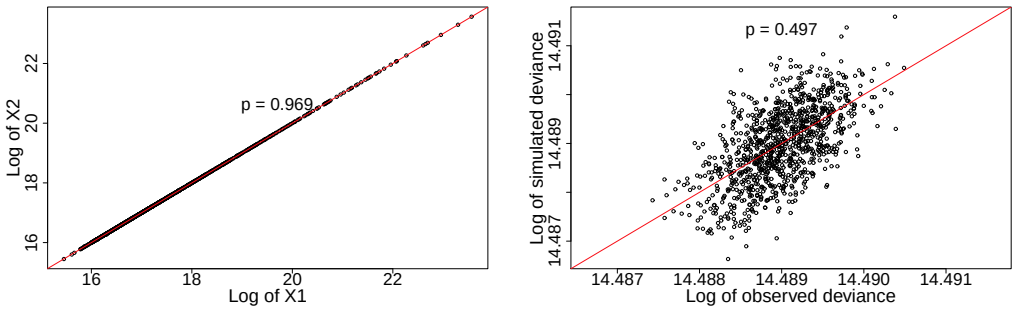
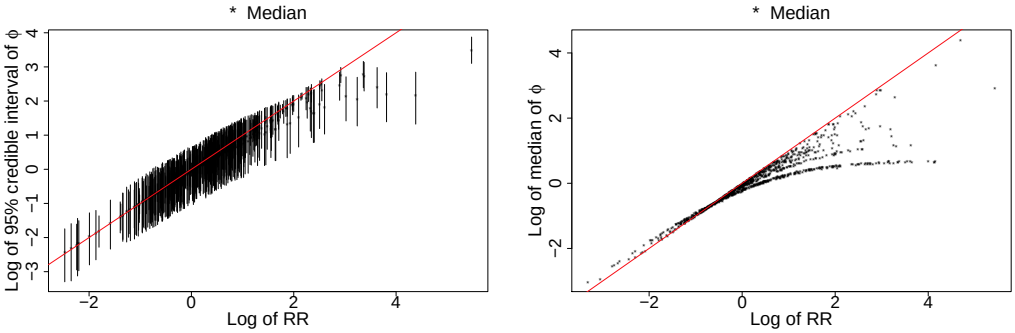


Figure C.3: Trace plots of α, β, ν and the logarithm of the target distribution for three chains, for C-G and Data 1.



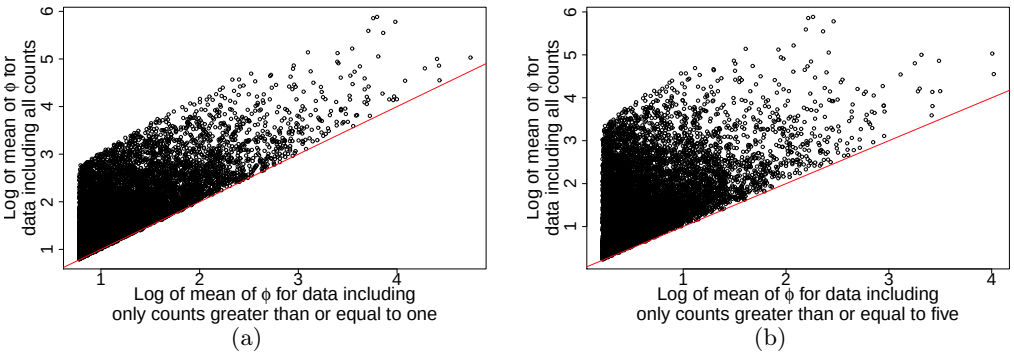
(a) (b)

Figure C.4: Bayesian p-value scatter plots for C-G and Data 1. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.



(a) (b)

Figure C.5: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for the C-G model. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 1, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for C-G model. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 1.



(a) (b)

Figure C.6: (a) Scatter plot of logarithm of posterior means of ϕ for Data 1 against corresponding values for Data 2, and (b) Scatter plot of logarithm of posterior means of ϕ for Data 1 against corresponding values for Data 3. The plots were made with the top 10000 corresponding values from the two data sets under consideration.

Table C.1: Values of original counts, mean replicate counts, expected counts, RR and ϕ for fifty randomly selected drug and side-effect pairs using C-G model and Data 1.

Count	Mean Rep. Count	Expected Count	RR	ϕ
7	7.2	11.218	0.62	0.64
6	5.9	5.973	1.00	1.01
0	0.0	0.004	0.00	0.98
666	665.6	695.539	0.96	0.96
0	0.4	0.915	0.00	0.43
44	40.1	6.070	7.25	6.61
0	0.0	0.039	0.00	0.92
111	58.3	0.721	154.02	81.10
0	0.0	0.030	0.00	0.92
28	1.3	0.031	889.14	41.99
32	2.1	0.047	685.52	46.34
7	7.0	6.866	1.02	1.02
1	1.1	1.245	0.80	0.86
1	1.0	0.997	1.00	1.00
1	0.9	0.765	1.31	1.16
1	0.6	0.369	2.71	1.71
1	0.9	0.80	1.22	1.12
0	0.2	0.339	0.00	0.64
26	1.4	0.037	707.23	38.47
7	0.2	0.018	399.26	11.33
6	5.8	4.814	1.25	1.21
2022	2012.9	132.310	15.28	15.21
6	0.1	0.010	576.05	10.20
1021	1022.0	1701.710	0.60	0.60
637	636.3	606.791	1.05	1.05
1211	1210.3	739.652	1.64	1.64
933	933.3	854.123	1.09	1.09
1039	1037.8	380.965	2.73	2.73
56	10.3	0.147	381.61	70.79
1	0.6	0.409	2.44	1.57
46	4.6	0.073	628.21	64.17
2	1.9	1.625	1.23	1.16
3664	3655.6	325.306	11.26	11.24
1608	1608.4	1421.430	1.13	1.13
82	11.4	0.105	780.57	109.03
1	0.4	0.206	4.85	1.91
1032	1031.8	519.620	1.99	1.98
594	593.5	392.751	1.51	1.51
3	2.9	2.468	1.22	1.16
2	1.5	0.845	2.37	1.77
66	65.6	55.168	1.20	1.19
0	0.5	2.436	0.00	0.20
19	19.7	93.931	0.20	0.21
306	71.0	0.198	1543.39	359.07
306	71.0	0.198	1543.39	359.07
6	0.0	0.004	1577.25	10.09
10	0.4	0.023	431.67	15.66
1	0.4	0.198	5.05	1.96
3	2.6	1.432	2.09	1.78
5	5.1	4.914	1.02	1.03

Table C.2a: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	358.53	320.87	1	1
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	349.02	311.32	2	2
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	323.51	293.66	3	3
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	267.66	238.76	4	4
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	256.19	230.43	5	5
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	184.3	163.13	6	6
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	152.79	147.32	10	7
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	167.41	147.07	8	8
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	170.1	144.86	7	9
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	156.35	141.35	9	10
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	148.51	139.79	11	11
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	139.78	122.77	12	12
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	510	95	129.01	122.5	14	13
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	135.44	118.88	13	14
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	552	104	121.83	114.78	17	15
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	126.83	113.36	16	16
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	128.91	109.11	15	17
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	120.37	102.14	18	18
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	437	94	113.33	101.49	20	19
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	114.11	97.77	19	20
SELENIUM SULFIDE	HAIR DISCOLOR	301	2.236	134.64	119.43	526	107	104.37	93.05	23	21
COPPER	UTER DIS	4457	46.398	96.06	93.26	975	151	94.76	91.97	30	22
PHENFORMIN	ACIDOSIS LACTIC	390	3.302	118.11	106.61	681	124	98.53	89.48	27	23
OPCON A	PAIN EYE	267	2.015	132.49	116.61	547	109	100.08	88.99	26	24
DTP VACCINE	SCREAMING SYND	966	9.643	100.18	93.95	897	147	93.9	88.19	31	25
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	108.42	87.88	21	26
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	105.76	87.46	22	27
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	101.96	85.48	25	28
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	103.26	85.04	24	29
POLIOVIRUS VACCINE, LIVE, ORAL	SCREAMING SYND	344	3.28	104.89	93.91	831	148	87.35	78.75	34	30
TETRAHYDROZOLINE	PAIN EYE	187	1.43	130.77	112.59	557	117	89.85	77.83	33	31
LOMEFLOXACIN HYDROCHLORIDE	PHOTOSENSITIVITY	481	5.137	93.64	85.27	1023	166	83.08	76.13	38	32
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	95.73	75.47	28	33
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	92.35	75.08	32	34

Table C.2b: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	95.5	74.3	29	35
POLIOVIRUS VACCINE, LIVE, ORAL	SIDS	146	1.038	140.6	118.45	484	108	86.33	73.53	35	36
MITOMYCIN	UREMIA	218	1.995	109.29	95.26	785	142	82.28	71.86	42	37
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM INTRACRAN	476	5.426	87.73	79.99	1150	184	78.28	71.37	46	38
CHOLINE	KERATITIS	116	0.703	164.99	135.12	372	91	85.71	71.27	36	39
MENOTROPINS	OVAR DIS	151	1.175	128.48	108.06	583	122	82.37	70	41	40
CEFOXITIN	ENTEROCOL PSEUDOMEM	138	1.016	135.79	113.16	517	116	82.86	69.82	39	41
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	84.46	67.31	37	42
BENDECTIN	ECTROMELIA	111	0.721	154.02	126.27	415	100	80.76	67.11	43	43
UROKINASE	CHILLS	831	10.942	75.95	70.83	1426	220	71.64	66.8	51	44
DIPHTHERIA-TETANUS TOXOID-PERTUSSIS VA	SCREAMING SYND	147	1.22	120.47	101.62	654	130	78.45	66.58	45	45
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	82.7	65.74	40	46
TETRAHYDROZOLINE	CONJUNCTIVITIS	230	2.454	93.73	81.91	1021	174	74.12	65.21	50	47
HAEMOPHILUS B POLYSACCHARIDE VACCINE	MENINGITIS	176	1.711	102.86	87.66	851	164	74.45	64.4	49	48
COPPER	SALPINGITIS	1484	21.323	69.6	66.08	1635	238	67.56	64.19	58	49
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	80.63	62.31	44	50
SUPROFEN	PAIN BACK	399	5.188	76.91	69.39	1404	227	68.47	62.01	56	51
NONOXYNOL	LEUKORRHEA	603	8.344	72.27	66.51	1516	235	66.97	61.98	60	52
MICONAZOLE	VAGINITIS	2262	34.851	64.9	62.24	1818	271	63.67	61.06	64	53
COPPER	DEVICE MIGRATION	1490	22.868	65.16	61.88	1811	272	63.42	60.23	67	54
COPPER	PAIN PELVIC	1962	30.696	63.92	61.12	1861	278	62.59	59.92	73	55
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	75.42	59.2	48	56
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	77	58.9	47	57
INTRAUTERINE DEVICE	UTER DIS	251	3.127	80.26	70.35	1328	224	66.38	58.59	62	58
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM CEREBR	695	10.419	66.71	61.81	1744	274	62.8	58.33	71	59
PILOCARPINE	MIOSIS	103	0.812	126.83	103.44	601	128	70.24	57.61	53	60
TRIAMCINOLONE	ATROPHY INJECT SITE	439	6.303	69.65	63.14	1630	260	63.17	57.54	70	61
GONADOTROPIN, CHORIONIC	OVAR DIS	86	0.567	151.69	119.94	423	106	70.3	56.8	52	62
TIOCONAZOLE	VAGINITIS	268	3.558	75.32	66.33	1442	236	63.66	56.5	65	63
FLUNISOLIDE	NASAL SEPTUM DIS	75	0.431	174.04	136.91	344	90	69.25	55.41	55	64
COPPER	FERTIL DEC FEM	695	10.997	63.2	58.56	1899	294	59.69	55.35	79	65
OXYTOCIN	FETAL DIS	95	0.759	125.14	100.11	616	133	67.04	55.01	59	66
ALCOHOL	ALCOHOL INTOLER	169	2.006	84.26	71.8	1233	215	63.4	54.27	68	67

Table C.2c: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
MICONAZOLE	VULVOVAGINITIS	234	3.168	73.87	64.71	1471	248	61.35	54	76	68
TRAZODONE	PRIAPISM	347	5.117	67.81	60.77	1701	282	60.08	53.96	78	69
NONOXYNOL	DEVICE MIGRATION	800	13.275	60.27	56.12	2045	311	57.55	53.85	85	70
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	70.17	53.31	54	71
RUBELLA VIRUS VACCINE, LIVE	LYMPHADENO	115	1.162	98.97	80.9	925	179	63.36	52.78	69	72
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	67.7	52.33	57	73
OXYTOCIN	LABOR ABNORM	105	1.043	100.68	81.5	887	176	61.99	51.12	75	74
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	66.71	50.1	61	75
DORNASE ALFA	HEMOPTYSIS	64	0.336	190.37	145.75	301	83	64.67	50.08	63	76
FLOSEQUINAN	DEATH	109	1.204	90.5	73.89	1092	203	58.74	48.6	82	77
TICE BCG	CYSTITIS	69	0.475	145.35	111.64	457	119	61.19	48.18	77	78
NAFARELIN ACETATE	HEM VAGINAL	506	9.008	56.17	51.29	2260	349	52.29	48.02	95	79
ATRACURIUM BESYLATE	PARALYSIS FLACCID	100	1.051	95.13	77.06	988	193	58.68	47.84	83	80
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	63.54	47.75	66	81
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	62.6	47.67	72	82
PRILOCAINE	PARESTH CIRCUMORAL	200	2.995	66.77	57.75	1742	304	54.87	47.64	91	83
BUTORPHANOL	DRUG DEPEND	457	8.165	55.97	50.95	2270	353	51.85	47.23	96	84
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	62.3	47.23	74	85
OXYTETRACYCLINE	DISCOLOR TOOTH	74	0.599	123.49	96.79	633	138	58.99	46.82	81	86
SOMATROPIN	OTITIS MED	101	1.147	88.07	71.5	1143	218	56.29	46.25	87	87
INSULIN NOVOLIN 70/30	HYPERGLYCEM	484	8.959	54.02	49.22	2408	370	50.43	46.22	108	88
PROCAINAMIDE	LE SYND	471	8.679	54.27	49.43	2386	367	50.53	46.01	107	89
ISOTRETINOIN	BLIND NIGHT	140	1.94	72.15	60.3	1524	286	54.11	45.91	92	90
FOSCARNET SODIUM	HYPOCALCEM	99	1.149	86.18	69.64	1184	226	55.04	45.11	90	91
COPPER	UTER RUPT	270	4.665	57.88	51.02	2170	352	50.76	45.09	104	92
RITONAVIR	PARESTH CIRCUMORAL	217	3.566	60.84	52.99	2005	332	51.41	45.01	98	93
ISOTRETINOIN	CHEILITIS	380	6.983	54.42	48.97	2372	374	49.73	44.96	115	94
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	59.21	44.52	80	95
PHOTOPLEX	PHOTOSENSITIVITY	156	2.359	66.13	55.96	1766	314	51.76	44.3	97	96
INFLUENZA VIRUS VACCINE	GUILLAIN BARRE SYND	62	0.451	137.47	104.21	507	127	56.07	43.85	89	97
CLOMIPHENE	OVAR DIS	163	2.567	63.5	53.76	1882	326	50.64	43.6	106	98
DTP VACCINE	SIDS	207	3.495	59.23	51.21	2095	350	49.86	43.47	113	99
CEFACLOR	SERUM SICK	3164	69.702	45.39	43.81	3063	438	44.96	43.34	134	100

C.2 Data 2

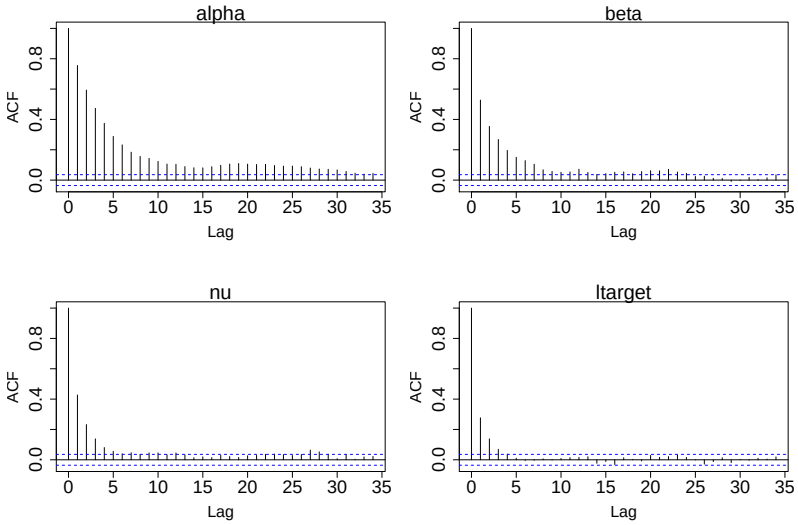


Figure C.7: Acf plots of α, β, ν and the logarithm of the target distribution for C-G and Data 2.

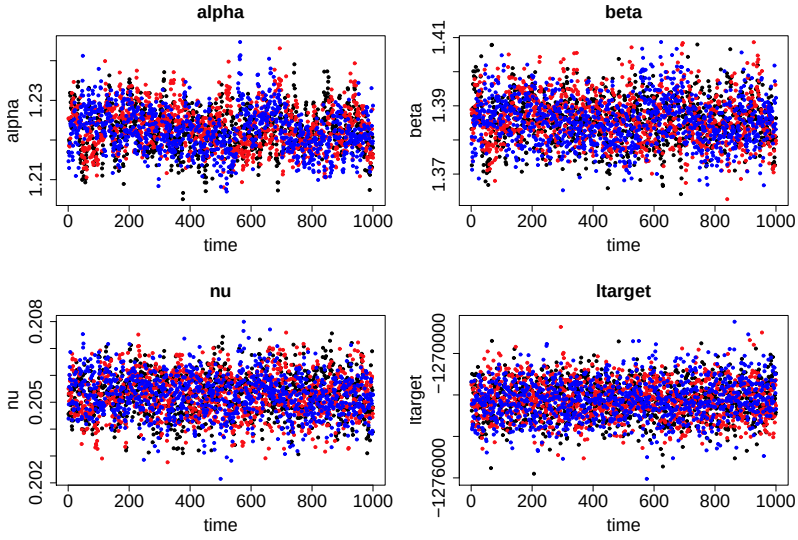
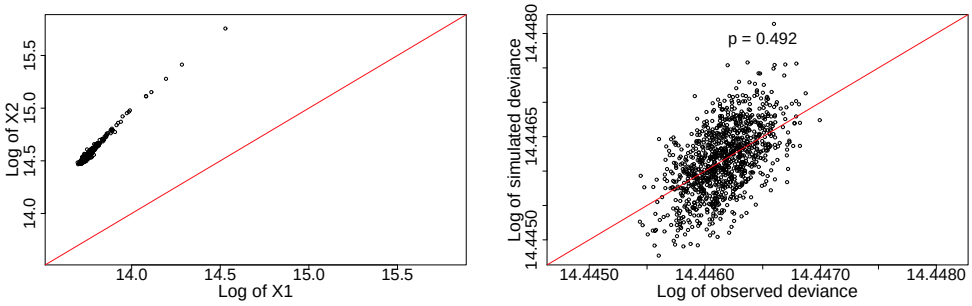
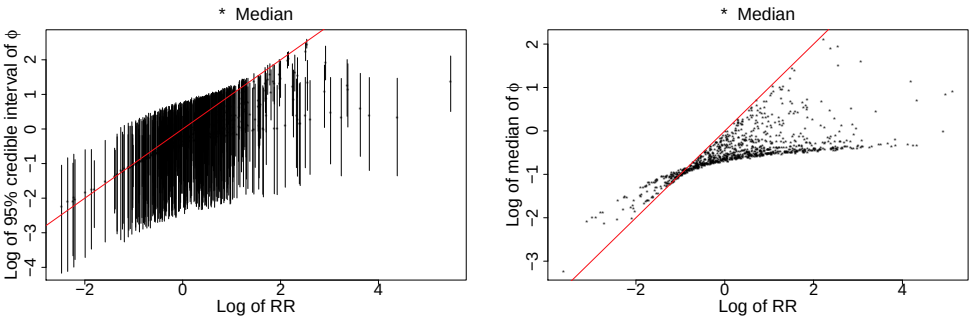


Figure C.8: Trace plots of α, β, ν and the logarithm of the target distribution for three chains, for C-G and Data 2.



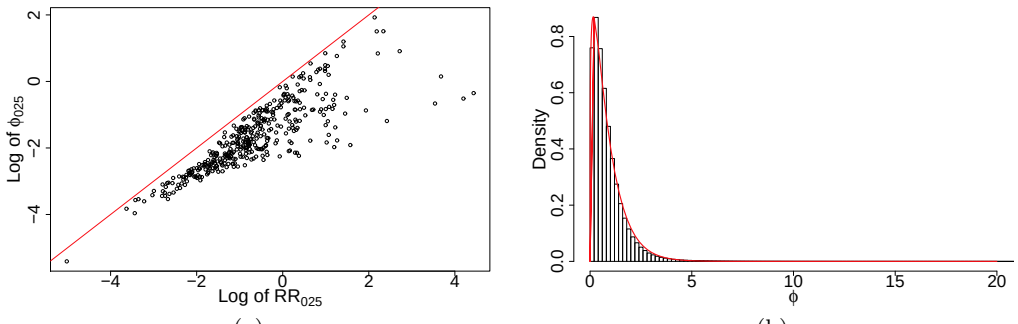
(a) (b)

Figure C.9: Bayesian p-value scatter plots for C-G and Data 2. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.



(a) (b)

Figure C.10: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for the C-G model. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 2, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for C-G model. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 2.



(a) (b)

Figure C.11: (a) Scatter plot of logarithm of ϕ_{025} against the logarithm of RR_{025} for C-G and Data 2, and (b) Histogram of ϕs for C-G and Data 2. The superimposed density curve (in red) represents the Gamma distribution corresponding to the MCMC estimates of the posterior means of φ and ψ.

Table C.3: Values of original counts, mean replicate counts, expected counts, RR and ϕ for fifty randomly selected drug and side-effect pairs using C-G model and Data 2.

Count	Mean Rep. Count	Expected Count	RR	ϕ
1	0.2	0.000	2816.96	0.99
18	12.3	5.957	3.02	1.72
3	1.7	0.339	8.84	1.09
1	0.3	0.001	694.57	0.99
175	157.7	43.911	3.99	3.54
64	59.9	42.377	1.51	1.38
527	490.7	79.506	6.63	6.15
2	4.4	4.698	0.43	0.53
1	1.7	0.471	2.12	0.77
5	6.2	5.663	0.88	0.71
15	12.6	8.162	1.84	1.26
22	17.1	9.189	2.39	1.62
579	577.2	452.946	1.28	1.27
1481	1321.4	54.248	27.3	24.34
3	4.1	3.064	0.98	0.74
22	20.5	17.797	1.24	1.05
58	2.5	0.098	594.08	9.07
12	16.5	53.081	0.23	0.27
803	773.9	142.740	5.63	5.39
7	5.0	2.600	2.69	1.17
6	0.6	0.010	576.05	1.69
1	2.1	0.880	1.14	0.73
1	0.3	0.002	559.28	0.99
1	1.2	0.160	6.25	0.87
10	0.7	0.017	590.60	2.21
1	1.8	0.616	1.62	0.75
150	7.0	0.228	658.94	22.13
13	12.4	10.760	1.21	0.95
1	1.7	0.570	1.75	0.77
1	3.3	2.780	0.36	0.53
7	2.2	0.418	16.76	1.62
46	2.0	0.073	628.21	7.41
1	2.5	1.519	0.66	0.63
1	2.4	1.298	0.77	0.65
604	601.8	532.731	1.13	1.13
1	2.8	1.708	0.59	0.63
8	7.4	5.231	1.53	1.00
1	1.9	0.597	1.67	0.78
1006	1002.0	749.688	1.34	1.33
1196	1198.6	2962.320	0.40	0.40
1094	1065.8	202.805	5.39	5.24
642	642.2	775.209	0.83	0.83
773	704.1	62.563	12.36	11.21
1	1.6	0.431	2.32	0.75
1	3.4	3.125	0.32	0.52
37	1.5	0.041	895.62	6.20
1	3.2	2.429	0.41	0.57
1	2.7	1.623	0.62	0.63
3	4.3	3.55	0.85	0.69
18	19.6	30.366	0.59	0.59

Table C.4a: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 2. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	114.76	105.72	1	1
COPPER	UTER DIS	4457	46.398	96.06	93.26	975	151	83.86	78.55	2	2
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	510	95	83.4	74.3	3	3
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	81.66	71.71	4	4
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	552	104	72.2	63.12	5	5
DTP VACCINE	SCREAMING SYND	966	9.643	100.18	93.95	897	147	59.09	51.17	6	6
MICONAZOLE	VAGINITIS	2262	34.851	64.9	62.24	1818	271	54.45	49.55	7	7
COPPER	PAIN PELVIC	1962	30.696	63.92	61.12	1861	278	52.47	47.38	10	8
COPPER	SALPINGITIS	1484	21.323	69.6	66.08	1635	238	52.91	47.17	9	9
COPPER	DEVICE MIGRATION	1490	22.868	65.16	61.88	1811	272	50.37	44.93	11	10
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	53.32	42.99	8	11
UROKINASE	CHILLS	831	10.942	75.95	70.83	1426	220	46.97	40.36	13	12
CEFACLOR	SERUM SICK	3164	69.702	45.39	43.81	3063	438	41.42	38.4	17	13
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	47.11	37.05	12	14
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	44.84	35.86	14	15
NONOXYNOL	DEVICE MIGRATION	800	13.275	60.27	56.12	2045	311	40.05	34.34	21	16
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM CEREBR	695	10.419	66.71	61.81	1744	274	40.65	34.17	19	17
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	44.26	34.13	15	18
NONOXYNOL	LEUKORRHEA	603	8.344	72.27	66.51	1516	235	40.03	33.49	22	19
COPPER	FERTIL DEC FEM	695	10.997	63.2	58.56	1899	294	39.12	33.17	24	20
INSULIN HUMAN	HYPERGLYCEM	6200	170.45	36.37	35.47	4270	560	34.99	33.03	30	21
LOMEFLOXACIN HYDROCHLORIDE	PHOTOSENSITIVITY	481	5.137	93.64	85.27	1023	166	40.7	32.99	18	22
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	42.53	32.73	16	23
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM INTRACRAN	476	5.426	87.73	79.99	1150	184	39.12	32.04	23	24
BUTOCONAZOLE NITRATE	VAGINITIS	1072	22.835	46.94	44.14	2902	434	36.33	31.83	27	25
CEFUROXIME	COLITIS PSEUDOMEM	1063	22.881	46.46	43.7	2953	440	35.93	31.43	28	26
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	40.47	31.31	20	27
PHENFORMIN	ACIDOSIS LACTIC	390	3.302	118.11	106.61	681	124	39.07	31.13	25	28
TRETINOIN	DERM EXFOL	1213	29.963	40.48	38.21	3643	508	33.1	29.28	36	29
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	37.2	29.04	26	30
POLIOVIRUS VACCINE, LIVE, ORAL	SCREAMING SYND	344	3.28	104.89	93.91	831	148	34.53	27.45	31	31
MINOXIDIL	HIRSUTISM	1035	26.322	39.32	36.93	3802	532	31.32	27.37	40	32
CLONIDINE	DERM CONTACT	964	23.784	40.53	38.01	3633	513	31.64	27.33	39	33
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	437	94	35.23	27.28	29	34

Table C.4b: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 2. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
TRIAMCINOLONE	ATROPHY INJECT SITE	439	6.303	69.65	63.14	1630	260	33.71	27.18	34	35
SUPROFEN	PAIN BACK	399	5.188	76.91	69.39	1404	227	33.52	27.05	35	36
PRED-G	PAIN EYE	644	13.125	49.07	45.33	2739	417	32.37	26.94	37	37
NAFARELIN ACETATE	HEM VAGINAL	506	9.008	56.17	51.29	2260	349	32.14	26.3	38	38
SELENIUM SULFIDE	HAIR DISCOLOR	301	2.236	134.64	119.43	526	107	33.74	26.17	33	39
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	34.32	25.71	32	40
MINOXIDIL	HAIR DIS	928	24.654	37.64	35.25	4048	563	29.54	25.62	46	41
INSULIN NOVOLIN 70/30	HYPERGLYCEM	484	8.959	54.02	49.22	2408	370	30.86	25.31	41	42
PROCAINAMIDE	LE SYND	471	8.679	54.27	49.43	2386	367	30.74	25	43	43
TIMOPTIC	PAIN EYE	695	16.875	41.19	38.16	3539	509	29.49	24.87	48	44
BUTORPHANOL	DRUG DEPEND	457	8.165	55.97	50.95	2270	353	30.78	24.73	42	45
CEFACLOR	ARTHROSIS	1010	30.298	33.34	31.29	4809	673	27.27	23.75	53	46
TRAZODONE	PRIAPISM	347	5.117	67.81	60.77	1701	282	29.26	23.32	50	47
OPCON A	PAIN EYE	267	2.015	132.49	116.61	547	109	30.58	23.22	44	48
NONOXYNOL	VAGINITIS	1141	36.497	31.26	29.45	5245	720	26.45	23.17	55	49
TRETINOIN	SKIN DRY	750	21.028	35.67	33.15	4389	618	27.05	23.05	54	50
METHADONE	WITHDRAW SYND	1233	41.143	29.97	28.32	5534	759	25.8	22.78	57	51
ISOTRETINOIN	CHEILITIS	380	6.983	54.42	48.97	2372	374	27.78	22.22	51	52
CEFIXIME	COLITIS PSEUDOMEM	421	8.596	48.98	44.32	2749	432	27.49	22.14	52	53
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	29.8	21.96	45	54
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	29.48	21.87	49	55
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	29.54	21.73	47	56
CEFACLOR	ERYTHEMA MULT	1481	54.248	27.3	25.92	6279	836	24.3	21.63	67	57
WARFARIN	PROTHROMBIN INC	762	23.392	32.58	30.27	4966	706	25.31	21.53	60	58
TUBERCULIN, PURIFIED PROTEIN DERIVATIV	INJECT SITE REACT	411	9.536	43.1	39.01	3301	500	25.35	20.39	59	59
NONOXYNOL	PREGN UNINTEND	1173	43.993	26.66	25.14	6471	876	23.17	20.32	70	60
WARFARIN	PROTHROMBIN DEC	4355	193.608	22.49	21.83	8172	1028	21.71	20.29	78	61
TIOCONAZOLE	VAGINITIS	268	3.558	75.32	66.33	1442	236	26.18	20.25	56	62
ALENDRONATE SODIUM	PAIN BONE	492	13.442	36.6	33.4	4231	609	24.45	20.07	65	63
BENOXAPROFEN	NAIL DIS	382	8.665	44.09	39.7	3200	489	24.91	19.86	63	64
ENOXAPARIN SODIUM	HEM	342	7.257	47.13	42.17	2891	454	24.42	19.31	66	65
RISPERIDONE	PROLACTIN INC	360	7.978	45.12	40.49	3090	475	24.65	19.28	64	66
INTRAUTERINE DEVICE	UTER DIS	251	3.127	80.26	70.35	1328	224	25.54	19.09	58	67

Table C.4c: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 2. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
MOXALACTAM	PROTHROMBIN DEC	683	23.759	28.75	26.6	5836	805	22.41	18.92	75	68
MITOMYCIN	UREMIA	218	1.995	109.29	95.26	785	142	24.96	18.75	62	69
TETRAHYDROZOLINE	CONJUNCTIVITIS	230	2.454	93.73	81.91	1021	174	25.17	18.72	61	70
MINOXIDIL	HAIR DISCOLOR	433	12.27	35.29	32.03	4451	647	22.87	18.53	72	71
COPPER	UTER RUPT	270	4.665	57.88	51.02	2170	352	23.72	18.07	69	72
CLINDAMYCIN	COLITIS	625	22.373	27.94	25.75	6081	842	21.5	17.92	79	73
LEVONORGESTREL	REACT UNEVAL	5043	259.501	19.43	18.9	9866	1209	18.95	17.79	105	74
SCOPOLAMINE	WITHDRAW SYND	784	30.874	25.39	23.64	6869	935	20.89	17.72	83	75
MICONAZOLE	VULVOVAGINITIS	234	3.168	73.87	64.71	1471	248	23.72	17.69	68	76
CLINDAMYCIN	COLITIS PSEUDOMEM	672	25.375	26.48	24.51	6527	901	20.99	17.49	82	77
INSULIN HUMAN	HYPOGLYCEM	1810	87.027	20.8	19.84	9040	1148	19.33	17.41	102	78
LOPERAMIDE	PHARYNGITIS	590	21.605	27.31	25.13	6276	877	20.88	17.34	85	79
SCOPOLAMINE	MYDRIASIS	269	5.256	51.18	45.09	2582	422	22.54	17.24	73	80
PHENYTOIN	DRUG LEVEL DEC	1295	59.805	21.65	20.48	8585	1111	19.48	17.19	99	81
COPPER	ENDOMETR DIS	334	8.68	38.48	34.45	3924	577	21.79	17.17	76	82
TETRAHYDROZOLINE	PAIN EYE	187	1.43	130.77	112.59	557	117	23.09	16.75	71	83
TRETINOIN	SKIN DIS	382	11.665	32.75	29.49	4926	718	20.77	16.52	86	84
VALPROIC ACID	NPN INC	475	17.088	27.8	25.34	6124	867	20.03	16.41	93	85
NITRODISC	APPLICAT SITE REACT	326	8.823	36.95	32.98	4176	621	21.04	16.39	81	86
PILOCARPINE	PAIN EYE	303	7.996	37.89	33.64	4017	603	20.65	16.01	89	87
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	22.42	16	74	88
MINITRAN	APPLICAT SITE REACT	340	9.98	34.07	30.46	4674	698	20.43	15.95	90	89
PRED-G	CONJUNCTIVITIS	329	9.708	33.89	30.28	4703	704	20.01	15.8	94	90
ISOTRETINOIN	HYPERLIPEM	726	32.758	22.16	20.58	8344	1103	18.4	15.8	114	91
MINOXIDIL	SKIN DRY	1038	50.729	20.46	19.22	9239	1185	18.06	15.8	124	92
TETRACYCLINE	TOOTH DIS	248	5.312	46.68	41.04	2932	466	20.73	15.71	87	93
RITONAVIR	PARESTH CIRCUMORAL	217	3.566	60.84	52.99	2005	332	21.15	15.69	80	94
DIGOXIN	DIGITALIS INTOX	1474	77.322	19.06	18.09	10121	1267	17.56	15.66	129	95
MINOXIDIL	ALOPECIA	4479	262.608	17.06	16.56	11732	1407	16.64	15.61	145	96
PROCAINAMIDE	ANA	300	8.28	36.23	32.25	4289	642	20.06	15.49	92	97
NITRO-DUR	APPLICAT SITE REACT	592	25.651	23.08	21.25	7892	1057	18.37	15.33	116	98
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	21.77	15.17	77	99
EFIDAC/24	INSOMNIA	393	14.23	27.62	24.95	6171	883	18.8	15.09	108	100

C.3 Data 3

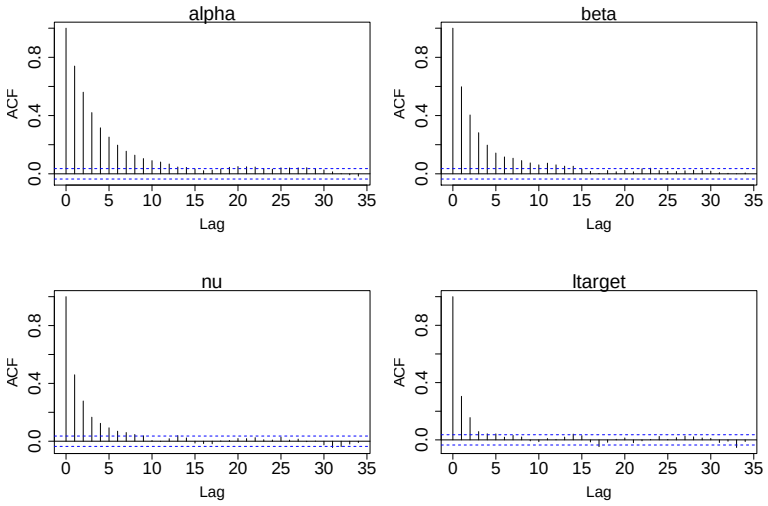


Figure C.12: Acf plots of α, β, ν and the logarithm of the target distribution for C-G and Data 3.

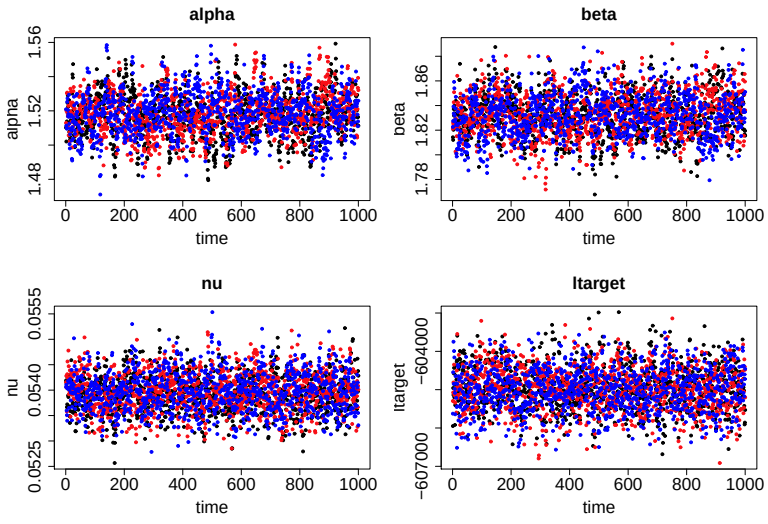
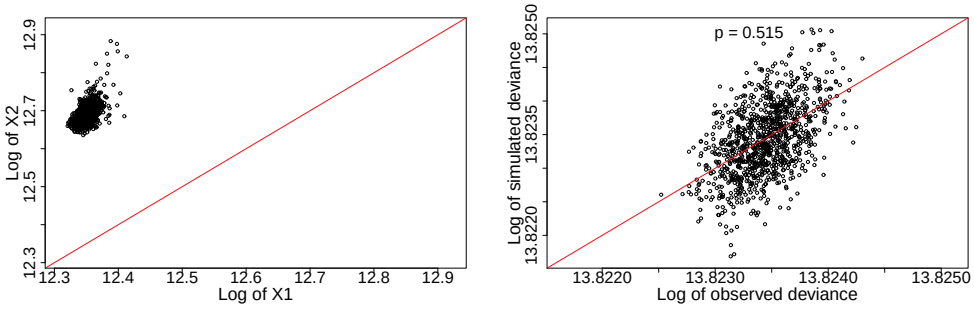
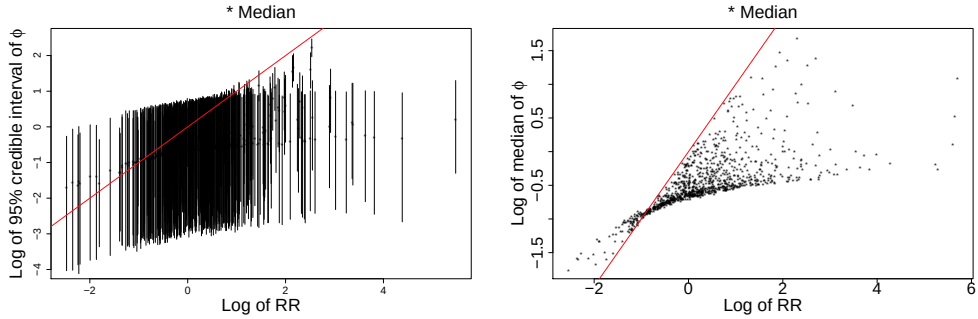


Figure C.13: Trace plots of α, β, ν and the logarithm of the target distribution for three chains, for C-G and Data 3.



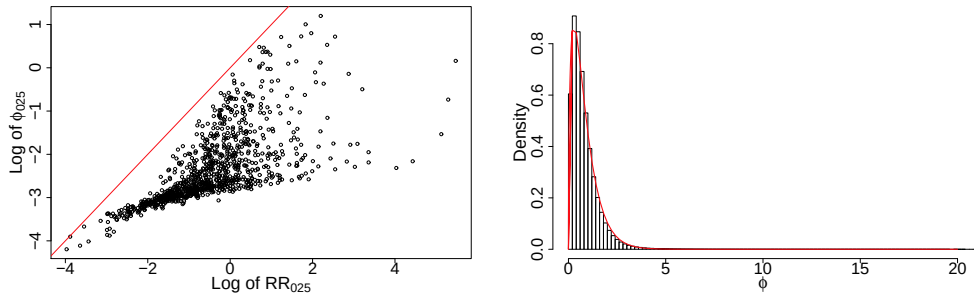
(a) (b)

Figure C.14: Bayesian p-value scatter plots for C-G and Data 3. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.



(a) (b)

Figure C.15: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for the C-G model. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 3, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for C-G model. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 3.



(a) (b)

Figure C.16: (a) Scatter plot of logarithm of ϕ_{025} against the logarithm of RR_{025} for C-G and Data 3, and (b) Histogram of ϕs for C-G and Data 3. The superimposed density curve (in red) represents the Gamma distribution corresponding to the MCMC estimates of the posterior means of φ and ψ.

Table C.5: Values of original counts, mean replicate counts, expected counts, RR and ϕ for fifty randomly selected drug and side-effect pairs using C-G model and Data 3.

Count	Mean Rep. Count	Expected Count	RR	ϕ
10	10.5	3.771	2.65	0.82
684	581.8	141.379	4.84	4.03
23	24.3	19.812	1.16	0.78
527	518.1	477.337	1.10	1.07
89	87.6	88.794	1.00	0.89
527	389.9	79.506	6.63	4.79
28	3.3	0.031	889.14	1.55
1109	942.5	162.189	6.84	5.75
5	10.8	5.020	1.00	0.67
3164	2149.5	69.702	45.39	30.72
6	11.6	6.055	0.99	0.65
10	15.0	9.937	1.01	0.69
53	27.8	11.698	4.53	1.58
26	3.4	0.037	707.23	1.50
9	14.5	9.014	1.00	0.66
58	4.7	0.098	594.08	2.37
896	871.8	533.355	1.68	1.62
107	105.1	106.067	1.01	0.90
151	146.8	142.483	1.06	0.96
9	14.1	8.989	1.00	0.66
6	8.4	2.228	2.69	0.77
11	15.6	11.050	1.00	0.68
199	89.4	20.056	9.92	4.01
26	38.1	65.969	0.39	0.44
1062	1051.5	904.794	1.17	1.15
126	101.1	58.529	2.15	1.56
20	29.1	38.710	0.52	0.52
695	686.6	684.336	1.02	0.99
5	10.2	4.241	1.18	0.68
8	11.8	5.329	1.50	0.72
17	16.7	9.548	1.78	0.85
7	12.7	7.001	1.00	0.67
10	9.5	2.704	3.70	0.86
1404	1333.1	455.788	3.08	2.91
544	556.8	1218.600	0.45	0.45
557	565.9	917.305	0.61	0.61
1400	1333.9	472.530	2.96	2.81
843	834.2	756.036	1.12	1.09
26	35.2	46.838	0.56	0.56
579	530.5	198.300	2.92	2.60
580	571.6	505.762	1.15	1.11
9	14.0	9.005	1.00	0.67
306	9.0	0.198	1543.39	9.55
6	2.1	0.004	1577.25	0.95
14	14.1	6.678	2.10	0.85
55	67.5	119.829	0.46	0.49
387	15.1	0.542	714.05	11.69
293	8.6	0.185	1583.73	9.16
11	15.8	11.036	1.00	0.66
10	14.7	9.043	1.11	0.71

Table C.6a: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 3. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
COPPER	UTER DIS	4457	46.398	96.06	93.26	321	150	55.61	48.96	1	1
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	165	79	54.62	46.44	2	2
MICONAZOLE	VAGINITIS	2262	34.851	64.9	62.24	532	265	33	27.49	3	3
INSULIN HUMAN	HYPERGLYCEM	6200	170.45	36.37	35.47	1123	545	30.35	27.25	7	4
CEFACLOR	SERUM SICK	3164	69.702	45.39	43.81	845	426	30.63	26.35	5	5
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	198	94	32.61	25.85	4	6
COPPER	PAIN PELVIC	1962	30.696	63.92	61.12	543	272	30.46	25.24	6	7
COPPER	SALPINGITIS	1484	21.323	69.6	66.08	487	233	26.97	21.55	9	8
COPPER	DEVICE MIGRATION	1490	22.868	65.16	61.88	530	266	26.4	21.22	10	9
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	158	77	27.44	21.11	8	10
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	212	103	25.65	19.78	11	11
DTP VACCINE	SCREAMING SYND	966	9.643	100.18	93.95	303	146	22.27	16.86	12	12
WARFARIN	PROTHROMBIN DEC	4355	193.608	22.49	21.83	2019	989	19.17	16.86	13	13
LEVONORGESTREL	REACT UNEVAL	5043	259.501	19.43	18.9	2392	1159	17.21	15.24	19	14
TRETINOIN	DERM EXFOL	1213	29.963	40.48	38.21	982	494	19.12	14.82	14	15
CEFUROXIME	COLITIS PSEUDOMEM	1063	22.881	46.46	43.7	817	428	18.84	14.62	16	16
BUTOCONAZOLE NITRATE	VAGINITIS	1072	22.835	46.94	44.14	806	422	19	14.61	15	17
UROKINASE	CHILLS	831	10.942	75.95	70.83	426	216	18.66	13.79	17	18
CEFACLOR	ERYTHEMA MULT	1481	54.248	27.3	25.92	1609	809	16.93	13.58	21	19
MINOXIDIL	ALOPECIA	4479	262.608	17.06	16.56	2834	1346	15.12	13.32	30	20
MINOXIDIL	HIRSUTISM	1035	26.322	39.32	36.93	1006	518	17.35	13.16	18	21
METHADONE	WITHDRAW SYND	1233	41.143	29.97	28.32	1426	737	16.53	12.94	23	22
CLONIDINE	DERM CONTACT	964	23.784	40.53	38.01	980	499	16.86	12.8	22	23
NONOXYNOL	DEVICE MIGRATION	800	13.275	60.27	56.12	590	304	17.14	12.8	20	24
NONOXYNOL	VAGINITIS	1141	36.497	31.26	29.45	1352	699	16.36	12.6	24	25
PERMETHRIN	NO DRUG EFFECT	4029	249.042	16.18	15.68	3032	1450	14.28	12.48	33	26
INSULIN HUMAN	HYPOGLYCEM	1810	87.027	20.8	19.84	2194	1102	15.04	12.32	31	27
LEVONORGESTREL	METRORRHAGIA	7530	524.28	14.36	14.04	3510	1664	13.52	12.26	38	28
CEFACLOR	ARTHROSIS	1010	30.298	33.34	31.29	1261	653	15.9	12.17	27	29
MINOXIDIL	HAIR DIS	928	24.654	37.64	35.25	1073	548	15.94	11.96	25	30
NONOXYNOL	PREGN UNINTEND	1173	43.993	26.66	25.14	1650	845	15.17	11.69	29	31
NICOTINE	APPLICAT SITE REACT	5085	358.599	14.18	13.79	3563	1699	12.97	11.44	42	32
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM CEREBR	695	10.419	66.71	61.81	514	268	15.91	11.34	26	33
COPPER	FERTIL DEC FEM	695	10.997	63.2	58.56	553	287	15.67	11.28	28	34

Table C.6b: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 3. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
INSULIN	HYPERGLYCEM	2429	148.433	16.36	15.72	2992	1446	13.36	11.23	40	35
PHENYTOIN	DRUG LEVEL DEC	1295	59.805	21.65	20.48	2100	1067	13.86	11	35	36
DIGOXIN	DIGITALIS INTOX	1474	77.322	19.06	18.09	2458	1215	13.29	10.71	41	37
NONOXYNOL	LEUKORRHEA	603	8.344	72.27	66.51	457	231	14.51	10.1	32	38
DIATRIZOIC ACID	URTICARIA	5404	442.648	12.21	11.88	4306	2040	11.36	10.08	55	39
COPPER	METRORRHAGIA	2022	132.31	15.28	14.62	3240	1582	12.23	10.07	47	40
PRED-G	PAIN EYE	644	13.125	49.07	45.33	765	405	13.89	10.04	34	41
TIMOPTIC	PAIN EYE	695	16.875	41.19	38.16	958	495	13.81	10.04	36	42
TRETINOIN	SKIN DRY	750	21.028	35.67	33.15	1154	602	13.79	9.89	37	43
WARFARIN	PROTHROMBIN INC	762	23.392	32.58	30.27	1298	686	13.48	9.84	39	44
MINOXIDIL	SKIN DRY	1038	50.729	20.46	19.22	2240	1137	12.35	9.5	44	45
DTP VACCINE	FEVER	2494	192.186	12.98	12.47	3992	1891	11.1	9.48	60	46
CLOZAPINE	LEUKOCYTOSIS	1352	79.763	16.95	16.05	2854	1402	11.99	9.45	50	47
TRIAZOLAM	AMNESIA	1108	59.977	18.47	17.39	2563	1273	11.89	9.09	51	48
SCOPOLAMINE	WITHDRAW SYND	784	30.874	25.39	23.64	1736	904	12.24	9.02	46	49
ESTRADIOL	APPLICAT SITE REACT	3664	325.306	11.26	10.9	4715	2283	10.22	8.9	77	50
IOTHALAMIC ACID	URTICARIA	2924	248.904	11.75	11.32	4495	2174	10.36	8.87	74	51
ENALAPRIL	COUGH INC	2033	157.884	12.88	12.32	4028	1929	10.65	8.79	69	52
INSULIN	HYPOGLYCEM	1267	81.804	15.49	14.64	3186	1576	11.07	8.65	61	53
MOXALACTAM	PROTHROMBIN DEC	683	23.759	28.75	26.6	1500	781	12.06	8.58	48	54
COPPER	PREGN UNINTEND	1312	88.225	14.87	14.07	3356	1659	10.83	8.49	65	55
LOMEFLOXACIN HYDROCHLORIDE	PHOTOSENSITIVITY	481	5.137	93.64	85.27	335	164	12.58	8.42	43	56
PHENYTOIN	DRUG LEVEL INC	2321	206.125	11.26	10.8	4717	2318	9.72	8.15	84	57
NAFARELIN ACETATE	HEM VAGINAL	506	9.008	56.17	51.29	636	341	12.02	8.15	49	58
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM INTRACRAN	476	5.426	87.73	79.99	364	182	12.26	8.13	45	59
ISOTRETINOIN	HYPERLIPEM	726	32.758	22.16	20.58	2051	1059	11.04	8.12	62	60
CLINDAMYCIN	COLITIS PSEUDOMEM	672	25.375	26.48	24.51	1667	870	11.55	8.12	52	61
ALCOHOL	OVERDOSE INTENT	797	40.254	19.8	18.43	2330	1195	10.94	8.07	64	62
LORAZEPAM	DRUG DEPEND	1023	64.888	15.77	14.81	3115	1562	10.5	8.01	71	63
CLINDAMYCIN	COLITIS	625	22.373	27.94	25.75	1555	815	11.27	7.98	57	64
ALPRAZOLAM	WITHDRAW SYND	1671	138.118	12.1	11.52	4348	2119	9.76	7.85	83	65
CLARITHROMYCIN	TASTE PERVERS	1027	67.011	15.33	14.4	3225	1616	10.25	7.85	76	66
INSULIN NOVOLIN 70/30	HYPERGLYCEM	484	8.959	54.02	49.22	674	360	11.5	7.79	53	67

Table C.6c: Top hundred drug and adverse event pairs selected by C-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 3. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
INTERFERON BETA	FLU SYND	921	56.21	16.39	15.34	2989	1492	10.34	7.78	75	68
SELENIUM SULFIDE	NO DRUG EFFECT	722	34.737	20.78	19.29	2196	1133	10.68	7.73	68	69
GLYBURIDE	HYPOGLYCEM	919	57.378	16.02	14.99	3058	1542	10.15	7.68	78	70
PROCAINAMIDE	LE SYND	471	8.679	54.27	49.43	666	357	11.32	7.48	56	71
LOPERAMIDE	PHARYNGITIS	590	21.605	27.31	25.13	1608	846	10.8	7.47	66	72
ALBUTEROL	ASTHMA	2226	215.526	10.33	9.9	5244	2576	8.93	7.43	102	73
LEVONORGESTREL	AMENORRHEA	1583	140.522	11.27	10.71	4713	2343	9.15	7.41	98	74
INTERFERON BETA	INJECT SITE REACT	1370	113.854	12.03	11.4	4370	2151	9.35	7.39	91	75
TRIAMCINOLONE	ATROPHY INJECT SITE	439	6.303	69.65	63.14	483	254	11.11	7.34	59	76
BUTORPHANOL	DRUG DEPEND	457	8.165	55.97	50.95	639	344	11.11	7.24	58	77
HEPARIN	HEM CEREBR	734	39.875	18.41	17.08	2575	1302	10.05	7.22	81	78
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	12	5	11.42	7.21	54	79
LOVASTATIN	CREATINE PK INC	1131	89.674	12.61	11.89	4143	2039	9.26	7.15	94	80
ALENDRONATE SODIUM	PAIN BONE	492	13.442	36.6	33.4	1114	594	10.56	7.11	70	81
NITRO-DUR	APPLICAT SITE REACT	592	25.651	23.08	21.25	1962	1016	10.08	7.1	80	82
ISOTRETINOIN	SKIN DRY	673	34.791	19.34	17.91	2410	1232	9.95	7.09	82	83
FLUOXETINE HYDROCHLORIDE	SUICIDE ATTEMPT	2452	259.476	9.45	9.08	5871	2831	8.4	7.06	121	84
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	111	56	10.94	6.93	63	85
PHENFORMIN	ACIDOSIS LACTIC	390	3.302	118.11	106.61	247	123	10.72	6.79	67	86
SUPROFEN	PAIN BACK	399	5.188	76.91	69.39	421	223	10.44	6.78	72	87
EFIDAC/24	NO DRUG EFFECT	932	70.661	13.19	12.35	3905	1918	8.97	6.78	101	88
HEPARIN	HEM	1150	98.976	11.62	10.95	4549	2267	8.72	6.69	110	89
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	31	21	10.44	6.66	73	90
CLOZAPINE	SALIVA INC	549	24.183	22.7	20.84	1999	1040	9.63	6.59	86	91
ANGIOVIST 282	URTICARIA	678	41.814	16.21	14.99	3020	1541	9.11	6.58	100	92
CEFIXIME	COLITIS PSEUDOMEM	421	8.596	48.98	44.32	768	420	10.12	6.58	79	93
CEFAMANDOLE	PAIN	780	55.197	14.13	13.15	3577	1781	8.87	6.46	107	94
TUBERCULIN, PURIFIED PROTEIN DERIVATIV	INJECT SITE REACT	411	9.536	43.1	39.01	899	486	9.69	6.43	85	95
THEOPHYLLINE	DRUG LEVEL INC	1646	173.187	9.5	9.05	5826	2844	8	6.41	135	96
LORAZEPAM	WITHDRAW SYND	1297	127.043	10.21	9.66	5313	2654	8.12	6.39	128	97
MINOXIDIL	HAIR DISCOLOR	433	12.27	35.29	32.03	1170	629	9.61	6.33	87	98
ENALAPRIL	ANGIOEDEMA	1009	88.869	11.35	10.66	4666	2362	8.35	6.32	123	99
VALPROIC ACID	NPN INC	475	17.088	27.8	25.34	1563	836	9.51	6.3	89	100

Appendix D

P-G Model Results

D.1 Data 1

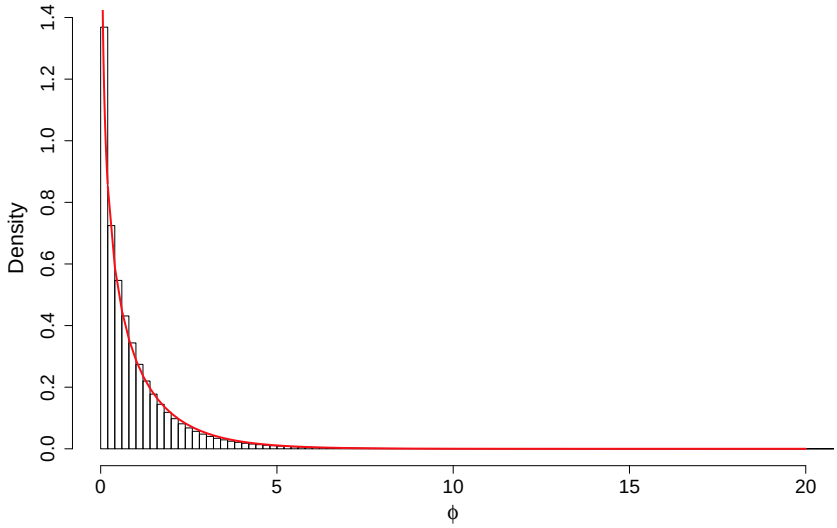


Figure D.1: Histogram of ϕs for P-G and Data 1. The superimposed density curve (in red) represents the Gamma distribution corresponding to the MCMC estimates of the posterior means of α and β.

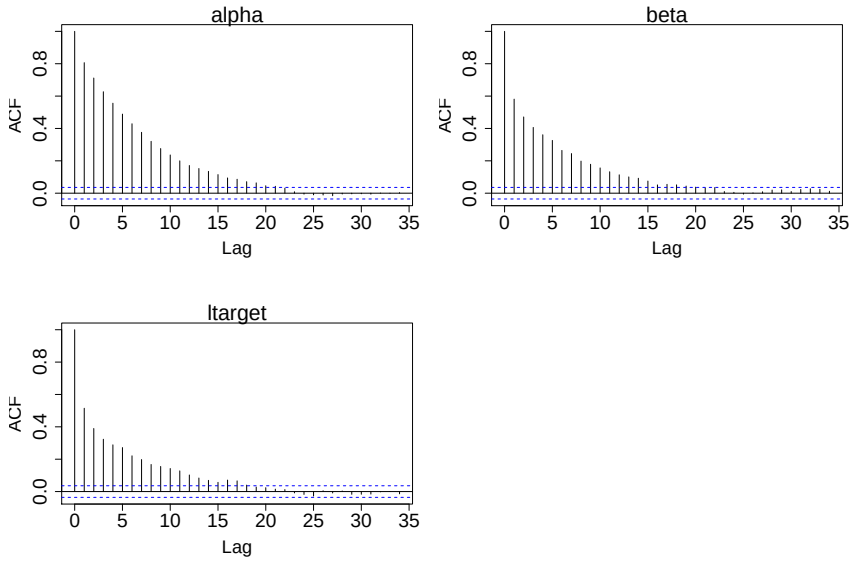


Figure D.2: Acf plots of α, β and the logarithm of the target distribution for P-G and Data 1.

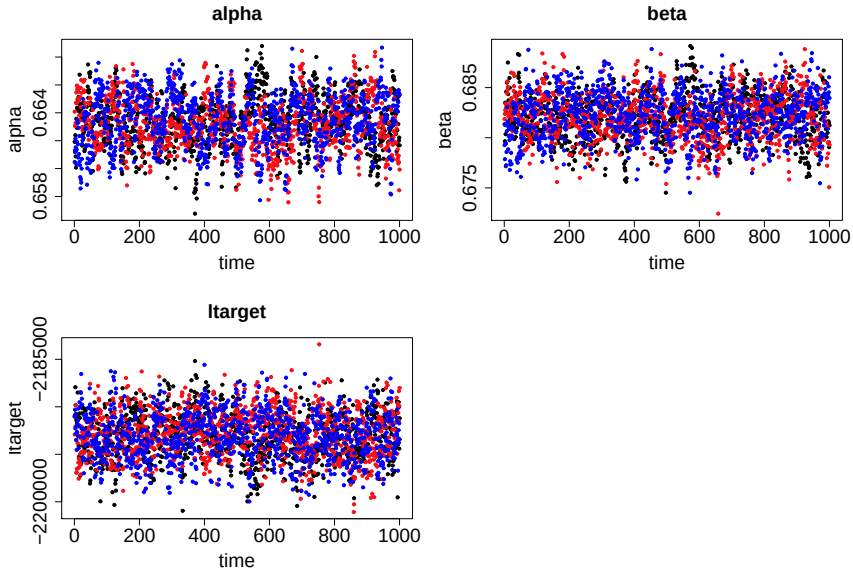
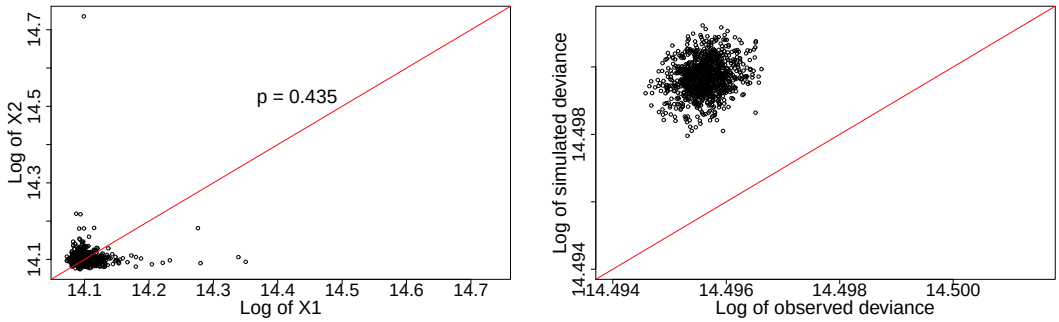
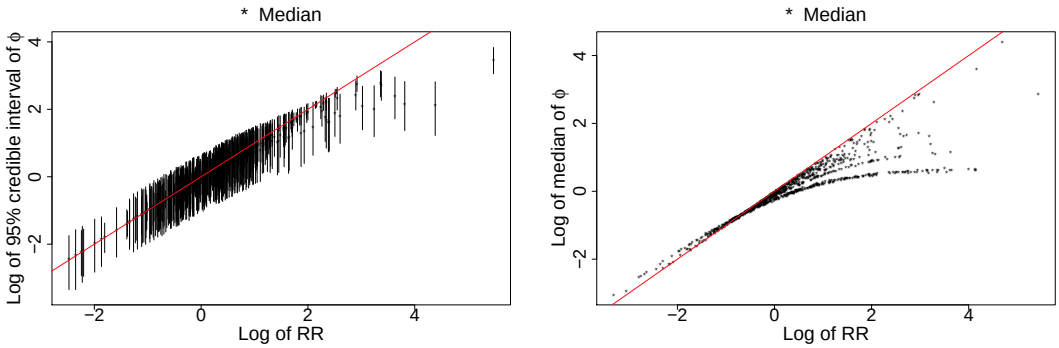


Figure D.3: Trace plots of α, β and the logarithm of the target distribution for three chains, for P-G and Data 1.



(a) (b)

Figure D.4: Bayesian p-value scatter plots for P-G and Data 1. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.



(a) (b)

Figure D.5: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for the P-G model. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 1, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for P-G model. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 1.

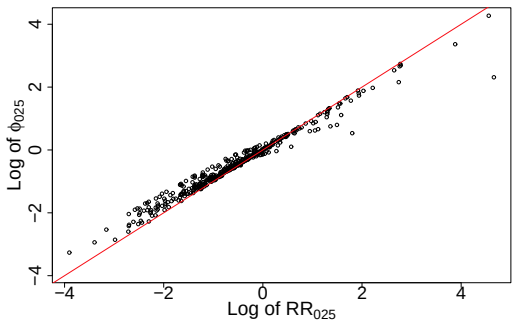


Figure D.6: Scatter plot of logarithm of ϕ_{025} against the logarithm of RR_{025} for P-G and Data 1.

Table D.1: Values of original counts, mean replicate counts, expected counts, RR and ϕ for fifty randomly selected drug and side-effect pairs using P-G model and Data 1.

Count	Mean Rep. Count	Expected Count	RR	ϕ
7	7.3	11.218	0.62	0.65
6	6.0	5.973	1.00	1.01
0	0.0	0.004	0.00	0.94
666	665.7	695.539	0.96	0.96
0	0.4	0.915	0.00	0.41
44	40.2	6.070	7.25	6.62
0	0.0	0.039	0.00	0.95
111	57.5	0.721	154.02	79.55
0	0.0	0.030	0.00	0.96
28	1.3	0.031	889.14	40.3
32	2.1	0.047	685.52	44.7
7	6.9	6.866	1.02	1.00
1	1.1	1.245	0.80	0.86
1	1.0	0.997	1.00	0.98
1	0.9	0.765	1.31	1.17
1	0.6	0.369	2.71	1.56
1	0.9	0.820	1.22	1.11
0	0.2	0.339	0.00	0.65
26	1.4	0.037	707.23	37.21
7	0.2	0.018	399.26	10.92
6	5.9	4.814	1.25	1.21
2022	2011.7	132.310	15.28	15.21
6	0.1	0.010	576.05	9.69
1021	1020.1	1701.710	0.60	0.60
637	636.5	606.791	1.05	1.05
1211	1211.4	739.652	1.64	1.64
933	933.1	854.123	1.09	1.09
1039	1039.6	380.965	2.73	2.73
56	10.0	0.147	381.61	68.35
1	0.6	0.409	2.44	1.53
46	4.5	0.073	628.21	61.73
2	1.9	1.625	1.23	1.16
3664	3658.6	325.306	11.26	11.25
1608	1609.9	1421.430	1.13	1.13
82	11.0	0.105	780.57	105.38
1	0.4	0.206	4.85	1.89
1032	1030.6	519.620	1.99	1.98
594	592.5	392.751	1.51	1.51
3	2.9	2.468	1.22	1.15
2	1.5	0.845	2.37	1.72
66	65.6	55.168	1.20	1.19
0	0.5	2.436	0.00	0.20
19	19.5	93.931	0.20	0.21
306	69.2	0.198	1543.39	348.6
306	69.2	0.198	1543.39	348.6
6	0.0	0.004	1577.25	9.74
10	0.4	0.023	431.67	15.28
1	0.4	0.198	5.05	1.87
3	2.5	1.432	2.09	1.74
5	5.0	4.914	1.02	1.02

Table D.2a: Top hundred drug and adverse event pairs selected by P-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	348.02	310.9	1	1
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	337.76	300.84	2	2
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	316.29	286.65	3	3
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	261.16	233.31	4	4
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	251.17	227.51	5	5
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	180.34	158.82	6	6
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	152.57	146.71	10	7
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	164.44	143.55	8	8
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	164.8	140.67	7	9
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	154.56	139.7	9	10
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	148.08	139.68	11	11
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	510	95	128.71	122.38	14	12
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	137.47	120.45	12	13
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	133.43	117.17	13	14
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	552	104	121.43	114.51	17	15
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	125.49	112.64	15	16
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	125.46	104.96	16	17
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	437	94	112.12	100.7	19	18
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	117.81	98.86	18	19
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	111.85	94.92	20	20
SELENIUM SULFIDE	HAIR DISCOLOR	301	2.236	134.64	119.43	526	107	103.34	92.6	22	21
COPPER	UTER DIS	4457	46.398	96.06	93.26	975	151	94.68	91.95	28	22
PHENFORMIN	ACIDOSIS LACTIC	390	3.302	118.11	106.61	681	124	97.94	89.04	27	23
DTP VACCINE	SCREAMING SYND	966	9.643	100.18	93.95	897	147	93.59	87.88	29	24
OPCON A	PAIN EYE	267	2.015	132.49	116.61	547	109	99.16	87.53	26	25
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	103.2	85.07	23	26
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	99.8	84.72	25	27
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	104.75	84.12	21	28
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	100.42	82.51	24	29
POLIOVIRUS VACCINE, LIVE, ORAL	SCREAMING SYND	344	3.28	104.89	93.91	831	148	86.75	78.13	34	30
TETRAHYDROZOLINE	PAIN EYE	187	1.43	130.77	112.59	557	117	88.48	76.48	33	31
LOMEFLOXACIN HYDROCHLORIDE	PHOTOSENSITIVITY	481	5.137	93.64	85.27	1023	166	82.75	75.68	37	32
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	92.42	72.94	30	33
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	89.57	72.29	32	34

Table D.2b: Top hundred drug and adverse event pairs selected by P-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
POLIOVIRUS VACCINE, LIVE, ORAL	SIDS	146	1.038	140.6	118.45	484	108	85.1	71.99	35	35
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	92.29	71.98	31	36
MITOMYCIN	UREMIA	218	1.995	109.29	95.26	785	142	81.59	71.55	39	37
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM INTRACRAN	476	5.426	87.73	79.99	1150	184	77.96	71.32	44	38
MENOTROPINS	OVAR DIS	151	1.175	128.48	108.06	583	122	81.5	69.33	40	39
CHOLINE	KERATITIS	116	0.703	164.99	135.12	372	91	83.86	68.96	36	40
CEFOXITIN	ENTEROCOL PSEUDOMEM	138	1.016	135.79	113.16	517	116	81.46	68.87	41	41
UROKINASE	CHILLS	831	10.942	75.95	70.83	1426	220	71.52	66.84	51	42
BENDECTIN	ECTROMELIA	111	0.721	154.02	126.27	415	100	79.21	65.88	43	43
DIPHTHERIA-TETANUS TOXOID-PERTUSSIS VA	SCREAMING SYND	147	1.22	120.47	101.62	654	130	77.29	65.54	46	44
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	82.02	65.53	38	45
COPPER	SALPINGITIS	1484	21.323	69.6	66.08	1635	238	67.45	64.18	57	46
TETRAHYDROZOLINE	CONJUNCTIVITIS	230	2.454	93.73	81.91	1021	174	73.32	64.12	50	47
HAEMOPHILUS B POLYSACCHARIDE VACCINE	MENINGITIS	176	1.711	102.86	87.66	851	164	73.49	63.26	48	48
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	80	63.14	42	49
NONOXYNOL	LEUKORRHEA	603	8.344	72.27	66.51	1516	235	66.87	61.67	58	50
SUPROFEN	PAIN BACK	399	5.188	76.91	69.39	1404	227	67.96	61.61	54	51
MICONAZOLE	VAGINITIS	2262	34.851	64.9	62.24	1818	271	63.66	61.08	63	52
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	77.43	60.05	45	53
COPPER	DEVICE MIGRATION	1490	22.868	65.16	61.88	1811	272	63.29	60.05	64	54
COPPER	PAIN PELVIC	1962	30.696	63.92	61.12	1861	278	62.51	59.72	70	55
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM CEREBR	695	10.419	66.71	61.81	1744	274	62.53	58.04	69	56
INTRAUTERINE DEVICE	UTER DIS	251	3.127	80.26	70.35	1328	224	65.93	57.95	60	57
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	73.47	57.95	49	58
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	74.38	57.63	47	59
TRIAMCINOLONE	ATROPHY INJECT SITE	439	6.303	69.65	63.14	1630	260	62.89	57.22	68	60
PILOCARPINE	MIOSIS	103	0.812	126.83	103.44	601	128	69.02	56.83	52	61
TIOCONAZOLE	VAGINITIS	268	3.558	75.32	66.33	1442	236	63.23	55.91	65	62
GONADOTROPIN, CHORIONIC	OVAR DIS	86	0.567	151.69	119.94	423	106	68.82	55.58	53	63
COPPER	FERTIL DEC FEM	695	10.997	63.2	58.56	1899	294	59.46	55.11	79	64
OXYTOCIN	FETAL DIS	95	0.759	125.14	100.11	616	133	66.09	54.17	59	65
ALCOHOL	ALCOHOL INTOLER	169	2.006	84.26	71.8	1233	215	62.93	54.15	67	66
TRAZODONE	PRIAPISM	347	5.117	67.81	60.77	1701	282	59.86	53.76	78	67

Table D.2c: Top hundred drug and adverse event pairs selected by P-G model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
NONOXYNOL	DEVICE MIGRATION	800	13.275	60.27	56.12	2045	311	57.41	53.65	83	68
FLUNISOLIDE	NASAL SEPTUM DIS	75	0.431	174.04	136.91	344	90	67.46	53.42	56	69
MICONAZOLE	VULVOVAGINITIS	234	3.168	73.87	64.71	1471	248	60.83	53.36	74	70
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	67.77	52.08	55	71
RUBELLA VIRUS VACCINE, LIVE	LYMPHADENO	115	1.162	98.97	80.9	925	179	62.41	51.85	71	72
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	65.29	50.05	61	73
OXYTOCIN	LABOR ABNORM	105	1.043	100.68	81.5	887	176	60.97	49.69	73	74
DORNASE ALFA	HEMOPTYSIS	64	0.336	190.37	145.75	301	83	63.18	49.62	66	75
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	64.17	48.66	62	76
FLOSEQUINAN	DEATH	109	1.204	90.5	73.89	1092	203	57.94	47.97	81	77
NAFARELIN ACETATE	HEM VAGINAL	506	9.008	56.17	51.29	2260	349	52.2	47.72	94	78
ATRACURIUM BESYLATE	PARALYSIS FLACCID	100	1.051	95.13	77.06	988	193	57.81	47.2	82	79
TICE BCG	CYSTITIS	69	0.475	145.35	111.64	457	119	59.9	47.2	77	80
BUTORPHANOL	DRUG DEPEND	457	8.165	55.97	50.95	2270	353	51.71	47.13	95	81
PRILOCAINE	PARESTH CIRCUMORAL	200	2.995	66.77	57.75	1742	304	54.39	47.11	90	82
OXYTETRACYCLINE	DISCOLOR TOOTH	74	0.599	123.49	96.79	633	138	57.98	45.96	80	83
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	60.29	45.91	75	84
INSULIN NOVOLIN 70/30	HYPERGLYCEM	484	8.959	54.02	49.22	2408	370	50.16	45.85	104	85
PROCAINAMIDE	LE SYND	471	8.679	54.27	49.43	2386	367	50.38	45.84	102	86
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	60.25	45.61	76	87
ISOTRETINOIN	BLIND NIGHT	140	1.94	72.15	60.3	1524	286	53.37	45.21	92	88
SOMATROPIN	OTITIS MED	101	1.147	88.07	71.5	1143	218	55.28	45.21	86	89
COPPER	UTER RUPT	270	4.665	57.88	51.02	2170	352	50.65	44.99	99	90
ISOTRETINOIN	CHEILITIS	380	6.983	54.42	48.97	2372	374	49.59	44.92	108	91
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	61.05	44.88	72	92
RITONAVIR	PARESTH CIRCUMORAL	217	3.566	60.84	52.99	2005	332	51.12	44.54	97	93
FOSCARNET SODIUM	HYPOCALCEM	99	1.149	86.18	69.64	1184	226	54.16	44.44	91	94
PHOTOPLEX	PHOTOSENSITIVITY	156	2.359	66.13	55.96	1766	314	51.38	43.93	96	95
DTP VACCINE	SIDS	207	3.495	59.23	51.21	2095	350	49.64	43.62	106	96
CEFACLOR	SERUM SICK	3164	69.702	45.39	43.81	3063	438	44.97	43.44	131	97
CLOMIPHENE	OVAR DIS	163	2.567	63.5	53.76	1882	326	50.17	43.18	103	98
PRED-G	PAIN EYE	644	13.125	49.07	45.33	2739	417	46.64	43.11	122	99
BUTOCONAZOLE NITRATE	VAGINITIS	1072	22.835	46.94	44.14	2902	434	45.58	42.9	126	100

Appendix E

C-IG Model Results

E.1 Data 1

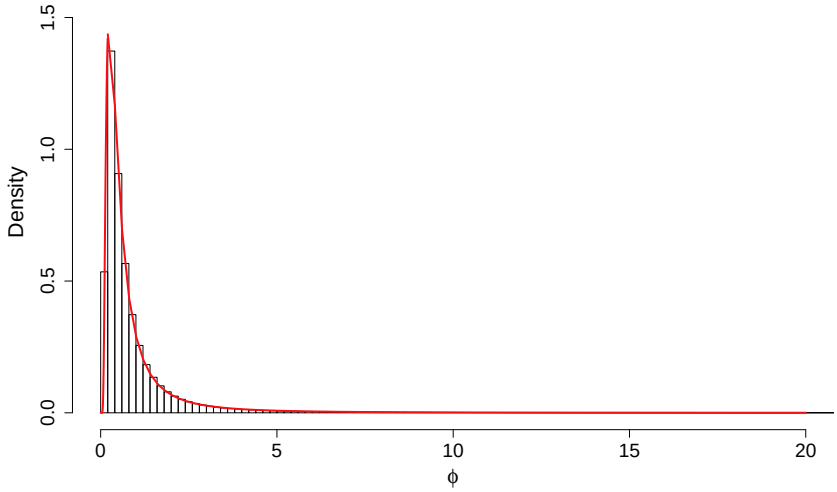


Figure E.1: Histogram of ϕs for C-IG and Data 1. The superimposed density curve (in red) represents the Gamma distribution corresponding to the MCMC estimates of the posterior means of φ and ψ.

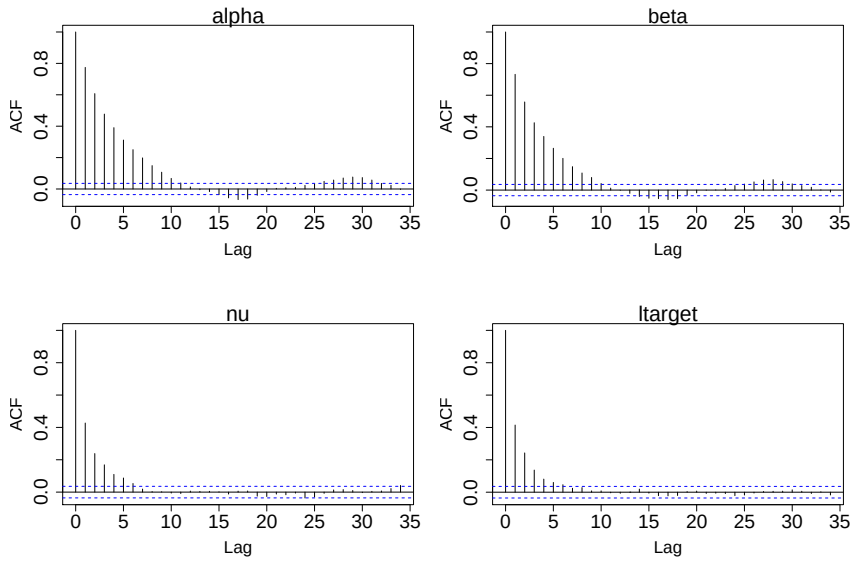


Figure E.2: Acf plots of φ, ψ, ν and the logarithm of the target distribution for C-IG and Data 1.

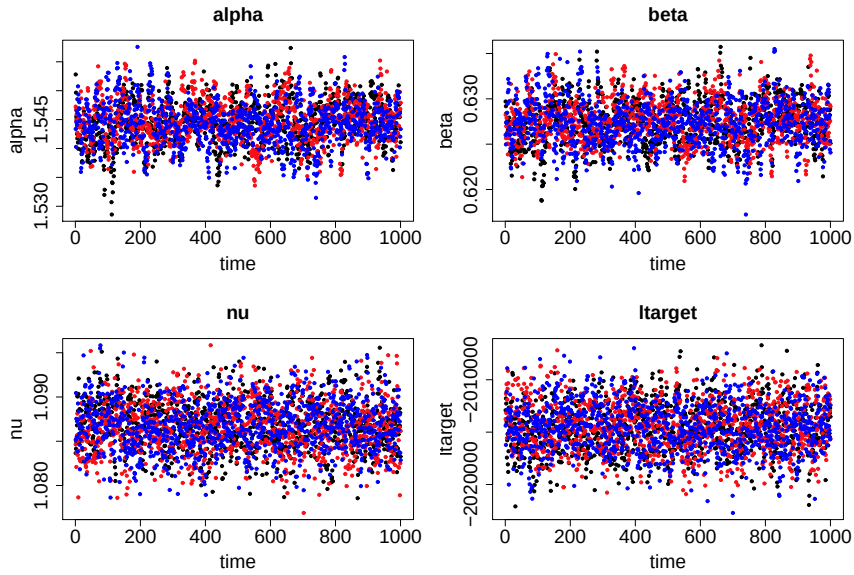
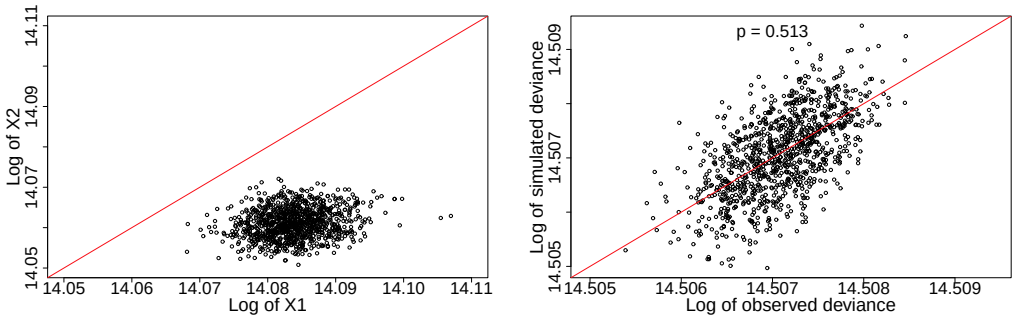
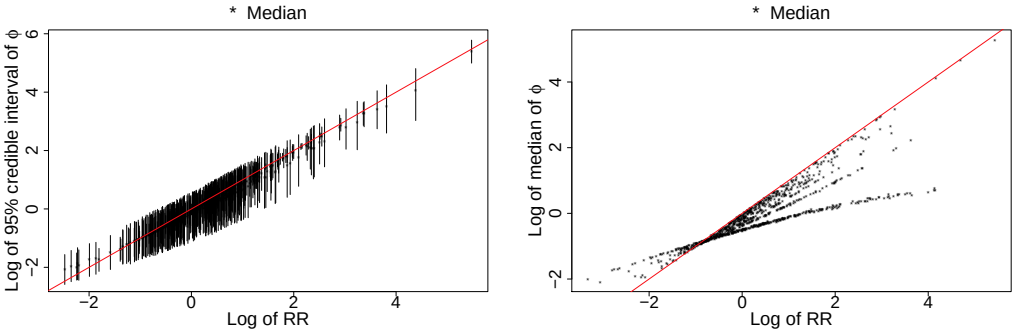


Figure E.3: Trace plots of φ, ψ, ν and the logarithm of the target distribution for three chains, for C-IG and Data 1.



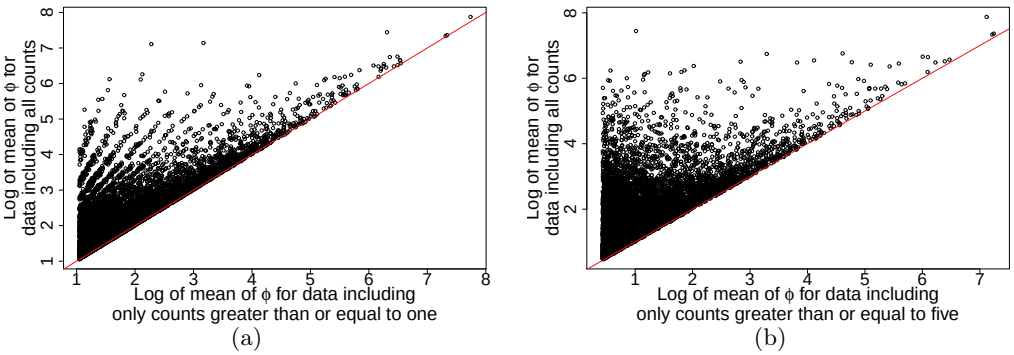
(a) (b)

Figure E.4: Bayesian p-value scatter plots for C-IG and Data 1. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.



(a) (b)

Figure E.5: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for the C-IG model. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 1, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for C-IG model. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 1.



(a) (b)

Figure E.6: (a) Scatter plot of logarithm of posterior means of ϕ for Data 1 against corresponding values for Data 2, and (b) Scatter plot of logarithm of mean of ϕ for Data 1 against corresponding values for Data 3. The plots were made with the top 10000 corresponding values from the two data sets under consideration.

Table E.1: Values of original counts, mean replicate counts, expected counts, RR and ϕ for fifty randomly selected drug and side-effect pairs using C-IG model and Data 1.

Count	Mean Rep. Count	Expected Count	RR	ϕ
7	6.7	11.218	0.62	0.60
6	5.3	5.973	1.00	0.90
0	0.0	0.004	0.00	1.04
666	665.2	695.539	0.96	0.96
0	0.4	0.915	0.00	0.50
44	42.4	6.070	7.25	7.00
0	0.0	0.039	0.00	0.93
111	109.4	0.721	154.02	152.10
0	0.00	0.030	0.00	0.92
28	26.7	0.031	889.14	848.22
32	30.4	0.047	685.52	652.89
7	6.2	6.866	1.02	0.92
1	0.9	1.245	0.80	0.70
1	0.8	0.997	1.00	0.80
1	0.7	0.765	1.31	0.91
1	0.5	0.369	2.71	1.43
1	0.7	0.820	1.22	0.87
0	0.2	0.339	0.00	0.63
26	24.6	0.037	707.23	668.36
7	5.6	0.018	399.26	320.48
6	5.2	4.814	1.25	1.08
2022	2021.1	132.310	15.28	15.27
6	4.7	0.010	576.05	451.76
1021	1021.2	1701.710	0.60	0.60
637	635.7	606.791	1.05	1.05
1211	1209.4	739.652	1.64	1.64
933	932.4	854.123	1.09	1.09
1039	1037.8	380.965	2.73	2.73
56	54.4	0.147	381.61	371.58
1	0.5	0.409	2.44	1.30
46	44.5	0.073	628.21	610.41
2	1.5	1.625	1.23	0.94
3664	3663.3	325.306	11.26	11.26
1608	1607.9	1421.430	1.13	1.13
82	80.7	0.105	780.57	768.57
1	0.3	0.206	4.85	1.82
1032	1030.1	519.620	1.99	1.98
594	592.9	392.751	1.51	1.51
3	2.3	2.468	1.22	0.96
2	1.2	0.845	2.37	1.50
66	65.0	55.168	1.20	1.18
0	0.8	2.436	0.00	0.35
19	20.6	93.931	0.20	0.22
306	303.9	0.198	1543.39	1534.57
306	303.9	0.198	1543.39	1534.57
6	4.5	0.004	1577.25	1221.66
10	8.6	0.023	431.67	373.13
1	0.3	0.198	5.05	1.84
3	2.2	1.432	2.09	1.54
5	4.5	4.914	1.02	0.91

Table E.2a: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	2608.96	2037.79	1	1
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1572.53	1405.13	3	2
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	1532.35	1371.54	4	3
TROLAMINE	OTITIS EXT	14	0.007	1895.68	947.84	4	4	1643.85	939.7	2	4
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	711.35	644.78	11	5
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	764.01	615.81	10	6
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	776.1	615.65	9	7
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	674.91	602.98	13	8
HYDROCORTISONE-NEOMYCIN-POLYMYXIN B	PAIN EAR	37	0.041	895.62	629.35	13	6	848.96	597.62	7	9
BSS	OPHTHALMITIS	28	0.031	889.14	571.59	14	10	833.56	571.76	8	10
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	650.64	555.34	15	11
PANCRELIPASE	STENO INTEST COLON	34	0.047	727.21	491.94	18	13	686.62	486.88	12	12
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	627.95	486.86	18	13
COLFOSCERIL PALMITATE	HEM LUNG	41	0.06	681.25	481.86	25	15	650.23	476.81	16	14
KOATE	HIV SYND	7	0.004	1578.59	451.03	7	17	1160.35	453.69	5	15
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	604.22	452.32	19	16
BSS PLUS	OPHTHALMITIS	32	0.047	685.52	449.87	24	20	642.83	445.76	17	17
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	573.9	440.61	20	18
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	486.96	438.27	23	19
CHLOROXINE	HAIR DISCOLOR	26	0.037	707.23	462.42	21	16	656.26	437.09	14	20
HEMOFIL M	HEPATITIS HBSAG	6	0.004	1577.25	525.75	8	12	1103.44	406.05	6	21
METIPRANOLOL	IRITIS	22	0.035	623.42	368.39	34	22	573.82	373.22	21	22
SODIUM HYALURONATE	IRITIS	21	0.038	551.61	341.47	39	23	501.76	324.79	22	23
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	343.8	305.04	45	24
SODIUM HYALURONATE	OPHTHALMITIS	20	0.038	529.86	317.91	43	24	476.61	301.07	25	25
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	351.69	299.79	43	26
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	334.42	292.53	47	27
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	389.08	290.55	33	28
LOTRISONE	SKIN STRIAE	17	0.032	526.76	278.87	44	32	469.59	284.78	27	29
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	368.36	282.62	37	30
EMBOLEX	VASOSPASM	19	0.038	505.85	292.86	45	29	455.23	275.07	28	31
MIVACURIUM CHLORIDE	CHOLINESTERASE DEC	16	0.032	502.05	282.4	46	31	446.01	261.32	29	32
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	310.16	257.62	49	33
MYSTECLIN F	ANOMALY TOOTH	28	0.07	401.12	257.86	66	35	374.86	257.4	35	34

Table E.2b: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
DOXACURIUM CHLORIDE	PARALYSIS FLACCID	16	0.033	484.79	272.69	49	33	429.05	257.09	30	35
DOXYLAMINE	HANGOVER	44	0.125	351.27	247.48	88	37	335.38	248.7	46	36
BENZOCAINE	METHEMOGLOBIN	29	0.077	376.93	246.96	74	38	354.55	245.26	41	37
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	288.33	244.31	62	38
PHENIRAMINE	MYDRIASIS	26	0.067	387.95	253.66	70	36	359.88	243.81	39	39
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	292.55	241.33	58	40
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	318.19	238.29	48	41
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	292.93	238.18	56	42
DESOXIMETASONE	SKIN STRIAE	10	0.017	590.6	236.24	36	45	483.42	237.74	24	43
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	297.29	236.07	53	44
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	288.26	229.95	63	45
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	297.65	229.07	52	46
PANCRELIPASE	STENO INTEST	16	0.038	423.38	238.15	61	44	374.44	216.86	36	47
SODIUM HYALURONATE	CORNEAL OPACITY	30	0.094	319.8	213.2	103	49	300.87	209.61	50	48
APROTININ BOVINE	THROM CORONARY	13	0.029	452.15	208.68	57	50	387.27	209.08	34	49
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	233.96	206.42	80	50
OXYTOCIN	UTER ATONY	32	0.105	306.03	200.83	117	52	289.16	201.76	61	51
BUDESONIDE	NASAL SEPTUM DIS	28	0.091	309.31	198.84	113	53	289.59	197.76	60	52
ISOSORBIDE MONONITRATE	HEADACHE VASC	40	0.141	283.71	198.59	140	54	269.48	197.7	68	53
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	248.36	195.52	76	54
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	212.94	193.03	96	55
HEMOFIL	HIV SYND	22	0.069	318.09	187.96	106	60	291.53	192.07	59	56
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	217.74	191.28	93	57
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	245.44	190.67	78	58
HEPARIN SODIUM IN DEXTROSE	ANTICOAG DEC	29	0.105	275.05	180.21	147	62	257.23	180.52	73	59
PYRIDOSTIGMINE	CHOLINERG SYND	10	0.023	439.56	175.82	58	65	360.24	179.2	38	60
CHYMOTRYPSIN	KERATITIS	22	0.072	304.72	180.06	119	63	280.06	178.41	66	61
HEMOFIL M	HIV SYND	10	0.023	431.67	172.67	60	68	353.62	177.5	42	62
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	230.42	176.37	82	63
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	203.98	174.39	99	64
DELADUMONE OB	HIRSUTISM	21	0.071	295.96	183.21	129	61	269.81	172.08	67	65
NORGESTREL	CARCINOMA CERVIX SIT	6	0.009	677.3	225.77	27	48	471.96	169.3	26	66
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	198.75	167.35	100	67

Table E.2c: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
CARBOPROST	HEM POSTPARTUM	14	0.041	339.82	169.91	93	69	297.68	167.22	51	68
HYALURONIDASE	STRABISMUS	32	0.13	246.93	162.05	180	72	232.42	164.4	81	69
CREON	STENO INTEST COLON	6	0.01	576.05	192.02	37	57	408.72	156.41	31	70
BSS	CORNEAL OPACITY	21	0.077	271.35	167.98	150	71	248.2	155.79	77	71
BSS PLUS	UVEITIS	17	0.059	287.91	152.43	134	80	255.38	153.69	74	72
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	162.64	153.4	129	73
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	158.32	153.02	139	74
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	170.15	152.56	120	75
BUSULFAN	VENOOCCLUS LIVER SYN	21	0.081	260.1	161.01	160	73	237	152.23	79	76
FLUCLOXACILLIN	STENO INTEST COLON	8	0.017	468.79	175.8	51	66	357.85	151.89	40	77
LOTRISONE	ATROPHY SKIN	26	0.109	237.95	155.58	197	76	220.58	147.76	92	78
METIPRANOLOL	UVEITIS	25	0.105	238.78	152.82	193	78	221.14	147.7	91	79
HETASTARCH	HEMOTHORAX	21	0.083	253.02	156.63	170	75	229.02	145.26	83	80
LOTRISONE	DERM FUNG	15	0.053	285.44	152.24	138	82	250.9	144.71	75	81
DORNASE ALFA	HEMOPTYSIS	64	0.336	190.37	145.75	301	83	184.6	143.3	110	82
BSS PLUS	IRITIS	13	0.042	308.04	142.17	115	86	266.1	143.02	69	83
SOMATREM	NEOPL CNS	52	0.273	190.7	139.36	299	88	184.28	141.56	111	84
FLUNISOLIDE	NASAL SEPTUM DIS	75	0.431	174.04	136.91	344	90	170.16	136.79	119	85
PROTIRELIN	URIN URGENCY	9	0.025	357.13	158.72	80	74	283.38	136.12	65	86
CHOLINE	KERATITIS	116	0.703	164.99	135.12	372	91	162.61	136.01	130	87
CHYMOTRYPSIN	CORNEAL OPACITY	11	0.035	311.16	141.44	112	87	261.72	135.37	72	88
PRILOCAINE	HYPALGESIA	21	0.09	234.17	144.96	201	85	214.75	134.49	95	89
NORGESTREL	NEOPL CERVIX	5	0.008	636.53	127.31	32	99	400.76	133.84	32	90
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	437	94	148.03	133.03	154	91
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	510	95	136.88	130.13	167	92
CHOLINE	IRITIS	22	0.101	217.17	128.33	227	96	198.23	127.74	101	93
BENDECTIN	ECTROMELIA	111	0.721	154.02	126.27	415	100	151.51	126.51	147	94
CARBACHOL	CORNEAL OPACITY	22	0.104	211.66	125.07	239	101	194.22	125.31	105	95
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	552	104	131.36	124.07	177	96
PANALBA K.M.	ANOMALY TOOTH	8	0.022	369.52	138.57	76	89	284.03	123.34	64	97
CHYMOTRYPSIN	IRITIS	7	0.018	399.26	114.08	67	114	294.92	122.91	55	98
MAXITROL	HEALING ABNORM	14	0.056	248.73	124.36	176	102	216.86	120.96	94	99
BETHANECHOL CHLORIDE	CHOLINERG SYND	12	0.045	269.03	134.52	152	92	228.4	120.92	84	100

E.2 Data 2

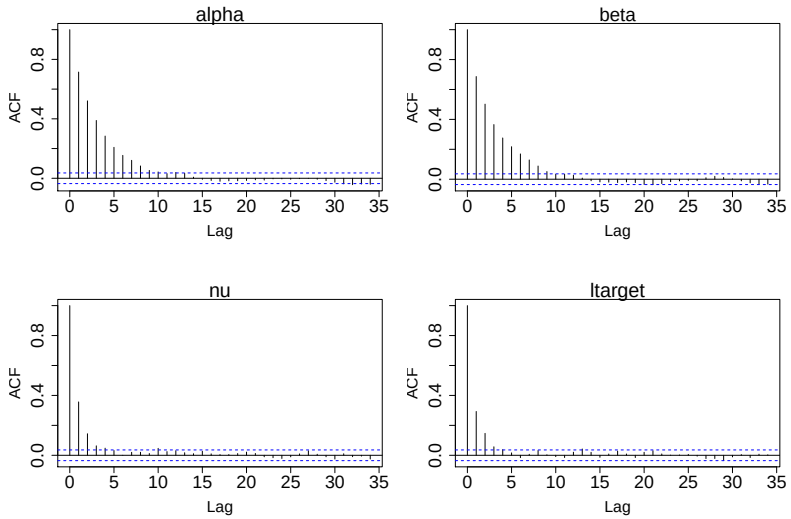


Figure E.7: Acf plots of φ, ψ, ν and the logarithm of the target distribution for C-IG and Data 2.

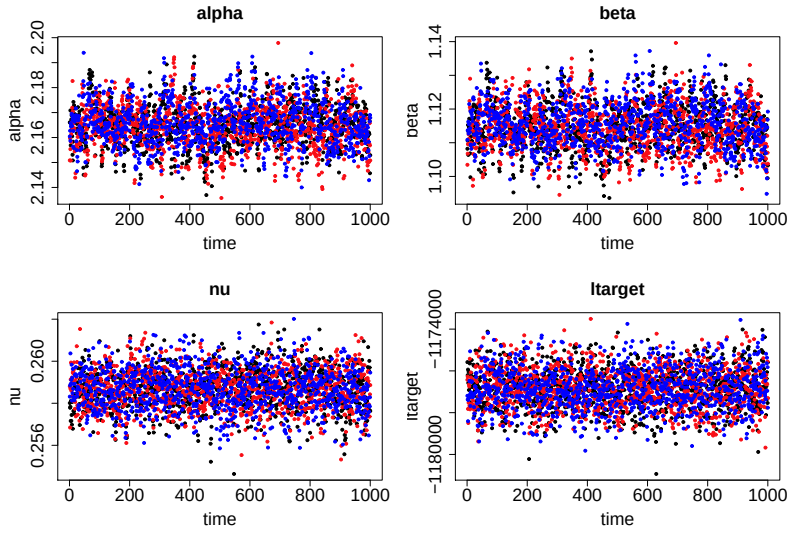
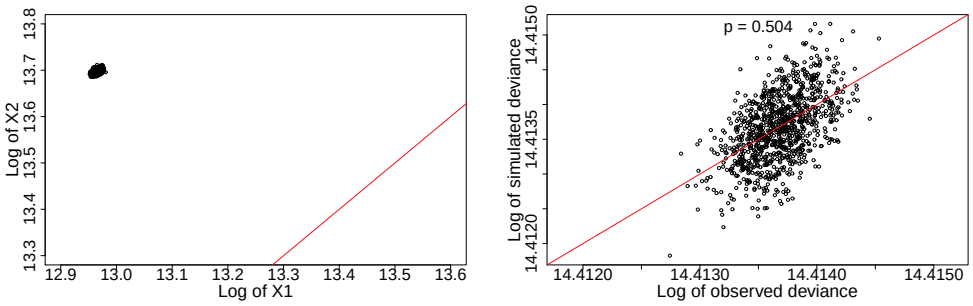
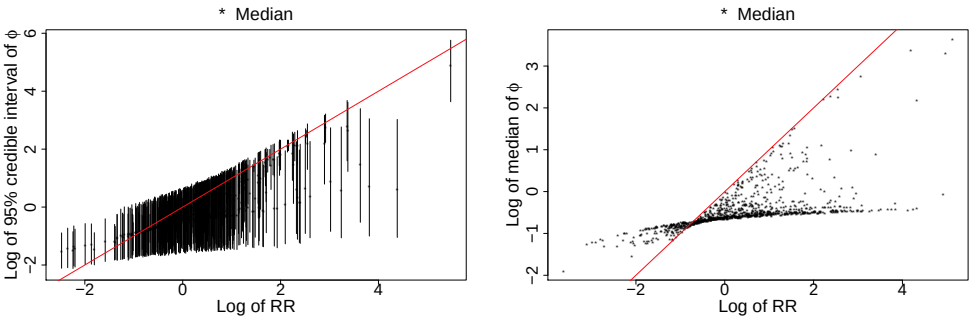


Figure E.8: Trace plots of φ, ψ, ν and the logarithm of the target distribution for three chains, for C-IG and Data 2.



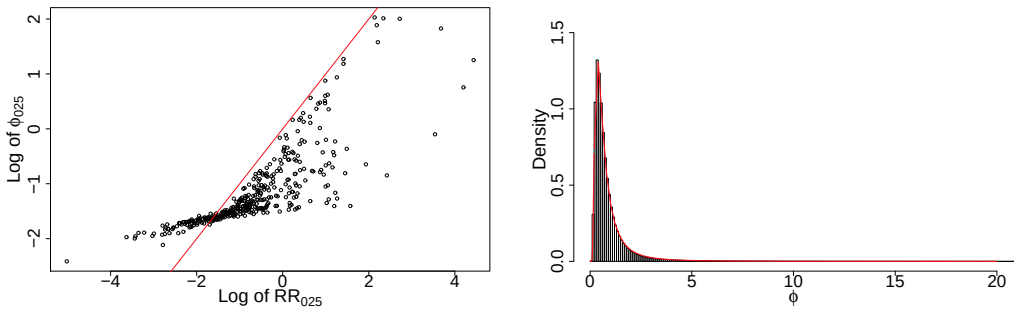
(a) (b)

Figure E.9: Bayesian p-value scatter plots for C-IG and Data 2. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.



(a) (b)

Figure E.10: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for the C-IG model. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 2, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for C-IG model. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 2.



(a) (b)

Figure E.11: (a) Scatter plot of logarithm of ϕ_{025} against the logarithm of RR_{025} for C-IG and Data 2, and (b) Histogram of ϕs for C-IG and Data 2. The superimposed density curve (in red) represents the Gamma distribution corresponding to the MCMC estimates of the posterior means of φ and ψ.

Table E.3: Values of original counts, mean replicate counts, expected counts, RR and ϕ for fifty randomly selected drug and side-effect pairs using C-IG a model and Data 2.

Count	Mean Rep. Count	Expected Count	RR	ϕ
1	0.1	0.000	2816.96	1.15
18	12.7	5.957	3.02	1.87
3	1.4	0.339	8.84	1.18
1	0.2	0.001	694.57	1.14
175	168.7	43.911	3.99	3.80
64	58.8	42.377	1.51	1.36
527	519.3	79.506	6.63	6.52
2	4.2	4.698	0.43	0.55
1	1.4	0.471	2.12	0.79
5	5.5	5.663	0.88	0.67
15	11.4	8.162	1.84	1.20
22	16.8	9.189	2.39	1.65
579	573.7	452.946	1.28	1.26
1481	1473.2	54.248	27.30	27.11
3	3.6	3.064	0.98	0.70
22	18.7	17.797	1.24	0.97
58	49.2	0.098	594.08	492.22
12	19.0	53.081	0.23	0.33
803	794.9	142.740	5.63	5.55
7	4.4	2.600	2.69	1.13
6	0.6	0.010	576.05	8.14
1	1.8	0.880	1.14	0.73
1	0.2	0.002	559.28	1.17
1	0.9	0.160	6.25	0.89
10	2.1	0.017	590.60	61.75
1	1.5	0.616	1.62	0.78
150	141.6	0.228	658.94	617.70
13	11.0	10.760	1.21	0.88
1	1.5	0.570	1.75	0.81
1	3.1	2.780	0.36	0.57
7	2.4	0.418	16.76	2.74
46	37.7	0.073	628.21	496.75
1	2.3	1.519	0.66	0.66
1	2.1	1.298	0.77	0.68
604	598.4	532.731	1.13	1.12
1	2.4	1.708	0.59	0.65
8	6.3	5.231	1.53	0.91
1	1.6	0.597	1.67	0.78
1006	999.8	749.688	1.34	1.33
1196	1195.5	2962.320	0.40	0.40
1094	1087.5	202.805	5.39	5.35
642	638.1	775.209	0.83	0.82
773	764.6	62.563	12.36	12.19
1	1.4	0.431	2.32	0.79
1	3.2	3.125	0.32	0.56
37	28.7	0.041	895.62	656.98
1	2.9	2.429	0.41	0.60
1	2.4	1.623	0.62	0.65
3	3.8	3.550	0.85	0.66
18	18.3	30.366	0.59	0.56

Table E.4a: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 2. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	2209.21	1263.23	1	1
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1526.45	1202.43	2	2
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	1488.24	1181.98	3	3
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	692.25	567.73	4	4
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	653.88	517.17	7	5
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	609.31	433.17	9	6
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	670.42	396.29	5	7
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	665.22	391.01	6	8
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	473.51	382.77	14	9
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	520.5	282.58	10	10
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	473.31	259.61	15	11
HYDROCORTISONE-NEOMYCIN-POLYMYXIN B	PAIN EAR	37	0.041	895.62	629.35	13	6	610.71	256.65	8	12
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	329.98	252.26	20	13
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	319.86	242.12	22	14
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	469.94	228.61	16	15
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	324.94	223.41	21	16
COLFOSCERIL PALMITATE	HEM LUNG	41	0.06	681.25	481.86	25	15	487.53	219.13	12	17
PANCRELIPASE	STENO INTEST COLON	34	0.047	727.21	491.94	18	13	477.99	194.9	13	18
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	268.19	186.37	27	19
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	280.13	183.25	26	20
BSS	OPHTHALMITIS	28	0.031	889.14	571.59	14	10	516.13	173.51	11	21
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	262.41	171.23	28	22
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	224.5	169.18	37	23
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	207.55	168.98	43	24
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	303.04	167.59	24	25
BSS PLUS	OPHTHALMITIS	32	0.047	685.52	449.87	24	20	434.12	166.14	17	26
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	312.59	165.81	23	27
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	260.64	162.5	29	28
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	208.1	157.54	42	29
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	251.46	151.62	32	30
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	255.01	149.65	31	31
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	157.77	147.12	58	32
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	161.1	143.45	56	33
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	192.24	139.33	49	34

Table E.4b: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 2. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	165.07	131.08	55	35
DOXYLAMINE	HANGOVER	44	0.125	351.27	247.48	88	37	257.63	127.91	30	36
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	183.4	126.25	51	37
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	213.07	125.13	41	38
CHLOROXINE	HAIR DISCOLOR	26	0.037	707.23	462.42	21	16	385.28	124.2	18	39
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	236.67	123.41	35	40
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	510	95	136.05	122.3	68	41
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	246.34	119.29	33	42
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	203.91	117.04	44	43
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	552	104	130.3	115.88	71	44
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	437	94	143.47	113.55	65	45
SELENIUM SULFIDE	HAIR DISCOLOR	301	2.236	134.64	119.43	526	107	129.52	102.82	74	46
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	187.8	100.59	50	47
OPCON A	PAIN EYE	267	2.015	132.49	116.61	547	109	126.67	99.23	77	48
CHOLINE	KERATITIS	116	0.703	164.99	135.12	372	91	146.76	98.62	63	49
PHENFORMIN	ACIDOSIS LACTIC	390	3.302	118.11	106.61	681	124	114.82	94.08	87	50
BENDECTIN	ECTROMELIA	111	0.721	154.02	126.27	415	100	137.99	91.63	66	51
POLIOVIRUS VACCINE, LIVE, ORAL	SIDS	146	1.038	140.6	118.45	484	108	128.86	91.31	75	52
ISOSORBIDE MONONITRATE	HEADACHE VASC	40	0.141	283.71	198.59	140	54	200.35	90.61	45	53
COPPER	UTER DIS	4457	46.398	96.06	93.26	975	151	95.9	90.42	105	54
TETRAHYDROZOLINE	PAIN EYE	187	1.43	130.77	112.59	557	117	122.51	89.88	79	55
DORNASE ALFA	HEMOPTYSIS	64	0.336	190.37	145.75	301	83	155.66	89.08	59	56
FLUNISOLIDE	NASAL SEPTUM DIS	75	0.431	174.04	136.91	344	90	146.77	87.84	62	57
CEFOXITIN	ENTEROCOL PSEUDOMEM	138	1.016	135.79	113.16	517	116	124.08	87.31	78	58
DTP VACCINE	SCREAMING SYND	966	9.643	100.18	93.95	897	147	99.1	87.28	103	59
MENOTROPINS	OVAR DIS	151	1.175	128.48	108.06	583	122	118.76	85.15	84	60
GONADOTROPIN, CHORIONIC	OVAR DIS	86	0.567	151.69	119.94	423	106	129.83	82.41	73	61
POLIOVIRUS VACCINE, LIVE, ORAL	SCREAMING SYND	344	3.28	104.89	93.91	831	148	101.51	81.8	98	62
MYSTECLIN F	ANOMALY TOOTH	28	0.07	401.12	257.86	66	35	229.43	79.83	36	63
MITOMYCIN	UREMIA	218	1.995	109.29	95.26	785	142	103.35	78.93	96	64
BENZOCAINE	METHEMOGLOBIN	29	0.077	376.93	246.96	74	38	224.38	78.64	38	65
DIPHTHERIA-TETANUS TOXOID-PERTUSSIS VA	SCREAMING SYND	147	1.22	120.47	101.62	654	130	110.78	78.53	91	66
SOMATREM	NEOPL CNS	52	0.273	190.7	139.36	299	88	148.11	76.92	60	67

Table E.4c: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 2. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
LOMEFLOXACIN HYDROCHLORIDE	PHOTOSENSITIVITY	481	5.137	93.64	85.27	1023	166	91.48	76.41	108	68
OXYTOCIN	UTER ATONY	32	0.105	306.03	200.83	117	52	195.33	76.39	48	69
METIPRANOLOL	IRITIS	22	0.035	623.42	368.39	34	22	290.01	75.46	25	70
PILOCARPINE	MIOSIS	103	0.812	126.83	103.44	601	128	112.58	73.68	89	71
SODIUM HYALURONATE	CORNEAL OPACITY	30	0.094	319.8	213.2	103	49	195.47	73.36	47	72
OXYTOCIN	FETAL DIS	95	0.759	125.14	100.11	616	133	109.78	72.36	92	73
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM INTRACRAN	476	5.426	87.73	79.99	1150	184	85.33	71.07	120	74
TICE BCG	CYSTITIS	69	0.475	145.35	111.64	457	119	120.82	70.84	80	75
HAEMOPHILUS B POLYSACCHARIDE VACCINE	MENINGITIS	176	1.711	102.86	87.66	851	164	95.67	69.67	106	76
TETRAHYDROZOLINE	CONJUNCTIVITIS	230	2.454	93.73	81.91	1021	174	89.05	68.73	115	77
PHENIRAMINE	MYDRIASIS	26	0.067	387.95	253.66	70	36	213.36	67.65	40	78
UROKINASE	CHILLS	831	10.942	75.95	70.83	1426	220	74.81	65.87	142	79
INFLUENZA VIRUS VACCINE	GUILLAIN BARRE SYND	62	0.451	137.47	104.21	507	127	111.21	62.54	90	80
COPPER	SALPINGITIS	1484	21.323	69.6	66.08	1635	238	69	62.39	153	81
BUDESONIDE	NASAL SEPTUM DIS	28	0.091	309.31	198.84	113	53	178.83	62.25	52	82
OXYTETRACYCLINE	DISCOLOR TOOTH	74	0.599	123.49	96.79	633	138	103.92	61.76	95	83
HEPARIN SODIUM IN DEXTROSE	ANTICOAG DEC	29	0.105	275.05	180.21	147	62	166.91	61.63	54	84
SUPROFEN	PAIN BACK	399	5.188	76.91	69.39	1404	227	74.67	61.14	143	85
RUBELLA VIRUS VACCINE, LIVE	LYMPHADENO	115	1.162	98.97	80.9	925	179	88.62	60.71	116	86
NONOXYNOL	LEUKORRHEA	603	8.344	72.27	66.51	1516	235	70.91	60.24	147	87
MICONAZOLE	VAGINITIS	2262	34.851	64.9	62.24	1818	271	64.59	59.65	175	88
INTRAUTERINE DEVICE	UTER DIS	251	3.127	80.26	70.35	1328	224	76.64	59.6	138	89
HYALURONIDASE	STRABISMUS	32	0.13	246.93	162.05	180	72	157.88	59.14	57	90
OXYTOCIN	LABOR ABNORM	105	1.043	100.68	81.5	887	176	89.69	59.12	112	91
FLOSEQUINAN	SUDDEN DEATH	49	0.315	155.79	114.46	404	111	118.51	59.03	85	92
COPPER	DEVICE MIGRATION	1490	22.868	65.16	61.88	1811	272	64.69	58.17	174	93
COPPER	PAIN PELVIC	1962	30.696	63.92	61.12	1861	278	63.51	58.17	182	94
UROFOLLITROPIN	OVAR DIS	46	0.289	159.3	114.28	394	112	119.27	57.16	82	95
ALCOHOL	ALCOHOL INTOLER	169	2.006	84.26	71.8	1233	215	78.18	56.96	133	96
THIMEROSAL	DERM CONTACT	62	0.501	123.84	93.88	631	149	100.59	56.72	101	97
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM CEREBR	695	10.419	66.71	61.81	1744	274	65.57	56.49	169	98
SODIUM HYALURONATE	IRITIS	21	0.038	551.61	341.47	39	23	246.33	56.47	34	99
TIOCONAZOLE	VAGINITIS	268	3.558	75.32	66.33	1442	236	71.82	56.17	146	100

E.3 Data 3

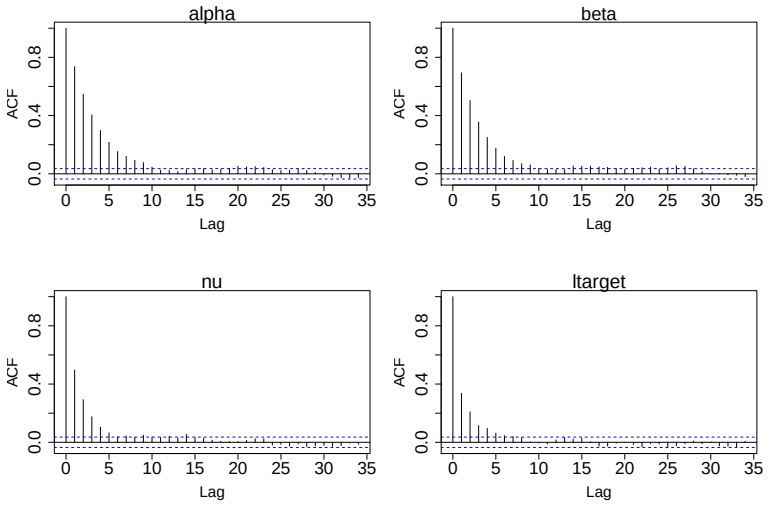


Figure E.12: Acf plots of φ, ψ, ν and the logarithm of the target distribution for C-IG and Data 3.

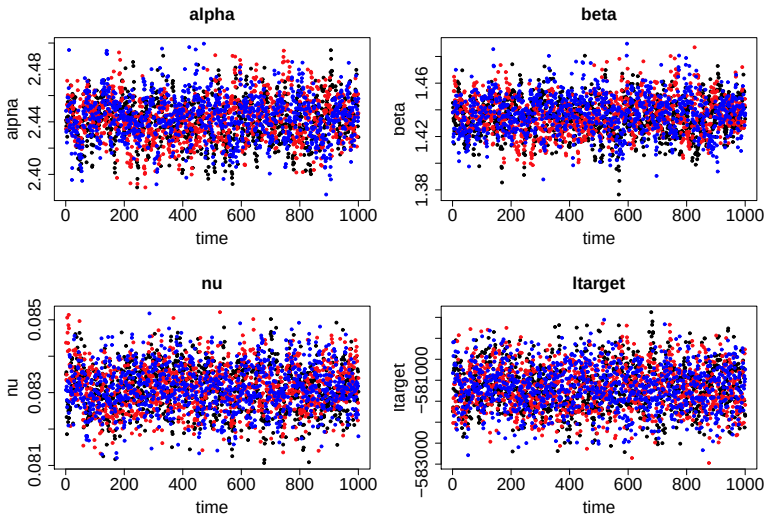
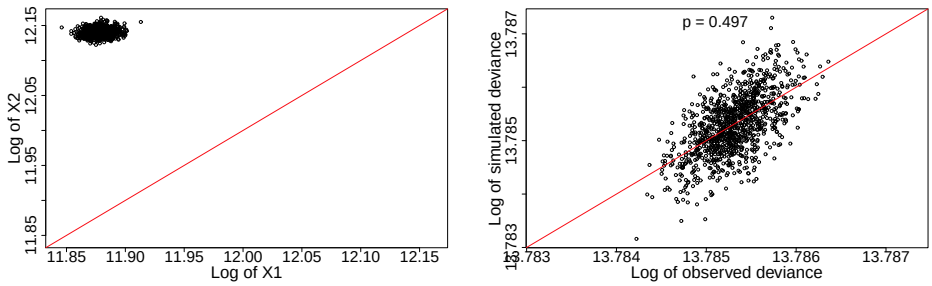
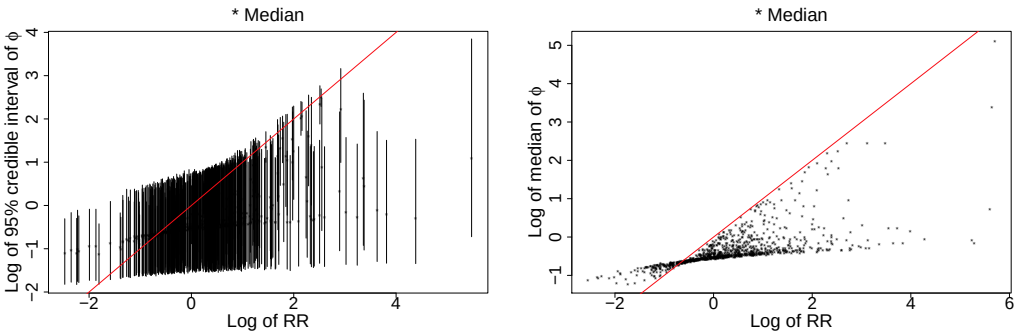


Figure E.13: Trace plots of φ, ψ, ν and the logarithm of the target distribution for three chains, for C-IG and Data 3.



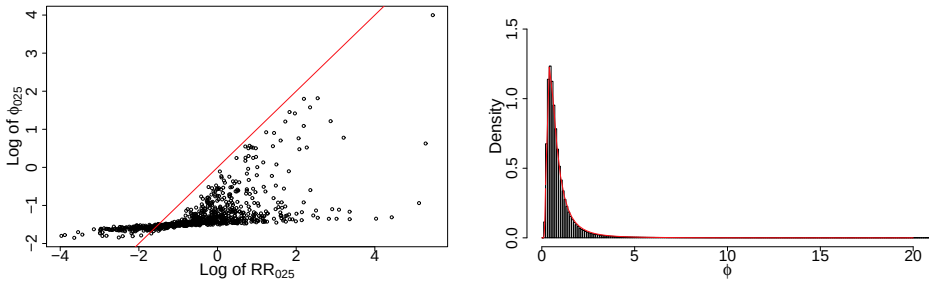
(a) (b)

Figure E.14: Bayesian p-value scatter plots for C-IG and Data 3. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.



(a) (b)

Figure E.15: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for the C-IG model. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 3, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for C-IG model. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 3.



(a) (b)

Figure E.16: (a) Scatter plot of logarithm of ϕ_{025} against the logarithm of RR_{025} for C-IG and Data 3, and (b) Histogram of ϕs for C-IG and Data 3. The superimposed density curve (in red) represents the Gamma distribution corresponding to the MCMC estimates of the posterior means of φ and ψ.

Table E.5: Values of original counts, mean replicate counts, expected counts, RR and ϕ for fifty randomly selected drug and side-effect pairs using C-IG model and Data 3.

Count	Mean Rep. Count	Expected Count	RR	ϕ
10	8.7	3.771	2.65	0.94
684	661.6	141.379	4.84	4.62
23	21.6	19.812	1.16	0.80
527	513	477.337	1.10	1.06
89	82.5	88.794	1.00	0.87
527	500.6	79.506	6.63	6.22
28	4.0	0.031	889.14	26.41
1109	1082.5	162.189	6.84	6.64
5	8.9	5.020	1.00	0.74
3164	3133.4	69.702	45.39	44.88
6	9.8	6.055	0.99	0.73
10	12.7	9.937	1.01	0.74
53	33.4	11.698	4.53	2.36
26	3.7	0.037	707.23	16.89
9	12.5	9.014	1.00	0.72
58	28.4	0.098	594.08	231.22
896	877.5	533.355	1.68	1.63
107	99.7	106.067	1.01	0.88
151	141.6	142.483	1.06	0.95
9	12.1	8.989	1.00	0.73
6	6.8	2.228	2.69	0.88
11	13.5	11.050	1.00	0.73
199	170.9	20.056	9.92	8.26
26	37.5	65.969	0.39	0.48
1062	1049.7	904.794	1.17	1.15
126	107.9	58.529	2.15	1.74
20	27.6	38.710	0.52	0.57
695	681.5	684.336	1.02	0.99
5	8.4	4.241	1.18	0.75
8	9.9	5.329	1.50	0.80
17	14.5	9.548	1.78	0.92
7	10.7	7.001	1.00	0.73
10	7.8	2.704	3.70	1.01
1404	1379.2	455.788	3.08	3.02
544	557.4	1218.600	0.45	0.45
557	558.3	917.305	0.61	0.60
1400	1375.5	472.530	2.96	2.91
843	827.9	756.036	1.12	1.09
26	32.9	46.838	0.56	0.58
579	562.2	198.300	2.92	2.78
580	567.4	505.762	1.15	1.11
9	12.1	9.005	1.00	0.73
306	275.9	0.198	1543.39	1365.45
6	1.3	0.004	1577.25	1.29
14	12.0	6.678	2.10	0.94
55	65.5	119.829	0.46	0.50
387	356.4	0.542	714.05	648.96
293	261.8	0.185	1583.73	1387.17
11	13.8	11.036	1.00	0.72
10	12.6	9.043	1.11	0.76

Table E.6a: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate estimate of ϕ, using Data 3. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	3	2	1356.75	871.02	1	1
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	6	3	1333.86	853.83	2	2
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	12	5	638.13	438.99	4	3
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	16	9	581.32	371.57	5	4
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	31	21	431.71	288.62	7	5
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	18	11	477.72	234.56	6	6
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	1	1	998.79	201.74	3	7
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	52	25	291.08	179.18	10	8
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	54	27	271.75	156.86	11	9
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	165	79	156.05	137.07	23	10
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	111	56	191.89	131.51	16	11
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	158	77	157.32	128.63	22	12
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	10	7	390.71	124.08	8	13
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	49	26	246.26	112.45	12	14
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	99	51	193.85	111.88	15	15
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	198	94	133.12	110.23	29	16
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	106	58	181.3	105.67	18	17
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	212	103	126.34	102.11	31	18
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	148	81	149.07	97.07	25	19
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	73	39	202.06	96.26	14	20
COPPER	UTER DIS	4457	46.398	96.06	93.26	321	150	95.16	85.86	43	21
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	174	93	129.9	83.73	30	22
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	9	8	338.02	82.78	9	23
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	118	67	154.75	81.19	24	24
DTP VACCINE	SCREAMING SYND	966	9.643	100.18	93.95	303	146	95.66	76.1	42	25
SELENIUM SULFIDE	HAIR DISCOLOR	301	2.236	134.64	119.43	203	106	116.54	75.44	32	26
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	59	34	191.45	72.61	17	27
PHENFORMIN	ACIDOSIS LACTIC	390	3.302	118.11	106.61	247	123	105.65	72.17	37	28
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	70	40	178.46	72.02	19	29
OPCON A	PAIN EYE	267	2.015	132.49	116.61	210	108	112.86	70.38	33	30
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	123	70	140.69	63.91	26	31
POLIOVIRUS VACCINE, LIVE, ORAL	SCREAMING SYND	344	3.28	104.89	93.91	285	147	92.37	60.68	46	32
LOMEFLOXACIN HYDROCHLORIDE	PHOTOSENSITIVITY	481	5.137	93.64	85.27	335	164	85.35	60.59	50	33
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM INTRACRAN	476	5.426	87.73	79.99	364	182	79.78	56.98	53	34

Table E.6b: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 3. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
TETRAHYDROZOLINE	PAIN EYE	187	1.43	130.77	112.59	214	116	101.88	56.7	38	35
UROKINASE	CHILLS	831	10.942	75.95	70.83	426	216	72.02	56.37	58	36
COPPER	SALPINGITIS	1484	21.323	69.6	66.08	487	233	67.63	56.35	62	37
MICONAZOLE	VAGINITIS	2262	34.851	64.9	62.24	532	265	63.67	54.89	70	38
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	69	43	160.72	54.48	21	39
COPPER	PAIN PELVIC	1962	30.696	63.92	61.12	543	272	62.58	53.61	74	40
COPPER	DEVICE MIGRATION	1490	22.868	65.16	61.88	530	266	63.28	52.65	71	41
MITOMYCIN	UREMIA	218	1.995	109.29	95.26	275	141	88.46	50.96	48	42
NONOXYNOL	LEUKORRHEA	603	8.344	72.27	66.51	457	231	67.49	49.96	63	43
POLIOVIRUS VACCINE, LIVE, ORAL	SIDS	146	1.038	140.6	118.45	189	107	101.54	49.77	39	44
TISSUE PLASMINOGEN ACTIVATOR, RECOMBIN	HEM CEREBR	695	10.419	66.71	61.81	514	268	63.05	48.14	72	45
SUPROFEN	PAIN BACK	399	5.188	76.91	69.39	421	223	69.02	46.89	61	46
MENOTROPINS	OVAR DIS	151	1.175	128.48	108.06	221	121	93.37	46.63	45	47
CEFOXITIN	ENTEROCOL PSEUDOMEM	138	1.016	135.79	113.16	200	115	96.23	46.06	41	48
COPPER	FERTIL DEC FEM	695	10.997	63.2	58.56	553	287	59.62	45.87	80	49
CHOLINE	KERATITIS	116	0.703	164.99	135.12	156	91	106.26	45.52	35	50
TETRAHYDROZOLINE	CONJUNCTIVITIS	230	2.454	93.73	81.91	334	172	77.35	45.18	57	51
NONOXYNOL	DEVICE MIGRATION	800	13.275	60.27	56.12	590	304	57.32	44.16	82	52
TRIAMCINOLONE	ATROPHY INJECT SITE	439	6.303	69.65	63.14	483	254	62.81	44.04	73	53
DIPHTHERIA-TETANUS TOXOID-PERTUSSIS VA	SCREAMING SYND	147	1.22	120.47	101.62	236	129	86.71	42.74	49	54
HAEMOPHILUS B POLYSACCHARIDE VACCINE	MENINGITIS	176	1.711	102.86	87.66	291	162	78.74	41.73	55	55
INTRAUTERINE DEVICE	UTER DIS	251	3.127	80.26	70.35	403	220	67.38	41.17	64	56
BENDECTIN	ECTROMELIA	111	0.721	154.02	126.27	168	99	98.46	40.39	40	57
TRAZODONE	PRIAPISM	347	5.117	67.81	60.77	502	276	59.98	40.24	78	58
TIOCONAZOLE	VAGINITIS	268	3.558	75.32	66.33	431	232	63.79	39.65	67	59
CEFACLOR	SERUM SICK	3164	69.702	45.39	43.81	845	426	44.79	39.51	108	60
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	71	46	139.96	39.01	27	61
BSS	KERATITIS	74	0.243	305.12	239.15	67	42	137.5	38.56	28	62
NAFARELIN ACETATE	HEM VAGINAL	506	9.008	56.17	51.29	636	341	51.7	37.64	88	63
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	19	14	207.29	36.82	13	64
BUTOCONAZOLE NITRATE	VAGINITIS	1072	22.835	46.94	44.14	806	422	44.94	36.42	105	65
BUTORPHANOL	DRUG DEPEND	457	8.165	55.97	50.95	639	344	50.95	36.37	90	66
MICONAZOLE	VULVOVAGINITIS	234	3.168	73.87	64.71	442	243	60.55	36.28	76	67

Table E.6c: Top hundred drug and adverse event pairs selected by C-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 3. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
CEFUROXIME	COLITIS PSEUDOMEM	1063	22.881	46.46	43.7	817	428	44.67	36.07	109	68
PROCAINAMIDE	LE SYND	471	8.679	54.27	49.43	666	357	49.63	35.34	92	69
INSULIN NOVOLIN 70/30	HYPERGLYCEM	484	8.959	54.02	49.22	674	360	49.46	35.2	94	70
PRED-G	PAIN EYE	644	13.125	49.07	45.33	765	405	46.02	34.76	103	71
ALCOHOL	ALCOHOL INTOLER	169	2.006	84.26	71.8	381	211	63.74	33.25	69	72
ISOTRETINOIN	CHEILITIS	380	6.983	54.42	48.97	664	364	48.52	33.17	98	73
INSULIN HUMAN	HYPERGLYCEM	6200	170.45	36.37	35.47	1123	545	36.12	33.13	139	74
TRETINOIN	DERM EXFOL	1213	29.963	40.48	38.21	982	494	39.07	31.96	127	75
COPPER	UTER RUPT	270	4.665	57.88	51.02	614	343	48.96	31.34	97	76
PILOCARPINE	MIOSIS	103	0.812	126.83	103.44	226	127	77.47	31.12	56	77
CLONIDINE	DERM CONTACT	964	23.784	40.53	38.01	980	499	38.93	30.58	128	78
CEFIXIME	COLITIS PSEUDOMEM	421	8.596	48.98	44.32	768	420	44.32	30.53	111	79
MINOXIDIL	HIRSUTISM	1035	26.322	39.32	36.93	1006	518	37.91	30.44	133	80
PRILOCAINE	PARESTH CIRCUMORAL	200	2.995	66.77	57.75	513	297	53.04	30.06	86	81
TIMOPTIC	PAIN EYE	695	16.875	41.19	38.16	958	495	38.67	29.48	131	82
RITONAVIR	PARESTH CIRCUMORAL	217	3.566	60.84	52.99	579	325	49.44	28.94	95	83
MINOXIDIL	HAIR DIS	928	24.654	37.64	35.25	1073	548	35.93	28.12	140	84
ENOXAPARIN SODIUM	HEM	342	7.257	47.13	42.17	800	441	41.86	28.04	118	85
RUBELLA VIRUS VACCINE, LIVE	LYMPHADENO	115	1.162	98.97	80.9	307	177	63.78	27.91	68	86
SCOPOLAMINE	MYDRIASIS	269	5.256	51.18	45.09	725	410	43.42	27.87	114	87
DTP VACCINE	SIDS	207	3.495	59.23	51.21	601	342	47.66	27.15	100	88
RISPERIDONE	PROLACTIN INC	360	7.978	45.12	40.49	853	462	39.9	27.05	124	89
TUBERCULIN, PURIFIED PROTEIN DERIVATIV	INJECT SITE REACT	411	9.536	43.1	39.01	899	486	38.88	27.04	129	90
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	23	18	174.03	26.98	20	91
GONADOTROPIN, CHORIONIC	OVAR DIS	86	0.567	151.69	119.94	169	105	80.36	26.87	52	92
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	86	55	105.85	26.66	36	93
BENOXAPROFEN	NAIL DIS	382	8.665	44.09	39.7	879	476	39.32	26.57	125	94
CEFACLOR	ARTHROSIS	1010	30.298	33.34	31.29	1261	653	32.05	25.61	163	95
PHOTOPLEX	PHOTOSENSITIVITY	156	2.359	66.13	55.96	518	307	49.04	25.26	96	96
TRETINOIN	SKIN DRY	750	21.028	35.67	33.15	1154	602	33.62	25.26	151	97
ISOTRETINOIN	BLIND NIGHT	140	1.94	72.15	60.3	458	279	51.33	24.83	89	98
OXYTOCIN	FETAL DIS	95	0.759	125.14	100.11	230	132	71.4	24.62	59	99
NONOXYNOL	VAGINITIS	1141	36.497	31.26	29.45	1352	699	30.13	24.51	179	100

Appendix F

P-IG Model Results

F.1 Data 1

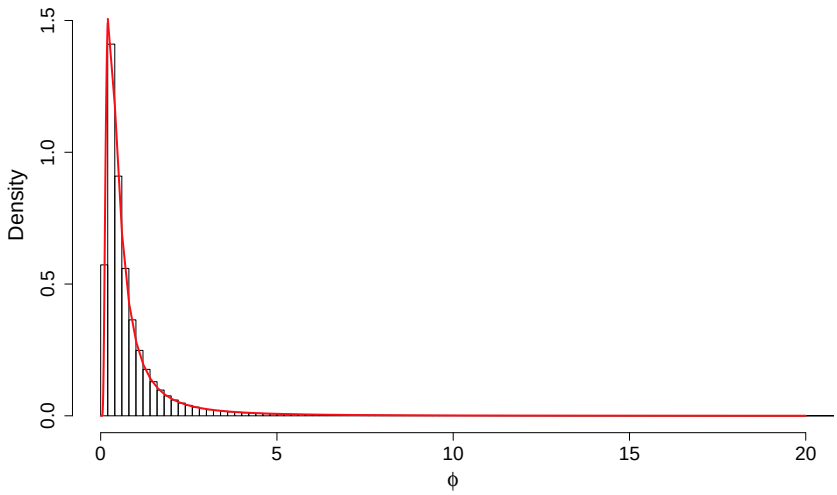


Figure F.1: Histogram of ϕs for P-IG and Data 1. The superimposed density curve (in red) represents the Inverse Gamma distribution corresponding to the MCMC estimates of the posterior means of φ and ψ.

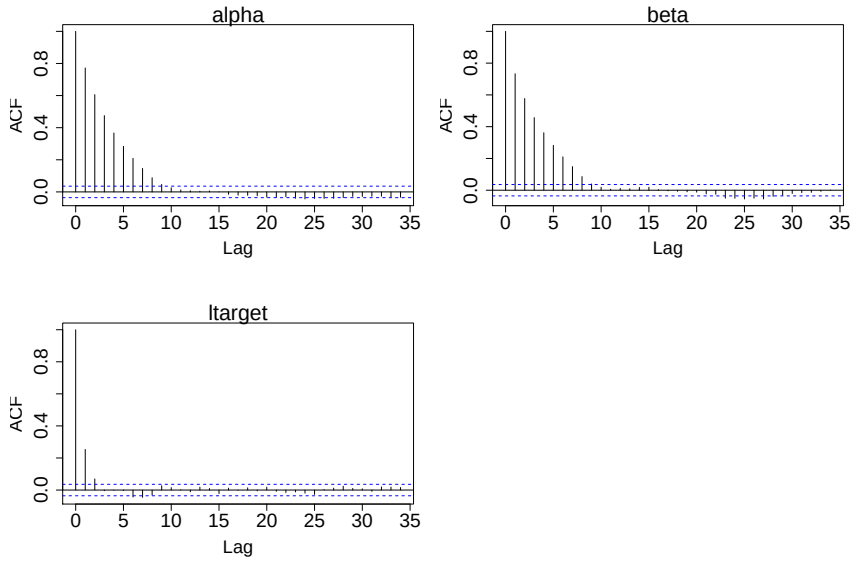


Figure F.2: Acf plots of φ, ψ and the logarithm of the target distribution for P-IG and Data 1.

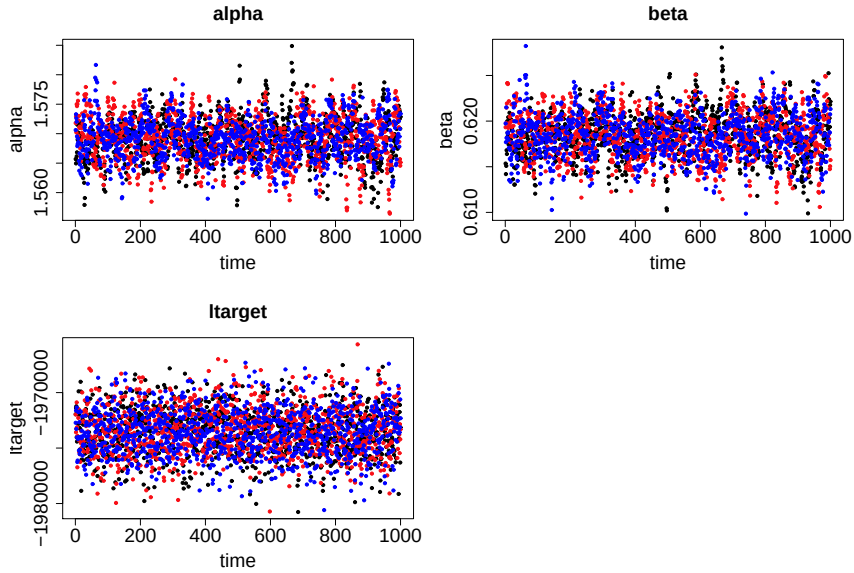
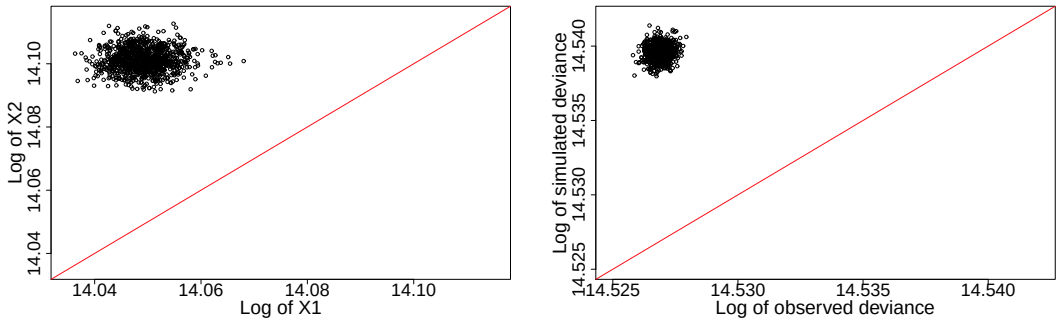
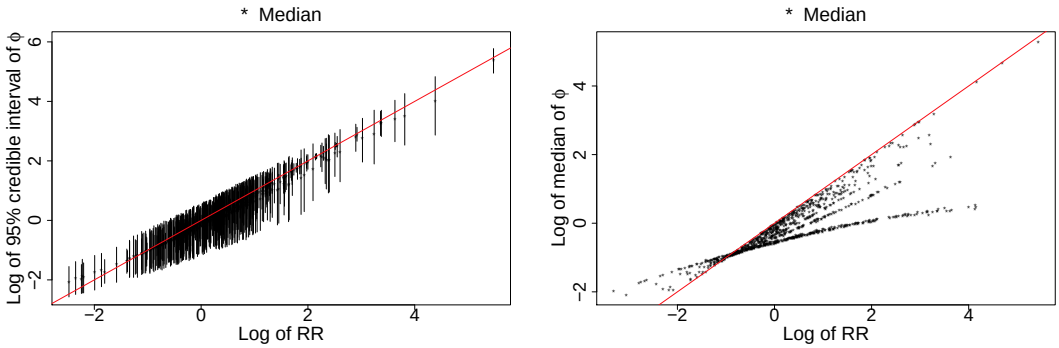


Figure F.3: Trace plots of φ, ψ and the logarithm of the target distribution for three chains, for P-IG and Data 1.



(a) (b)

Figure F.4: Bayesian p-value scatter plots for P-IG and Data 1. (a) Plot of logarithm of X_2 against logarithm of X_1, and (b) Plot of logarithm of SD against logarithm of OD. The red line is the line of equality.



(a) (b)

Figure F.5: (a) Logarithm of the 95% posterior intervals of ϕ against the logarithm of the reporting rate RR for the P-IG model. The plotted values are for 500 drug and adverse event pairs randomly sampled from Data 1, and (b) Logarithm of posterior medians of ϕ against the logarithm of the reporting rate RR for P-IG model. The plotted values are for 1000 drug and adverse event pairs randomly sampled from Data 1.

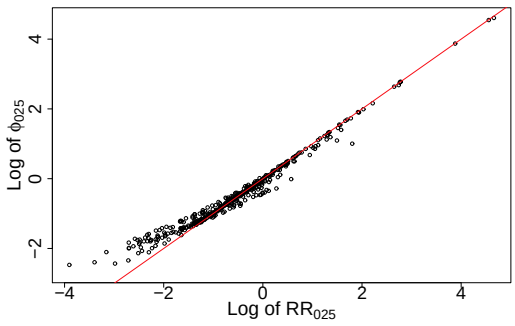


Figure F.6: Scatter plot of logarithm of ϕ_{025} against the logarithm of RR_{025} for P-IG and Data 1.

Table F.1: Values of original counts, mean replicate counts, expected counts, RR and ϕ for fifty randomly selected drug and side-effect pairs using P-IG model and Data 1.

Count	Mean Rep. Count	Expected Count	RR	ϕ
7	6.7	11.218	0.62	0.59
6	5.3	5.973	1.00	0.89
0	0.0	0.004	0.00	1.01
666	664.9	695.539	0.96	0.96
0	0.4	0.915	0.00	0.48
44	42.6	6.070	7.25	7.01
0	0.0	0.039	0.00	0.86
111	109.4	0.721	154.02	151.43
0	0.0	0.030	0.00	0.93
28	26.5	0.031	889.14	840.13
32	30.5	0.047	685.52	653.33
7	6.2	6.866	1.02	0.90
1	0.8	1.245	0.80	0.68
1	0.8	0.997	1.00	0.76
1	0.7	0.765	1.31	0.89
1	0.5	0.369	2.71	1.21
1	0.7	0.820	1.22	0.83
0	0.2	0.339	0.00	0.61
26	24.4	0.037	707.23	665.93
7	5.5	0.018	399.26	310.91
6	5.1	4.814	1.25	1.06
2022	2019.3	132.310	15.28	15.27
6	4.4	0.010	576.05	428.74
1021	1018.9	1701.710	0.60	0.60
637	636.1	606.791	1.05	1.05
1211	1210.2	739.652	1.64	1.64
933	932.0	854.123	1.09	1.09
1039	1039.0	380.965	2.73	2.72
56	54.4	0.147	381.61	370.99
1	0.5	0.409	2.44	1.19
46	44.3	0.073	628.21	605.84
2	1.5	1.625	1.23	0.89
3664	3662.6	325.306	11.26	11.26
1608	1609.1	1421.430	1.13	1.13
82	80.5	0.105	780.57	767.41
1	0.3	0.206	4.85	1.64
1032	1030.7	519.620	1.99	1.98
594	591.3	392.751	1.51	1.51
3	2.3	2.468	1.22	0.93
2	1.2	0.845	2.37	1.38
66	64.9	55.168	1.20	1.18
0	0.8	2.436	0.00	0.34
19	20.3	93.931	0.20	0.22
306	304.6	0.198	1543.39	1535.14
306	304.6	0.198	1543.39	1535.14
6	4.4	0.004	1577.25	1159.17
10	8.6	0.023	431.67	368.76
1	0.3	0.198	5.05	1.63
3	2.1	1.432	2.09	1.44
5	4.3	4.914	1.02	0.88

Table F.2a: Top hundred drug and adverse event pairs selected by P-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	2610.25	2028.38	1	1
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1571.77	1395.81	3	2
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	1532.51	1357.95	4	3
TROLAMINE	OTITIS EXT	14	0.007	1895.68	947.84	4	4	1622.86	899.6	2	4
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	710.74	642.88	11	5
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	776.94	607.99	9	6
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	762.71	606.3	10	7
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	675.26	602.72	13	8
HYDROCORTISONE-NEOMYCIN-POLYMYXIN B	PAIN EAR	37	0.041	895.62	629.35	13	6	840.21	593.15	7	9
BSS	OPHTHALMITIS	28	0.031	889.14	571.59	14	10	823.87	552.5	8	10
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	648.96	551.16	15	11
PANCRELIPASE	STENO INTEST COLON	34	0.047	727.21	491.94	18	13	685.58	476.5	12	12
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	621.95	476.42	18	13
COLFOSCERIL PALMITATE	HEM LUNG	41	0.06	681.25	481.86	25	15	648.52	469.94	16	14
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	599.25	442.55	19	15
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	486.7	440.07	23	16
BSS PLUS	OPHTHALMITIS	32	0.047	685.52	449.87	24	20	642.34	437.67	17	17
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	572.17	432.46	20	18
KOATE	HIV SYND	7	0.004	1578.59	451.03	7	17	1101.95	428.79	5	19
CHLOROXINE	HAIR DISCOLOR	26	0.037	707.23	462.42	21	16	652.62	425.5	14	20
METIPRANOLOL	IRITIS	22	0.035	623.42	368.39	34	22	563.3	351	21	21
HEMOFIL M	HEPATITIS HBSAG	6	0.004	1577.25	525.75	8	12	1032.36	332.92	6	22
SODIUM HYALURONATE	IRITIS	21	0.038	551.61	341.47	39	23	497.35	304.9	22	23
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	342.91	301.29	44	24
SODIUM HYALURONATE	OPHTHALMITIS	20	0.038	529.86	317.91	43	24	476.15	295.4	24	25
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	350.82	292.94	39	26
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	334.84	292.82	45	27
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	386.09	288.58	31	28
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	367.63	278.42	36	29
EMBOLEX	VASOSPASM	19	0.038	505.85	292.86	45	29	452.9	275.52	27	30
LOTRISONE	SKIN STRIAE	17	0.032	526.76	278.87	44	32	464.82	264.53	25	31
MIVACURIUM CHLORIDE	CHOLINESTERASE DEC	16	0.032	502.05	282.4	46	31	442.76	260.44	29	32
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	309.25	251.45	48	33
MYSTECLIN F	ANOMALY TOOTH	28	0.07	401.12	257.86	66	35	370.15	248.78	34	34

Table F.2b: Top hundred drug and adverse event pairs selected by P-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
DOXACURIUM CHLORIDE	PARALYSIS FLACCID	16	0.033	484.79	272.69	49	33	424.33	245.93	30	35
DOXYLAMINE	HANGOVER	44	0.125	351.27	247.48	88	37	334.55	243.9	46	36
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	288.46	241.59	55	37
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	292.42	239.16	53	38
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	292.97	237.57	52	39
PHENIRAMINE	MYDRIASIS	26	0.067	387.95	253.66	70	36	357.44	236.95	38	40
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	295.9	235.04	50	41
BENZOCAINE	METHEMOGLOBIN	29	0.077	376.93	246.96	74	38	348.7	234.31	40	42
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	316.95	231.23	47	43
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	287.99	229.11	56	44
DESOXIMETASONE	SKIN STRIAE	10	0.017	590.6	236.24	36	45	464.79	221.25	26	45
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	293.96	219.97	51	46
PANCRELIPASE	STENO INTEST	16	0.038	423.38	238.15	61	44	367.4	211.05	37	47
SODIUM HYALURONATE	CORNEAL OPACITY	30	0.094	319.8	213.2	103	49	298.1	204.35	49	48
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	233.6	204.25	80	49
APROTININ BOVINE	THROM CORONARY	13	0.029	452.15	208.68	57	50	378.65	201.12	33	50
OXYTOCIN	UTER ATONY	32	0.105	306.03	200.83	117	52	285.89	196.29	57	51
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	247.75	195.7	74	52
ISOSORBIDE MONONITRATE	HEADACHE VASC	40	0.141	283.71	198.59	140	54	269.95	193.04	66	53
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	212.55	191.74	92	54
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	217.65	191.13	89	55
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	244.2	188.46	78	56
BUDESONIDE	NASAL SEPTUM DIS	28	0.091	309.31	198.84	113	53	284.81	188.38	59	57
HEMOFIL	HIV SYND	22	0.069	318.09	187.96	106	60	285.78	180.33	58	58
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	229.5	174.72	82	59
HEPARIN SODIUM IN DEXTROSE	ANTICOAG DEC	29	0.105	275.05	180.21	147	62	255.58	173.52	70	60
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	203.3	173.28	97	61
CHYMOTRYPSIN	KERATITIS	22	0.072	304.72	180.06	119	63	275.24	170.74	62	62
DELADUMONE OB	HIRSUTISM	21	0.071	295.96	183.21	129	61	266.67	170.02	67	63
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	198.5	167.32	99	64
PYRIDOSTIGMINE	CHOLINERG SYND	10	0.023	439.56	175.82	58	65	348.25	165.75	41	65
HEMOFIL M	HIV SYND	10	0.023	431.67	172.67	60	68	346.71	160.79	43	66
HYALURONIDASE	STRABISMUS	32	0.13	246.93	162.05	180	72	231.32	160.67	81	67

Table F.2c: Top hundred drug and adverse event pairs selected by P-IG model based on the lower bound (ϕ_{025}) of the 95% credible interval estimate of ϕ using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	ϕ	ϕ_{025}	RK3	RK33
CARBOPROST	HEM POSTPARTUM	14	0.041	339.82	169.91	93	69	288.75	158.93	54	68
NORGESTREL	CARCINOMA CERVIX SIT	6	0.009	677.3	225.77	27	48	448.56	158.76	28	69
BSS	CORNEAL OPACITY	21	0.077	271.35	167.98	150	71	245.3	153.79	76	70
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	162.68	152.96	122	71
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	169.92	152.56	113	72
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	158.29	152.44	131	73
BUSULFAN	VENOOCCLUS LIVER SYN	21	0.081	260.1	161.01	160	73	234.76	147.26	79	74
BSS PLUS	UVEITIS	17	0.059	287.91	152.43	134	80	250.05	147.17	73	75
FLUCLOXACILLIN	STENO INTEST COLON	8	0.017	468.79	175.8	51	66	347.82	145.89	42	76
DORNASE ALFA	HEMOPTYSIS	64	0.336	190.37	145.75	301	83	184.47	143.63	108	77
HETASTARCH	HEMOTHORAX	21	0.083	253.02	156.63	170	75	227.72	143.29	83	78
LOTRISONE	ATROPHY SKIN	26	0.109	237.95	155.58	197	76	218.63	141.33	87	79
METIPRANOLOL	UVEITIS	25	0.105	238.78	152.82	193	78	217.63	141.1	90	80
SOMATREM	NEOPL CNS	52	0.273	190.7	139.36	299	88	183.25	138.43	109	81
LOTRISONE	DERM FUNG	15	0.053	285.44	152.24	138	82	244.74	134.52	77	82
FLUNISOLIDE	NASAL SEPTUM DIS	75	0.431	174.04	136.91	344	90	169.3	134.11	116	83
BSS PLUS	IRITIS	13	0.042	308.04	142.17	115	86	258.89	133.77	69	84
CHOLINE	KERATITIS	116	0.703	164.99	135.12	372	91	161.99	133.4	125	85
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	437	94	147.98	132.5	145	86
PRILOCAINE	HYPALGESIA	21	0.09	234.17	144.96	201	85	210.88	130.56	94	87
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	510	95	136.92	130.13	161	88
CREON	STENO INTEST COLON	6	0.01	576.05	192.02	37	57	383.06	128.18	32	89
CHYMOTRYPSIN	CORNEAL OPACITY	11	0.035	311.16	141.44	112	87	253.13	127.25	71	90
BENDECTIN	ECTROMELIA	111	0.721	154.02	126.27	415	100	150.77	125.08	140	91
CHOLINE	IRITIS	22	0.101	217.17	128.33	227	96	196.13	123.95	101	92
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	552	104	131.31	123.6	171	93
CARBACHOL	CORNEAL OPACITY	22	0.104	211.66	125.07	239	101	192.87	122.65	103	94
PROTIRELIN	URIN URGENCY	9	0.025	357.13	158.72	80	74	274.12	121.15	63	95
SELENIUM SULFIDE	HAIR DISCOLOR	301	2.236	134.64	119.43	526	107	133.78	119.65	167	96
GONADOTROPIN, CHORIONIC	OVAR DIS	86	0.567	151.69	119.94	423	106	148.02	119.13	144	97
OPHTHAINE HCL	CORNEAL LESION	21	0.101	207.17	128.25	249	97	187.48	118.28	106	98
PANALBA K.M.	ANOMALY TOOTH	8	0.022	369.52	138.57	76	89	273.5	117.57	64	99
POLIOVIRUS VACCINE, LIVE, ORAL	SIDS	146	1.038	140.6	118.45	484	108	138.9	117.14	159	100

Appendix G

Model C-G Compared with other Methods

Table G.1a: First 100 of Drug and adverse event pairs common to the top 1000 pairs selected by RR, GPS and C-G based on the point estimates of RR, λ (EBGM) and ϕ respectively, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	791.17	615.57	3	3	95.5	74.3	29	35
TROLAMINE	OTITIS EXT	14	0.007	1895.68	947.84	4	4	210.56	119.84	38	62	21.32	12.13	642	1004
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1206.22	1073.85	1	1	349.02	311.32	2	2
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	1194.52	1065.95	2	2	358.53	320.87	1	1
HYDROCOR'NE-NEOM-POLY B	PAIN EAR	37	0.041	895.62	629.35	13	6	370.77	264.93	12	14	53.2	37.52	94	131
BSS	OPHTHALMITIS	28	0.031	889.14	571.59	14	10	310.66	237.21	17	18	41.27	28.45	164	224
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	481.99	379.26	8	9	95.73	75.47	28	33
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	502.1	401.82	7	8	108.42	87.88	21	26
PANCRELIPASE	STENO INTEST COLON	34	0.047	727.21	491.94	18	13	322.71	226.92	15	19	48.51	34.27	117	156
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	644.88	582.98	4	4	323.51	293.66	3	3
CHLOROXINE	HAIR DISCOLOR	26	0.037	707.23	462.42	21	16	272.13	181.59	23	32	37.77	25.45	203	277
BSS PLUS	OPHTHALMITIS	32	0.047	685.52	449.87	24	20	303.77	211.3	20	22	45.6	31.44	129	183
COLFOSCERIL PALMITATE	HEM LUNG	41	0.06	681.25	481.86	25	15	345.34	251.08	14	17	57.63	42.28	84	106
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	598.72	532.64	5	5	267.66	238.76	4	4
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	524.7	445.7	6	6	170.1	144.86	7	9
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	397.15	305.72	10	10	80.63	62.31	44	50
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	349.13	258.63	13	15	63.54	47.75	66	81
METIPRANOLOL	IRITIS	22	0.035	623.42	368.39	34	22	233.41	150.03	32	42	32.09	20.9	284	401
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	371.54	284.7	11	11	77	58.9	47	57
SODIUM HYALURONATE	IRITIS	21	0.038	551.61	341.47	39	23	216.18	137.36	36	49	30.56	19.68	312	441
SODIUM HYALURONATE	OPHTHALMITIS	20	0.038	529.86	317.91	43	24	206.43	129.74	42	52	28.84	18.12	345	517
LOTRISONE	SKIN STRIAE	17	0.032	526.76	278.87	44	32	185.63	111.66	52	69	25.05	15.4	474	692
EMBOLEX	VASOSPASM	19	0.038	505.85	292.86	45	29	196.33	121.75	46	59	27.63	17.47	375	543
MIVACURIUM CHLORIDE	CHOLINESTERASE DEC	16	0.032	502.05	282.4	46	31	175.3	103.72	55	77	23.76	14.59	529	753
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	452.71	407.21	9	7	256.19	230.43	5	5
DOXACURIUM CHLORIDE	PARALYSIS FLACCID	16	0.033	484.79	272.69	49	33	173.11	102.42	56	79	23.48	14.13	539	792
APROTININ BOVINE	THROM CORONARY	13	0.029	452.15	208.68	57	50	146.91	81.71	74	110	19.19	10.76	785	1243
PANCRELIPASE	STENO INTEST	16	0.038	423.38	238.15	61	44	164.43	97.32	61	86	23.26	13.97	547	805
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	276.51	208.57	21	24	66.71	50.1	61	75
MYSTECLIN F	ANOMALY TOOTH	28	0.07	401.12	257.86	66	35	217.27	147.29	35	44	38.89	27.02	185	252
PHENIRAMINE	MYDRIASIS	26	0.067	387.95	253.66	70	36	206.08	137.48	43	48	36.26	24.73	218	295
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	272.47	207.8	22	25	70.17	53.31	54	71
BENZOCAINE	METHEMOGLOBIN	29	0.077	376.93	246.96	74	38	213.19	145.49	37	45	40.05	27.79	174	239
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	307.86	258.55	19	16	128.91	109.11	15	17

Table G.1b: First 100 of Drug and adverse event pairs common to the top 1000 pairs selected by RR, GPS and C-G based on the point estimates of RR, λ (EBGM) and ϕ respectively, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
DOXYLAMINE	HANGOVER	44	0.125	351.27	247.48	88	37	238.87	175.58	30	33	56.39	41.71	86	108
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	320.84	283.33	16	12	184.3	163.13	6	6
CARBOPROST	HEM POSTPARTUM	14	0.041	339.82	169.91	93	69	138.53	78.83	79	114	20.47	12.14	695	1002
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	309.61	270.47	18	13	167.41	147.07	8	8
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	233.87	173.75	31	34	59.21	44.52	80	95
SODIUM HYALURONATE	CORNEAL OPACITY	30	0.094	319.8	213.2	103	49	196.04	134.65	47	51	40.4	27.89	171	236
HEMOFIL	HIV SYND	22	0.069	318.09	187.96	106	60	171.06	109.87	57	71	30.63	20.12	309	428
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	267.85	219.9	24	20	105.76	87.46	22	27
BUDESONIDE	NASAL SEPTUM DIS	28	0.091	309.31	198.84	113	53	186.89	126.65	51	55	37.9	25.96	199	268
BSS PLUS	IRITIS	13	0.042	308.04	142.17	115	86	127.12	70.71	97	134	19	10.8	798	1234
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	226.43	170.41	33	35	62.6	47.67	72	82
OXYTOCIN	UTER ATONY	32	0.105	306.03	200.83	117	52	195.37	135.86	48	50	42.48	30.06	150	199
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	245.5	194.18	28	28	82.7	65.74	40	46
CHYMOTRYPSIN	KERATITIS	22	0.072	304.72	180.06	119	63	167.06	107.38	59	74	30.6	19.77	311	435
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	249.61	201.14	27	26	92.35	75.08	32	34
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	254.6	208.79	26	23	103.26	85.04	24	29
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	241.31	191.75	29	29	84.46	67.31	37	42
DELADUMONE OB	HIRSUTISM	21	0.071	295.96	183.21	129	61	160.9	102.29	64	80	29	18.51	341	493
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	259.43	218.26	25	21	120.37	102.14	18	18
BSS PLUS	UVEITIS	17	0.059	287.91	152.43	134	80	143.04	86.09	76	104	24.14	14.97	507	730
LOTRISONE	DERM FUNG	15	0.053	285.44	152.24	138	82	133.33	77.51	84	118	21.42	12.45	638	963
ISOSORBIDE MONONITRATE	HEADACHE VASC	40	0.141	283.71	198.59	140	54	199.82	144.69	45	46	50.34	36.71	110	138
HEPARIN SODIUM IN DEXTROSE	ANTICOAG DEC	29	0.105	275.05	180.21	147	62	175.93	120.01	54	61	38.32	26.63	192	257
BSS	CORNEAL OPACITY	21	0.077	271.35	167.98	150	71	153.23	97.36	68	85	28.88	17.99	344	523
BETHANECHOL CHLORIDE	CHOLINERG SYND	12	0.045	269.03	134.52	152	92	114.36	61.91	114	169	17.51	9.67	913	1489
BUSULFAN	VENOOCCLUS LIVER SYN	21	0.081	260.1	161.01	160	73	149.52	95.01	72	89	28.76	18.13	349	516
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	209.95	164.94	39	36	75.42	59.2	48	56
HETASTARCH	HEMOTHORAX	21	0.083	253.02	156.63	170	75	147.12	93.45	73	92	28.5	18.18	357	513
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	202.19	155.66	44	38	67.7	52.33	57	73
MAXITROL	HEALING ABNORM	14	0.056	248.73	124.36	176	102	120.18	68.42	103	146	19.92	11.53	731	1104
HYALURONIDASE	STRABISMUS	32	0.13	246.93	162.05	180	72	169.25	117.74	58	64	40.97	29.06	167	212
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	190.14	144.6	50	47	62.3	47.23	74	85
METIPRANOLOL	UVEITIS	25	0.105	238.78	152.82	193	78	152.09	100.54	70	82	33.32	22.25	264	361

Table G.1c: First 100 of Drug and adverse event pairs common to the top 1000 pairs selected by RR, GPS and C-G based on the point estimates of RR, λ (EBGM) and ϕ respectively, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
LOTRISONE	ATROPHY SKIN	26	0.109	237.95	155.58	197	76	153.92	102.66	67	78	34.22	23.04	249	339
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	222.58	194.84	34	27	139.78	122.77	12	12
PRILOCAINE	HYPALGESIA	21	0.09	234.17	144.96	201	85	140.45	89.24	78	98	28.67	18.5	352	495
FLOXURIDINE	ENTERITIS	13	0.057	229	105.69	207	125	110.94	61.7	119	171	18.5	10.57	837	1271
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	207.82	182.29	40	31	135.44	118.88	13	14
CHOLINE	IRITIS	22	0.101	217.17	128.33	227	96	136.47	87.64	81	101	29.28	18.92	337	470
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	207.07	187.02	41	30	156.35	141.35	9	10
CARBACHOL	CORNEAL OPACITY	22	0.104	211.66	125.07	239	101	134.24	86.29	83	103	29.15	18.78	339	481
DEMECLOCYCLINE	ANOMALY TOOTH	17	0.08	211.34	111.89	240	118	120.88	72.81	101	128	23.34	14.07	543	795
OPHTHAINE HCL	CORNEAL LESION	21	0.101	207.17	128.25	249	97	130.12	82.65	90	109	28.08	17.68	369	538
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	192.06	164.53	49	37	114.11	97.77	19	20
POTASSIUM IODIDE	SALIV GLAND ENLARGE	12	0.059	204.26	102.13	259	129	100.46	54.44	143	220	17.02	9.37	958	1575
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	184.95	155.6	53	39	101.96	85.48	25	28
MOMETASONE FUROATE	LEUKODERMA	18	0.093	194.46	108.03	283	123	117.78	72.03	106	130	24.32	14.9	502	732
SOMATREM	NEOPL CNS	52	0.273	190.7	139.36	299	88	156.48	118.05	66	63	56.25	42.52	88	105
DORNASE ALFA	HEMOPTYSIS	64	0.336	190.37	145.75	301	83	161.72	125.57	62	57	64.67	50.08	63	76
DINOPROSTONE	UTER ATONY	16	0.086	186.95	105.16	311	126	109.56	64.81	123	160	22.04	13.14	606	881
PAMIDRONATE DISODIUM	HYPOPHOSPHATEM	14	0.076	184.06	92.03	320	153	102.41	58.28	139	195	19.51	11.22	760	1160
ISOVUE-M	MENINGISM	15	0.082	182.77	97.47	321	137	105.16	61.13	131	175	20.73	12.44	678	964
BSS PLUS	CORNEAL LESION	27	0.149	181.26	114.13	326	113	129.19	86.86	93	102	33.87	22.64	255	353
INDIUM, IN-111	MENINGITIS	17	0.095	179.59	95.08	333	144	109.59	66	122	154	22.93	14	559	802
CHOLINE	UVEITIS	21	0.119	176.9	109.51	337	121	117.33	74.57	107	125	27.34	17.53	381	541
FLUNISOLIDE	NASAL SEPTUM DIS	75	0.431	174.04	136.91	344	90	152.87	121.09	69	60	69.25	55.41	55	64
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	165.73	148.27	60	43	126.83	113.36	16	16
PANALBA K.M.	TOOTH DIS	33	0.194	170.07	113.38	359	115	129.89	90.87	92	97	38.89	27.87	184	237
ABCIXIMAB	HEM RETROPERIT	14	0.083	168.75	84.38	361	168	97.38	55.43	150	215	19.34	11.45	773	1119
TICE BCG	PYURIA	17	0.101	168.22	89.06	365	159	105.18	63.34	130	165	22.78	13.98	564	804
NAPHCON-A	MYDRIASIS	18	0.109	165.34	91.86	370	154	106.25	64.95	128	159	23.81	14.78	523	738
ETODOLAC	BILIRUBINURIA	23	0.139	165.02	100.45	371	132	115.17	74.79	111	124	29.32	19.49	334	451
CHOLINE	KERATITIS	116	0.703	164.99	135.12	372	91	152.08	126.25	71	56	85.71	71.27	36	39
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	161.58	152.26	63	40	148.51	139.79	11	11
IFOSFAMIDE	FANCONI SYND	13	0.081	161.42	74.5	385	201	91.86	51.11	173	239	17.86	9.93	882	1425
UROFOLLITROPIN	OVAR DIS	46	0.289	159.3	114.28	394	112	131.89	97.69	86	84	48.9	35.97	116	144

Table G.2a: First 100 of drug and adverse event pairs common to the top 1000 pairs selected by RR, GPS and C-G based on the lower bounds RR_{025}, λ_{025} and ϕ_{025} of the 95% confidence/credible interval estimates of RR, λ and ϕ respectively, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	791.17	615.57	3	3	95.5	74.3	29	35
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1206.22	1073.85	1	1	349.02	311.32	2	2
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	1194.52	1065.95	2	2	358.53	320.87	1	1
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	644.88	582.98	4	4	323.51	293.66	3	3
HYDROCOR'NE-NEOM-POLY B	PAIN EAR	37	0.041	895.62	629.35	13	6	370.77	264.93	12	14	53.2	37.52	94	131
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	502.1	401.82	7	8	108.42	87.88	21	26
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	481.99	379.26	8	9	95.73	75.47	28	33
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	598.72	532.64	5	5	267.66	238.76	4	4
BSS	OPHTHALMITIS	28	0.031	889.14	571.59	14	10	310.66	237.21	17	18	41.27	28.45	164	224
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	524.7	445.7	6	6	170.1	144.86	7	9
PANCRELIPASE	STENO INTEST COLON	34	0.047	727.21	491.94	18	13	322.71	226.92	15	19	48.51	34.27	117	156
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	397.15	305.72	10	10	80.63	62.31	44	50
COLFOSCERIL PALMITATE	HEM LUNG	41	0.06	681.25	481.86	25	15	345.34	251.08	14	17	57.63	42.28	84	106
CHLOROXINE	HAIR DISCOLOR	26	0.037	707.23	462.42	21	16	272.13	181.59	23	32	37.77	25.45	203	277
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	371.54	284.7	11	11	77	58.9	47	57
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	349.13	258.63	13	15	63.54	47.75	66	81
BSS PLUS	OPHTHALMITIS	32	0.047	685.52	449.87	24	20	303.77	211.3	20	22	45.6	31.44	129	183
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	452.71	407.21	9	7	256.19	230.43	5	5
METIPRANOLOL	IRITIS	22	0.035	623.42	368.39	34	22	233.41	150.03	32	42	32.09	20.9	284	401
SODIUM HYALURONATE	IRITIS	21	0.038	551.61	341.47	39	23	216.18	137.36	36	49	30.56	19.68	312	441
SODIUM HYALURONATE	OPHTHALMITIS	20	0.038	529.86	317.91	43	24	206.43	129.74	42	52	28.84	18.12	345	517
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	320.84	283.33	16	12	184.3	163.13	6	6
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	307.86	258.55	19	16	128.91	109.11	15	17
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	309.61	270.47	18	13	167.41	147.07	8	8
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	276.51	208.57	21	24	66.71	50.1	61	75
EMBOLEX	VASOSPASM	19	0.038	505.85	292.86	45	29	196.33	121.75	46	59	27.63	17.47	375	543
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	272.47	207.8	22	25	70.17	53.31	54	71
MIVACURIUM CHLORIDE	CHOLINESTERASE DEC	16	0.032	502.05	282.4	46	31	175.3	103.72	55	77	23.76	14.59	529	753
LOTRISONE	SKIN STRIAE	17	0.032	526.76	278.87	44	32	185.63	111.66	52	69	25.05	15.4	474	692
DOXACURIUM CHLORIDE	PARALYSIS FLACCID	16	0.033	484.79	272.69	49	33	173.11	102.42	56	79	23.48	14.13	539	792
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	267.85	219.9	24	20	105.76	87.46	22	27
MYSTECLIN F	ANOMALY TOOTH	28	0.07	401.12	257.86	66	35	217.27	147.29	35	44	38.89	27.02	185	252
PHENIRAMINE	MYDRIASIS	26	0.067	387.95	253.66	70	36	206.08	137.48	43	48	36.26	24.73	218	295
DOXYLAMINE	HANGOVER	44	0.125	351.27	247.48	88	37	238.87	175.58	30	33	56.39	41.71	86	108

Table G.2b: First 100 of drug and adverse event pairs common to the top 1000 pairs selected by RR, GPS and C-G based on the lower bounds RR_{025}, λ_{025} and ϕ_{025} of the 95% confidence/credible interval estimates of RR, λ and ϕ respectively, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
BENZOCAINE	METHEMOGLOBIN	29	0.077	376.93	246.96	74	38	213.19	145.49	37	45	40.05	27.79	174	239
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	259.43	218.26	25	21	120.37	102.14	18	18
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	254.6	208.79	26	23	103.26	85.04	24	29
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	233.87	173.75	31	34	59.21	44.52	80	95
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	245.5	194.18	28	28	82.7	65.74	40	46
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	249.61	201.14	27	26	92.35	75.08	32	34
PANCRELIPASE	STENO INTEST	16	0.038	423.38	238.15	61	44	164.43	97.32	61	86	23.26	13.97	547	805
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	241.31	191.75	29	29	84.46	67.31	37	42
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	226.43	170.41	33	35	62.6	47.67	72	82
SODIUM HYALURONATE	CORNEAL OPACITY	30	0.094	319.8	213.2	103	49	196.04	134.65	47	51	40.4	27.89	171	236
GEMCITABINE HYDROCHLO'DE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	222.58	194.84	34	27	139.78	122.77	12	12
OXYTOCIN	UTER ATONY	32	0.105	306.03	200.83	117	52	195.37	135.86	48	50	42.48	30.06	150	199
BUDESONIDE	NASAL SEPTUM DIS	28	0.091	309.31	198.84	113	53	186.89	126.65	51	55	37.9	25.96	199	268
ISOSORBIDE MONONITRATE	HEADACHE VASC	40	0.141	283.71	198.59	140	54	199.82	144.69	45	46	50.34	36.71	110	138
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	209.95	164.94	39	36	75.42	59.2	48	56
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	207.07	187.02	41	30	156.35	141.35	9	10
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	207.82	182.29	40	31	135.44	118.88	13	14
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	202.19	155.66	44	38	67.7	52.33	57	73
HEMOFIL	HIV SYND	22	0.069	318.09	187.96	106	60	171.06	109.87	57	71	30.63	20.12	309	428
DELADUMONE OB	HIRSUTISM	21	0.071	295.96	183.21	129	61	160.9	102.29	64	80	29	18.51	341	493
HEPARIN SODIUM IN DEXTROS	ANTICOAG DEC	29	0.105	275.05	180.21	147	62	175.93	120.01	54	61	38.32	26.63	192	257
CHYMOTRYPSIN	KERATITIS	22	0.072	304.72	180.06	119	63	167.06	107.38	59	74	30.6	19.77	311	435
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	190.14	144.6	50	47	62.3	47.23	74	85
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	192.06	164.53	49	37	114.11	97.77	19	20
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	184.95	155.6	53	39	101.96	85.48	25	28
BSS	CORNEAL OPACITY	21	0.077	271.35	167.98	150	71	153.23	97.36	68	85	28.88	17.99	344	523
HYALURONIDASE	STRABISMUS	32	0.13	246.93	162.05	180	72	169.25	117.74	58	64	40.97	29.06	167	212
BUSULFAN	VENOOCCLUS LIVER SYN	21	0.081	260.1	161.01	160	73	149.52	95.01	72	89	28.76	18.13	349	516
HETASTARCH	HEMOTHORAX	21	0.083	253.02	156.63	170	75	147.12	93.45	73	92	28.5	18.18	357	513
LOTRISONE	ATROPHY SKIN	26	0.109	237.95	155.58	197	76	153.92	102.66	67	78	34.22	23.04	249	339
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	161.58	152.26	63	40	148.51	139.79	11	11
METIPRANOLOL	UVEITIS	25	0.105	238.78	152.82	193	78	152.09	100.54	70	82	33.32	22.25	264	361
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	157.84	152.09	65	41	152.79	147.32	10	7

Table G.2c: First 100 of drug and adverse event pairs common to the top 1000 pairs selected by RR, GPS and C-G based on the lower bounds RR_{025}, λ_{025} and ϕ_{025} of the 95% confidence/credible interval estimates of RR, λ and ϕ respectively, using Data 1. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
BSS PLUS	UVEITIS	17	0.059	287.91	152.43	134	80	143.04	86.09	76	104	24.14	14.97	507	730
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	165.73	148.27	60	43	126.83	113.36	16	16
LOTRISONE	DERM FUNG	15	0.053	285.44	152.24	138	82	133.33	77.51	84	118	21.42	12.45	638	963
DORNASE ALFA	HEMOPTYSIS	64	0.336	190.37	145.75	301	83	161.72	125.57	62	57	64.67	50.08	63	76
PRILOCAINE	HYPALGESIA	21	0.09	234.17	144.96	201	85	140.45	89.24	78	98	28.67	18.5	352	495
SOMATREM	NEOPL CNS	52	0.273	190.7	139.36	299	88	156.48	118.05	66	63	56.25	42.52	88	105
FLUNISOLIDE	NASAL SEPTUM DIS	75	0.431	174.04	136.91	344	90	152.87	121.09	69	60	69.25	55.41	55	64
CHOLINE	KERATITIS	116	0.703	164.99	135.12	372	91	152.08	126.25	71	56	85.71	71.27	36	39
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	437	94	144.84	129.37	75	53	113.33	101.49	20	19
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	510	95	136.28	129.37	82	54	129.01	122.5	14	13
CHOLINE	IRITIS	22	0.101	217.17	128.33	227	96	136.47	87.64	81	101	29.28	18.92	337	470
OPHTHAINE HCL	CORNEAL LESION	21	0.101	207.17	128.25	249	97	130.12	82.65	90	109	28.08	17.68	369	538
BENDECTIN	ECTROMELIA	111	0.721	154.02	126.27	415	100	142.21	117.53	77	65	80.76	67.11	43	43
CARBACHOL	CORNEAL OPACITY	22	0.104	211.66	125.07	239	101	134.24	86.29	83	103	29.15	18.78	339	481
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	552	104	130.65	123	89	58	121.83	114.78	17	15
GONADOTROPIN, CHORIONIC	OVAR DIS	86	0.567	151.69	119.94	423	106	137.2	110.41	80	70	70.3	56.8	52	62
SELENIUM SULFIDE	HAIR DISCOLOR	301	2.236	134.64	119.43	526	107	131.12	116.94	87	66	104.37	93.05	23	21
POLIOVIRUS VACCINE, LIVE	SIDS	146	1.038	140.6	118.45	484	108	132.92	112.61	85	68	86.33	73.53	35	36
OPCON A	PAIN EYE	267	2.015	132.49	116.61	547	109	128.66	113.91	95	67	100.08	88.99	26	24
FLOSEQUINAN	SUDDEN DEATH	49	0.315	155.79	114.46	404	111	130.83	97.82	88	83	50.9	38.54	101	123
UROFOLLITROPIN	OVAR DIS	46	0.289	159.3	114.28	394	112	131.89	97.69	86	84	48.9	35.97	116	144
BSS PLUS	CORNEAL LESION	27	0.149	181.26	114.13	326	113	129.19	86.86	93	102	33.87	22.64	255	353
PANALBA K.M.	TOOTH DIS	33	0.194	170.07	113.38	359	115	129.89	90.87	92	97	38.89	27.87	184	237
CEFOXITIN	ENTEROCOL PSEUDOMEM	138	1.016	135.79	113.16	517	116	128.21	108.11	96	73	82.86	69.82	39	41
TETRAHYDROZOLINE	PAIN EYE	187	1.43	130.77	112.59	557	117	125.49	108.46	98	72	89.85	77.83	33	31
DEMECLOCYCLINE	ANOMALY TOOTH	17	0.08	211.34	111.89	240	118	120.88	72.81	101	128	23.34	14.07	543	795
TICE BCG	CYSTITIS	69	0.475	145.35	111.64	457	119	129.03	101.11	94	81	61.19	48.18	77	78
CHOLINE	UVEITIS	21	0.119	176.9	109.51	337	121	117.33	74.57	107	125	27.34	17.53	381	541
MENOTROPINS	OVAR DIS	151	1.175	128.48	108.06	583	122	122.22	103.87	99	76	82.37	70	41	40
MOMETASONE FUROATE	LEUKODERMA	18	0.093	194.46	108.03	283	123	117.78	72.03	106	130	24.32	14.9	502	732
PHENFORMIN	ACIDOSIS LACTIC	390	3.302	118.11	106.61	681	124	115.99	104.91	109	75	98.53	89.48	27	23
DINOPROSTONE	UTER ATONY	16	0.086	186.95	105.16	311	126	109.56	64.81	123	160	22.04	13.14	606	881
INFLUENZA VIRUS VACCINE	GUILLAIN BARRE SYND	62	0.451	137.47	104.21	507	127	121.28	93.84	100	91	56.07	43.85	89	97

Table G.3: Top ten drug and adverse event pairs uniquely selected by RR, GPS, C-G and LogP in their top 1000 pairs based on the point estimates of RR, λ (EBGM), ϕ and the values of LogP respectively, using Data 1.

Method	Drug	Event	Rank	Count	Expected	RR	RR_{025}	EBGM	λ_{025}	ϕ	ϕ_{025}	LogP
RR	NISOLDIPINE	HEPATITIS NONSPECIFI	1	1	0.0003	3561.94	0.00	2.4	0.95	1.83	0.25	3.55
	ACARBOSE	HEPATITIS NONSPECIFI	2	1	0.0004	2816.96	0.00	2.4	0.95	1.76	0.2	3.45
	ORTHO-NOVUM 1/80	CARCINOMA LIVER	5	1	0.0005	1840	0.00	2.39	0.95	1.84	0.27	3.26
	PLATELET CONCENTRATE, HUMAN	LIVER DAMAGE AGGRAV	10	1	0.0008	1320.94	0.00	2.38	0.94	1.8	0.22	3.12
	URSODIOL	LIVER DAMAGE AGGRAV	11	1	0.0008	1187.47	0.00	2.38	0.94	1.81	0.23	3.07
	EMLA	HYPALGESIA	12	1	0.0009	1114.14	0.00	2.37	0.94	1.8	0.23	3.05
	ORTHO-NOVUM SQ	CARCINOMA LARYNX	19	1	0.0014	720.74	0.00	2.35	0.94	1.78	0.2	2.86
	COLFOSCERIL PALMITATE	INTEST SMALL PER	22	2	0.0029	699.11	0.00	19.81	4.63	3.28	0.75	5.39
	AVC	BALANITIS	23	1	0.0014	694.57	0.00	2.35	0.93	1.8	0.22	2.84
	HYDROCORTIS'NE-NEOMYCIN-POLYMYXIN B	OTITIS EXT	29	2	0.0031	648.98	0	19.7	4.62	3.21	0.75	5.32
GPS (EBGM)	DACTINOMYCIN	SARCOMA	355	12	0.1268	94.63	47.31	63.39	34.35	15.55	8.68	19.49
	SODIUM HYALURONATE	CORNEAL LESION	378	12	0.1322	90.79	45.4	61.6	33.38	15.52	8.73	19.28
	OXYMETAZOLINE	NASAL SEPTUM DIS	390	11	0.1201	91.58	41.63	60.16	31.6	14.38	7.93	17.77
	ARA-C	ATAXIA CEREBELL	409	11	0.1258	87.42	39.74	58.28	30.64	14.39	7.66	17.55
	METHOXYFLURANE	POLYURIA	418	10	0.1115	89.65	35.86	57.3	29.14	13.2	7.07	16.13
	BENDECTIN	VENT SEPT DEF	421	10	0.1118	89.47	35.79	57.22	29.07	13.28	6.82	16.12
	RIFABUTIN	OPHTHALMITIS	429	11	0.1308	84.07	38.22	56.73	29.85	14.25	7.59	17.37
	PHENYLEPHRINE	IRITIS	437	10	0.1147	87.21	34.88	56.26	28.59	13.19	7.04	16.01
	BETAMETHASONE	SPERM ARREST	447	9	0.0982	91.66	40.74	55.79	27.21	12.18	6.33	14.67
	SULFISOXAZOLE	SIDS	450	11	0.1344	81.86	37.21	55.69	29.27	14.22	7.63	17.24
C-G (Phi)	NITRODUR II	APPLICAT SITE REACT	397	82	2.3929	34.27	27.16	33.34	26.68	26.96	21.58	92.63
	PILOCARPINE	GLAUCOMA	419	79	2.3374	33.8	26.52	32.85	26.18	26.56	21.26	88.82
	PROCHLORPERAZINE	TORTICOLLIS	443	66	1.8801	35.1	27.13	33.9	26.43	26.08	20.17	75.44
	MINOCYCLINE	INTRACRAN HYPERTENS	451	88	2.7479	32.02	25.47	31.26	25.22	25.99	21.09	96.82
	EPOGEN	THROM	454	79	2.4007	32.91	25.83	32.01	25.51	25.96	20.79	87.93
	METHYLPREDNISOLONE	NECRO BONE	462	91	2.8903	31.48	25.26	30.76	24.91	25.75	21.07	99.43
	LEUCOVORIN	MUCOUS MEM DIS	466	59	1.6682	35.37	26.38	34.01	26.11	25.46	19.78	67.74
	MASOPROCOL	EDEMA	471	61	1.7668	34.53	26.04	33.27	25.67	25.25	19.51	69.38
	ISOFLURANE	INCREASED EFFECT	501	87	2.9108	29.89	23.71	29.2	23.53	24.39	19.68	93.2
	EPINEPHRINE	KERATITIS	510	57	1.7256	33.03	24.92	31.79	24.3	24.1	18.45	63.84
LogP	LEVONORGESTREL	METRORRHAGIA	3	7530	524.2800	14.36	14.04	14.36	14.04	14.35	14.05	5673.72
	NICOTINE	APPLICAT SITE REACT	8	5085	358.5990	14.18	13.79	14.18	13.79	14.15	13.76	3805.87
	DIATRIZOIC ACID	URTICARIA	9	5404	442.6480	12.21	11.88	12.21	11.88	12.19	11.88	3719.83
	PERMETHRIN	NO DRUG EFFECT	11	4029	249.0420	16.18	15.68	16.17	15.68	16.12	15.64	3231.32
	INSULIN HUMAN	NO DRUG EFFECT	14	6194	1003.6500	6.17	6.02	6.17	6.02	6.17	6.03	2643.75
	ESTRADIOL	APPLICAT SITE REACT	16	3664	325.3060	11.26	10.9	11.26	10.9	11.24	10.89	2405.46
	IOTHALAMIC ACID	URTICARIA	20	2924	248.9040	11.75	11.32	11.74	11.32	11.72	11.3	1968.83
	INSULIN	HYPERGLYCEM	21	2429	148.4330	16.36	15.72	16.36	15.72	16.28	15.69	1960.18
	DTP VACCINE	FEVER	23	2494	192.1860	12.98	12.47	12.97	12.47	12.93	12.42	1778.66
	ESTRADIOL	RASH	24	4852	952.2200	5.1	4.95	5.09	4.97	5.09	4.95	1739.75

Table G.4: Top ten drug and adverse event pairs uniquely selected by RR, GPS and C-G in their top 1000 pairs based on the lower bounds RR_{025}, λ_{025} and ϕ_{025} of the 95% confidence/credible interval estimates of RR, λ and ϕ respectively, using Data 1.

Method	Drug	Event	Rank	Count	Expected	RR	RR_{025}	EBGM	λ_{025}	ϕ	ϕ_{025}
RR	CREON	STENO INTEST	41	4	0.0074	538.02	134.5	56.79	18.23	6.40	2.26
	SALICYLIC ACID	SALICYLISM	85	4	0.0113	355.05	88.76	53.61	17.19	6.27	2.22
	PANCREATIN	STENO INTEST COLON	104	4	0.0125	318.95	79.74	52.63	16.92	6.21	2.09
	ENOXACIN	TENDON RUPT	131	4	0.0137	292.03	73.01	51.77	16.61	6.28	2.29
	IOCARMATE MEGLUMINE	PARAPLEGIA	155	4	0.0151	264.64	66.16	50.76	16.31	6.20	2.21
	PHENACETIN	NECRO KIDNEY PAPILL	172	4	0.0158	252.83	63.21	50.27	16.15	6.22	2.12
	CEPHAPIRIN	RESPIRAT DISTRES SYN	186	4	0.0165	242.36	60.59	49.80	15.99	6.28	2.23
	HYALURONIDASE	NEURITIS RETROBULBAR	244	4	0.0191	209.48	52.37	48.10	15.45	6.28	2.20
	ECHOTHIOPHATE IODIDE	CHOLINERG SYND	258	4	0.0195	204.64	51.16	47.82	15.35	6.28	2.24
	NORMOSOL-R	ALKALOSIS	272	4	0.0201	198.93	49.73	47.47	15.27	6.25	2.31
GPS (EBGM)	CLOBETASOL PROPIONATE	LEUKODERMA	703	7	0.1002	69.84	19.95	42.41	27.45	9.55	4.54
	PENICILLAMINE	GLOMERULITIS	1022	22	0.5865	37.51	22.17	33.67	21.65	17.84	11.70
	FLUOROMETHOLONE	KERATITIS	923	17	0.4103	41.44	21.94	35.67	21.47	16.04	10.03
	MOTIVAL	DRUG DEPEND	820	13	0.2746	47.34	21.85	38.20	21.26	14.29	8.09
	METHOXYFLURANE	KIDNEY FAIL ACUTE	974	18	0.4540	39.65	22.03	34.58	21.12	16.47	10.14
	HAEMOPHILUS B POLYSACCHARIDE VACCINE	LARYNGITIS	890	14	0.3171	44.15	22.08	36.54	20.80	14.61	8.53
	NORINYL	THROM CEREBR	863	13	0.2848	45.65	21.07	37.06	20.61	14.08	8.19
	MEGESTROL	ADREN INSUFFIC	906	14	0.3215	43.54	21.77	36.11	20.55	14.59	8.63
	NORFLOXACIN	ENTEROCOL PSEUDOMEM	1122	22	0.6231	35.31	20.86	31.87	20.48	17.40	11.33
	CYCLOPENTOLATE	HEM EYE	760	10	0.1830	54.64	21.85	40.28	20.48	12.12	6.29
C-G (Phi)	INSULIN HUMAN	HYPOGLYCEM	682	1810	87.0265	20.8	19.84	20.78	19.84	20.65	19.7
	MINOXIDIL	SKIN DRY	714	1038	50.7291	20.46	19.22	20.43	19.22	20.20	19.00
	PILOCARPINE	CONJUNCTIVITIS	635	244	10.7042	22.79	19.99	22.64	19.94	21.49	18.96
	SELENIUM SULFIDE	NO DRUG EFFECT	697	722	34.7374	20.78	19.29	20.74	19.27	20.40	18.92
	PROTAMINE	HEM	659	355	16.2651	21.83	19.61	21.73	19.56	21.01	18.88
	LEVONORGESTREL	REACT UNEVAL	770	5043	259.501	19.43	18.9	19.43	18.90	19.38	18.84
	FINASTERIDE	DEATH	711	544	26.2034	20.76	19.04	20.70	19.02	20.23	18.63
	FLUOROURACIL	MUCOUS MEM DIS	622	157	6.5840	23.85	20.20	23.59	20.11	21.71	18.63
	PIROXICAM	ULCER PEPTIC	621	134	5.5321	24.22	20.25	23.92	20.12	21.72	18.41
	FINASTERIDE	GYNECOMASTIA	735	511	25.0667	20.39	18.63	20.33	18.62	19.86	18.27

G.1 Formula for Computing LogP

$$\begin{aligned}
\text{LogP}_{ij} &= -\log_{10}\left[Pr(X \geq N_{ij})\right] \quad \text{where} \quad X \sim \text{Poisson}(E_{ij}) \\
&= -\log_{10}\left[\sum_{n \geq N_{ij}} \frac{e^{-E_{ij}}(E_{ij})^n}{n!}\right] \\
&= -\log_{10}\left[1 - \sum_{n=0}^{(N_{ij}-1)} \frac{e^{-E_{ij}}(E_{ij})^n}{n!}\right]
\end{aligned}$$

which is approximated by (G.1) to avoid computational problems when $N_{ij} >> E_{ij}$ and LogP_{ij} is large ($\text{LogP}_{ij} > 12$) [19].

$$\begin{aligned}
&-\log_{10}\left[\sum_{n=N_{ij}}^{(N_{ij}+3)} \frac{e^{-E_{ij}}(E_{ij})^n}{n!}\right] \\
&= \frac{E_{ij}}{\log(10)} - N_{ij}\log_{10}(E_{ij}) + \log_{10}(N_{ij}!) \\
&-\log_{10}\left[1 + \frac{E_{ij}}{N_{ij}+1} + \frac{(E_{ij})^2}{(N_{ij}+1)(N_{ij}+2)} + \frac{(E_{ij})^3}{(N_{ij}+1)(N_{ij}+2)(N_{ij}+3)}\right]
\end{aligned} \tag{G.1}$$

Table G.5: Drug and adverse event pairs with various combinations of observed and expected counts. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11, RK2, RK22, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
EMLA	HYPALGESIA	1	0.0009	1114.14	0.00	12	154785	2.37	0.94	29826	55171	1.80	0.23	63130	186757
COPPER	UTER DIS	4457	46.3981	96.06	93.26	975	151	95.93	93.15	155	94	94.76	91.97	30	22
VINCRISTINE	PURPURA THROMBOPEN	2	11.4792	0.17	0.00	373725	375525	0.21	0.07	374503	372101	0.18	0.04	373694	360887
DILTIAZEM	RASH	1697	1374.8900	1.23	1.18	189850	33326	1.23	1.18	99124	41473	1.23	1.18	140633	34513
NICOTINE	RASH	3827	1187.1800	3.22	3.12	78434	10609	3.22	3.14	20191	12293	3.22	3.12	20403	8143
IBUPROFEN	NO DRUG EFFECT	1547	1161.5000	1.33	1.27	178655	30505	1.33	1.27	85474	37764	1.33	1.27	124225	30832

Table G.6a: Top fifty drug and adverse event pairs selected by C-G using the point estimate of ϕ. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	1194.52	1065.95	2	2	358.53	320.87	1	1
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1206.22	1073.85	1	1	349.02	311.32	2	2
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	644.88	582.98	4	4	323.51	293.66	3	3
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	598.72	532.64	5	5	267.66	238.76	4	4
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	452.71	407.21	9	7	256.19	230.43	5	5
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	320.84	283.33	16	12	184.3	163.13	6	6
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	524.7	445.7	6	6	170.1	144.86	7	9
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	309.61	270.47	18	13	167.41	147.07	8	8
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	207.07	187.02	41	30	156.35	141.35	9	10
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	157.84	152.09	65	41	152.79	147.32	10	7
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	161.58	152.26	63	40	148.51	139.79	11	11
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	222.58	194.84	34	27	139.78	122.77	12	12
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	207.82	182.29	40	31	135.44	118.88	13	14
TETRACYCLINE	DISCOLOR TOOTH	1437	10.485	137.05	130	510	95	136.28	129.37	82	54	129.01	122.5	14	13
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	307.86	258.55	19	16	128.91	109.11	15	17
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	165.73	148.27	60	43	126.83	113.36	16	16
NONOXYNOL	CERVIX DIS	1074	8.161	131.61	123.77	552	104	130.65	123	89	58	121.83	114.78	17	15
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	259.43	218.26	25	21	120.37	102.14	18	18
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	192.06	164.53	49	37	114.11	97.77	19	20
METFORMIN	ACIDOSIS LACTIC	309	2.074	149.02	132.62	437	94	144.84	129.37	75	53	113.33	101.49	20	19
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	502.1	401.82	7	8	108.42	87.88	21	26
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	267.85	219.9	24	20	105.76	87.46	22	27
SELENIUM SULFIDE	HAIR DISCOLOR	301	2.236	134.64	119.43	526	107	131.12	116.94	87	66	104.37	93.05	23	21
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	254.6	208.79	26	23	103.26	85.04	24	29
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	184.95	155.6	53	39	101.96	85.48	25	28
OPCON A	PAIN EYE	267	2.015	132.49	116.61	547	109	128.66	113.91	95	67	100.08	88.99	26	24
PHENFORMIN	ACIDOSIS LACTIC	390	3.302	118.11	106.61	681	124	115.99	104.91	109	75	98.53	89.48	27	23
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	481.99	379.26	8	9	95.73	75.47	28	33
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	791.17	615.57	3	3	95.5	74.3	29	35
COPPER	UTER DIS	4457	46.398	96.06	93.26	975	151	95.93	93.15	155	94	94.76	91.97	30	22

Table G.6b: Top fifty drug and adverse event pairs selected by C-G using the point estimate of ϕ. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
DTP VACCINE	SCREAMING SYND	966	9.643	100.18	93.95	897	147	99.55	93.41	146	93	93.9	88.19	31	25
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	249.61	201.14	27	26	92.35	75.08	32	34
TETRAHYDROZOLINE	PAIN EYE	187	1.43	130.77	112.59	557	117	125.49	108.46	98	72	89.85	77.83	33	31
POLIOVIRUS VACCINE, LIVE	SCREAMING SYND	344	3.28	104.89	93.91	831	148	102.99	92.52	138	96	87.35	78.75	34	30
POLIOVIRUS VACCINE, LIVE	SIDS	146	1.038	140.6	118.45	484	108	132.92	112.61	85	68	86.33	73.53	35	36
CHOLINE	KERATITIS	116	0.703	164.99	135.12	372	91	152.08	126.25	71	56	85.71	71.27	36	39
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	241.31	191.75	29	29	84.46	67.31	37	42
LOMEFLOXACIN HYDRO'DE	PHOTOSENSITIVITY	481	5.137	93.64	85.27	1023	166	92.54	84.54	169	107	83.08	76.13	38	32
CEFOXITIN	ENTEROCOL PSEUDOMEM	138	1.016	135.79	113.16	517	116	128.21	108.11	96	73	82.86	69.82	39	41
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	245.5	194.18	28	28	82.7	65.74	40	46
MENOTROPINS	OVAR DIS	151	1.175	128.48	108.06	583	122	122.22	103.87	99	76	82.37	70	41	40
MITOMYCIN	UREMIA	218	1.995	109.29	95.26	785	142	106.07	92.67	129	95	82.28	71.86	42	37
BENDECTIN	ECTROMELIA	111	0.721	154.02	126.27	415	100	142.21	117.53	77	65	80.76	67.11	43	43
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	397.15	305.72	10	10	80.63	62.31	44	50
DIPHTHERIA-TET TOX-PERT VA	SCREAMING SYND	147	1.22	120.47	101.62	654	130	114.79	97.31	113	87	78.45	66.58	45	45
TISSUE PLAS'GEN ACT, REC	HEM INTRACRAN	476	5.426	87.73	79.99	1150	184	86.75	79.22	191	113	78.28	71.37	46	38
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	371.54	284.7	11	11	77	58.9	47	57
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	209.95	164.94	39	36	75.42	59.2	48	56
HAEMOPHILUS B POLY'DE VAC	MENINGITIS	176	1.711	102.86	87.66	851	164	99.33	85.44	147	106	74.45	64.4	49	48
TETRAHYDROZOLINE	CONJUNCTIVITIS	230	2.454	93.73	81.91	1021	174	91.45	80.2	175	112	74.12	65.21	50	47

Table G.7a: Top fifty drug and adverse event pairs selected by GPS using the point estimate (EBGM) of λ. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1206.22	1073.85	1	1	349.02	311.32	2	2
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	1194.52	1065.95	2	2	358.53	320.87	1	1
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	791.17	615.57	3	3	95.5	74.3	29	35
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	644.88	582.98	4	4	323.51	293.66	3	3
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	598.72	532.64	5	5	267.66	238.76	4	4
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	524.7	445.7	6	6	170.1	144.86	7	9
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	502.1	401.82	7	8	108.42	87.88	21	26
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	481.99	379.26	8	9	95.73	75.47	28	33
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	452.71	407.21	9	7	256.19	230.43	5	5
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	397.15	305.72	10	10	80.63	62.31	44	50
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	371.54	284.7	11	11	77	58.9	47	57
HYDROCOR'NE-NEOM-POLY B	PAIN EAR	37	0.041	895.62	629.35	13	6	370.77	264.93	12	14	53.2	37.52	94	131
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	349.13	258.63	13	15	63.54	47.75	66	81
COLFOSCERIL PALMITATE	HEM LUNG	41	0.06	681.25	481.86	25	15	345.34	251.08	14	17	57.63	42.28	84	106
PANCRELIPASE	STENO INTEST COLON	34	0.047	727.21	491.94	18	13	322.71	226.92	15	19	48.51	34.27	117	156
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	320.84	283.33	16	12	184.3	163.13	6	6
BSS	OPHTHALMITIS	28	0.031	889.14	571.59	14	10	310.66	237.21	17	18	41.27	28.45	164	224
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	309.61	270.47	18	13	167.41	147.07	8	8
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	307.86	258.55	19	16	128.91	109.11	15	17
BSS PLUS	OPHTHALMITIS	32	0.047	685.52	449.87	24	20	303.77	211.3	20	22	45.6	31.44	129	183
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	276.51	208.57	21	24	66.71	50.1	61	75
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	272.47	207.8	22	25	70.17	53.31	54	71
CHLOROXINE	HAIR DISCOLOR	26	0.037	707.23	462.42	21	16	272.13	181.59	23	32	37.77	25.45	203	277
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	267.85	219.9	24	20	105.76	87.46	22	27
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	259.43	218.26	25	21	120.37	102.14	18	18
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	254.6	208.79	26	23	103.26	85.04	24	29
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	249.61	201.14	27	26	92.35	75.08	32	34
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	245.5	194.18	28	28	82.7	65.74	40	46
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	241.31	191.75	29	29	84.46	67.31	37	42
DOXYLAMINE	HANGOVER	44	0.125	351.27	247.48	88	37	238.87	175.58	30	33	56.39	41.71	86	108

Table G.7b: Top fifty drug and adverse event pairs selected by GPS using the point estimate (EBGM) of λ. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	233.87	173.75	31	34	59.21	44.52	80	95
METIPRANOLOL	IRITIS	22	0.035	623.42	368.39	34	22	233.41	150.03	32	42	32.09	20.9	284	401
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	226.43	170.41	33	35	62.6	47.67	72	82
GEMCITABINE HYDROCHLORIDE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	222.58	194.84	34	27	139.78	122.77	12	12
MYSTECLIN F	ANOMALY TOOTH	28	0.07	401.12	257.86	66	35	217.27	147.29	35	44	38.89	27.02	185	252
SODIUM HYALURONATE	IRITIS	21	0.038	551.61	341.47	39	23	216.18	137.36	36	49	30.56	19.68	312	441
BENZOCAINE	METHEMOGLOBIN	29	0.077	376.93	246.96	74	38	213.19	145.49	37	45	40.05	27.79	174	239
TROLAMINE	OTITIS EXT	14	0.007	1895.68	947.84	4	4	210.56	119.84	38	62	21.32	12.13	642	1004
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	209.95	164.94	39	36	75.42	59.2	48	56
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	207.82	182.29	40	31	135.44	118.88	13	14
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	207.07	187.02	41	30	156.35	141.35	9	10
SODIUM HYALURONATE	OPHTHALMITIS	20	0.038	529.86	317.91	43	24	206.43	129.74	42	52	28.84	18.12	345	517
PHENIRAMINE	MYDRIASIS	26	0.067	387.95	253.66	70	36	206.08	137.48	43	48	36.26	24.73	218	295
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	202.19	155.66	44	38	67.7	52.33	57	73
ISOSORBIDE MONONITRATE	HEADACHE VASC	40	0.141	283.71	198.59	140	54	199.82	144.69	45	46	50.34	36.71	110	138
EMBOLEX	VASOSPASM	19	0.038	505.85	292.86	45	29	196.33	121.75	46	59	27.63	17.47	375	543
SODIUM HYALURONATE	CORNEAL OPACITY	30	0.094	319.8	213.2	103	49	196.04	134.65	47	51	40.4	27.89	171	236
OXYTOCIN	UTER ATONY	32	0.105	306.03	200.83	117	52	195.37	135.86	48	50	42.48	30.06	150	199
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	192.06	164.53	49	37	114.11	97.77	19	20
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	190.14	144.60	50	47	62.3	47.23	74	85

Table G.8a: Top fifty drug and adverse event pairs selected by GPS using the lower bound (λ_{025}) of the 95% credible interval estimate of λ. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1206.22	1073.85	1	1	349.02	311.32	2	2
HEMOFIL	HIV TEST POS	306	0.198	1543.39	1371.9	9	3	1194.52	1065.95	2	2	358.53	320.87	1	1
PERDIEM	STENO ESOPH	65	0.024	2694.87	2072.98	3	1	791.17	615.57	3	3	95.5	74.3	29	35
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	644.88	582.98	4	4	323.51	293.66	3	3
PANALBA K.M.	DISCOLOR TOOTH	289	0.425	680.75	603.02	26	9	598.72	532.64	5	5	267.66	238.76	4	4
DIANEAL	PERITONITIS	150	0.228	658.94	553.51	28	11	524.7	445.7	6	6	170.1	144.86	7	9
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	452.71	407.21	9	7	256.19	230.43	5	5
ETIDOCAINE	TRISMUS	82	0.105	780.57	618.74	17	7	502.1	401.82	7	8	108.42	87.88	21	26
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	481.99	379.26	8	9	95.73	75.47	28	33
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.093	647.58	485.69	30	14	397.15	305.72	10	10	80.63	62.31	44	50
CHOLINE	OPHTHALMITIS	58	0.098	594.08	450.68	35	18	371.54	284.7	11	11	77	58.9	47	57
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.742	346.17	304.41	91	25	320.84	283.33	16	12	184.3	163.13	6	6
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.645	337.74	294.36	94	27	309.61	270.47	18	13	167.41	147.07	8	8
HYDROCOR'NE-NEOM-POLY B	PAIN EAR	37	0.041	895.62	629.35	13	6	370.77	264.93	12	14	53.2	37.52	94	131
DURANEST W/EPINEPHRINE	TRISMUS	46	0.073	628.21	450.67	33	19	349.13	258.63	13	15	63.54	47.75	66	81
CORTISPORIN	PAIN EAR	132	0.37	356.64	297.2	81	26	307.86	258.55	19	16	128.91	109.11	15	17
COLFOSCERIL PALMITATE	HEM LUNG	41	0.06	681.25	481.86	25	15	345.34	251.08	14	17	57.63	42.28	84	106
BSS	OPHTHALMITIS	28	0.031	889.14	571.59	14	10	310.66	237.21	17	18	41.27	28.45	164	224
PANCRELIPASE	STENO INTEST COLON	34	0.047	727.21	491.94	18	13	322.71	226.92	15	19	48.51	34.27	117	156
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.329	315.65	257.98	108	34	267.85	219.9	24	20	105.76	87.46	22	27
MYSTECLIN F	DISCOLOR TOOTH	134	0.458	292.78	244.72	130	39	259.43	218.26	25	21	120.37	102.14	18	18
BSS PLUS	OPHTHALMITIS	32	0.047	685.52	449.87	24	20	303.77	211.3	20	22	45.6	31.44	129	183
OXYTOCIN	HEM POSTPARTUM	103	0.346	297.96	242.99	124	40	254.6	208.79	26	23	103.26	85.04	24	29
BSS PLUS	CORNEAL OPACITY	52	0.129	402.19	293.91	65	28	276.51	208.57	21	24	66.71	50.1	61	75
DINOPROSTONE	UTER SPASM	56	0.147	381.61	286.21	72	30	272.47	207.8	22	25	70.17	53.31	54	71
METHYSERGIDE	FIBRO RETROPERIT	87	0.29	300.36	238.21	122	43	249.61	201.14	27	26	92.35	75.08	32	34
GEMCITABINE HYDROCHL'DE	CARCINOMA GI	225	0.952	236.39	205.92	200	51	222.58	194.84	34	27	139.78	122.77	12	12
BSS	KERATITIS	74	0.243	305.12	239.15	118	42	245.5	194.18	28	28	82.7	65.74	40	46
CHOLINE	CORNEAL OPACITY	77	0.26	295.96	230.62	128	46	241.31	191.75	29	29	84.46	67.31	37	42
TETRACYCLINE	ANOMALY TOOTH	380	1.776	213.96	192.57	234	56	207.07	187.02	41	30	156.35	141.35	9	10

Table G.8b: Top fifty drug and adverse event pairs selected by GPS using the lower bound (λ_{025}) of the 95% credible interval estimate of λ. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
DINOPROSTONE	LABOR ABNORM	231	1.052	219.5	191.94	224	58	207.82	182.29	40	31	135.44	118.88	13	14
CHLOROXINE	HAIR DISCOLOR	26	0.037	707.23	462.42	21	16	272.13	181.59	23	32	37.77	25.45	203	277
DOXYLAMINE	HANGOVER	44	0.125	351.27	247.48	88	37	238.87	175.58	30	33	56.39	41.71	86	108
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	233.87	173.75	31	34	59.21	44.52	80	95
NONOXYNOL	BALANITIS	51	0.166	306.82	228.61	116	47	226.43	170.41	33	35	62.6	47.67	72	82
SODIUM HYALURONATE	KERATITIS	70	0.274	255.2	196.87	167	55	209.95	164.94	39	36	75.42	59.2	48	56
DINOPROSTONE	FETAL DIS	167	0.81	206.1	175.25	252	67	192.06	164.53	49	37	114.11	97.77	19	20
BOTULINUM TOXIN A	PTOSIS	60	0.238	252.56	189.42	173	59	202.19	155.66	44	38	67.7	52.33	57	73
RIFABUTIN	UVEITIS	134	0.665	201.45	168.37	265	70	184.95	155.6	53	39	101.96	85.48	25	28
BENDECTIN	ANOMALY CONGEN	1106	6.785	163	153.42	378	77	161.58	152.26	63	40	148.51	139.79	11	11
DIETHYLSTILBESTROL	ANOMALY CONGEN	2811	17.749	158.37	152.57	398	79	157.84	152.09	65	41	152.79	147.32	10	7
METIPRANOLOL	IRITIS	22	0.035	623.42	368.39	34	22	233.41	150.03	32	42	32.09	20.9	284	401
MIVACURIUM CHLORIDE	INCREASED EFFECT	319	1.865	171.02	152.26	356	81	165.73	148.27	60	43	126.83	113.36	16	16
MYSTECLIN F	ANOMALY TOOTH	28	0.07	401.12	257.86	66	35	217.27	147.29	35	44	38.89	27.02	185	252
BENZOCAINE	METHEMOGLOBIN	29	0.077	376.93	246.96	74	38	213.19	145.49	37	45	40.05	27.79	174	239
ISOSORBIDE MONONITRATE	HEADACHE VASC	40	0.141	283.71	198.59	140	54	199.82	144.69	45	46	50.34	36.71	110	138
TICE BCG	GRANULOMA	55	0.23	239.15	178.27	192	64	190.14	144.6	50	47	62.3	47.23	74	85
PHENIRAMINE	MYDRIASIS	26	0.067	387.95	253.66	70	36	206.08	137.48	43	48	36.26	24.73	218	295
SODIUM HYALURONATE	IRITIS	21	0.038	551.61	341.47	39	23	216.18	137.36	36	49	30.56	19.68	312	441
OXYTOCIN	UTER ATONY	32	0.105	306.03	200.83	117	52	195.37	135.86	48	50	42.48	30.06	150	199

Table G.9a: Top fifty drug and adverse event pairs selected by the point estimate of RR. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
NISOLDIPINE	HEPATITIS NONSPECIFI	1	0.0003	3561.94	0	1	154780	2.4	0.95	29410	54652	1.83	0.25	61020	176165
ACARBOSE	HEPATITIS NONSPECIFI	1	0.0004	2816.96	0	2	154781	2.4	0.95	29473	54836	1.76	0.2	66479	209700
PERDIEM	STENO ESOPH	65	0.0241	2694.87	2072.98	3	1	791.17	615.57	3	3	95.5	74.3	29	35
TROLAMINE	OTITIS EXT	14	0.0074	1895.68	947.84	4	4	210.56	119.84	38	62	21.32	12.13	642	1004
ORTHO-NOVUM 1/80	CARCINOMA LIVER	1	0.0005	1840	0	5	154782	2.39	0.95	29583	54731	1.84	0.27	60769	171588
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1206.22	1073.85	1	1	349.02	311.32	2	2
KOATE	HIV SYND	7	0.0044	1578.59	451.03	7	17	107.86	47.23	125	277	10.77	4.83	2231	4217
HEMOFIL M	HEPATITIS HBSAG	6	0.0038	1577.25	525.75	8	12	92.71	37.64	168	404	9.37	3.93	2826	5759
HEMOFIL	HIV TEST POS	306	0.1983	1543.39	1371.9	9	3	1194.52	1065.95	2	2	358.53	320.87	1	1
PLATELET CON, HUMAN	LIVER DAMAGE AGGRAV	1	0.0008	1320.94	0	10	154783	2.38	0.94	29728	55175	1.8	0.22	63086	195535
URSODIOL	LIVER DAMAGE AGGR	1	0.0008	1187.47	0	11	154784	2.38	0.94	29782	55065	1.81	0.23	62640	188914
EMLA	HYPALGESIA	1	0.0009	1114.14	0	12	154785	2.37	0.94	29826	55171	1.8	0.23	63130	186757
HYDRO'NE-NEOM-POLY B	PAIN EAR	37	0.0413	895.62	629.35	13	6	370.77	264.93	12	14	53.2	37.52	94	131
BSS	OPHTHALMITIS	28	0.0315	889.14	571.59	14	10	310.66	237.21	17	18	41.27	28.45	164	224
BC POWDER	SALICYLISM	3	0.0034	885.18	0	15	154786	43.33	11.44	675	2181	4.75	1.39	9818	27036
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	481.99	379.26	8	9	95.73	75.47	28	33
ETIDOCAINE	TRISMUS	82	0.1051	780.57	618.74	17	7	502.1	401.82	7	8	108.42	87.88	21	26
PANCRELIPASE	STENO INTEST COLON	34	0.0468	727.21	491.94	18	13	322.71	226.92	15	19	48.51	34.27	117	156
ORTHO-NOVUM SQ	CARCINOMA LARYNX	1	0.0014	720.74	0	19	154787	2.35	0.94	30164	55426	1.78	0.2	65355	210519
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	644.88	582.98	4	4	323.51	293.66	3	3
CHLOROXINE	HAIR DISCOLOR	26	0.0368	707.23	462.42	21	16	272.13	181.59	23	32	37.77	25.45	203	277
COLFOSCERIL PALMITATE	INTEST SMALL PER	2	0.0029	699.11	0	22	154788	19.81	4.63	2282	7285	3.28	0.75	19648	65118
AVC	BALANITIS	1	0.0014	694.57	0	23	154789	2.35	0.93	30201	55765	1.8	0.22	63388	195229
BSS PLUS	OPHTHALMITIS	32	0.0467	685.52	449.87	24	20	303.77	211.3	20	22	45.6	31.44	129	183
COLFOSCERIL PALMITATE	HEM LUNG	41	0.0602	681.25	481.86	25	15	345.34	251.08	14	17	57.63	42.28	84	106
PANALBA K.M.	DISCOLOR TOOTH	289	0.4245	680.75	603.02	26	9	598.72	532.64	5	5	267.66	238.76	4	4
NORGESTREL	CARCINOMA CERVIX SIT	6	0.0089	677.3	225.77	27	48	85.66	34.83	197	457	9.37	4.09	2828	5415
DIANEAL	PERITONITIS	150	0.2276	658.94	553.51	28	11	524.7	445.7	6	6	170.1	144.86	7	9
HYDRO'NE-NEOM-POLY B	OTITIS EXT	2	0.0031	648.98	0	29	154790	19.7	4.62	2305	7295	3.21	0.75	20600	64747
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.0927	647.58	485.69	30	14	397.15	305.72	10	10	80.63	62.31	44	50

Table G.9b: Top fifty drug and adverse event pairs selected by the point estimate of RR. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
OCTREOTIDE ACETATE	LIVER DAMAGE AGGR	1	0.0015	646.31	0	31	154791	2.35	0.93	30276	55642	1.82	0.23	61642	186034
NORGESTREL	NEOPL CERVIX	5	0.0079	636.53	127.31	32	99	71.72	26.42	282	693	7.8	3.09	3931	8249
DURANEST W/EPINEPHRINE	TRISMUS	46	0.0732	628.21	450.67	33	19	349.13	258.63	13	15	63.54	47.75	66	81
METIPRANOLOL	IRITIS	22	0.0353	623.42	368.39	34	22	233.41	150.03	32	42	32.09	20.9	284	401
CHOLINE	OPHTHALMITIS	58	0.0976	594.08	450.68	35	18	371.54	284.7	11	11	77	58.9	47	57
DESOXIMETASONE	SKIN STRIAE	10	0.0169	590.6	236.24	36	45	129.98	66.08	91	153	15.18	8.12	1189	1944
CREON	STENO INTEST COLON	6	0.0104	576.05	192.02	37	57	83.7	34.01	200	478	9.47	4.13	2782	5330
DELADUMONE OB	UTER ATONY	1	0.0018	559.28	0	38	154792	2.34	0.93	30463	55725	1.84	0.22	60495	189568
SODIUM HYALURONATE	IRITIS	21	0.0381	551.61	341.47	39	23	216.18	137.36	36	49	30.56	19.68	312	441
PHENAPHEN W/CODEIN NO. 3	MYELOID MAT ARREST	1	0.0018	549.33	0	40	154793	2.34	0.93	30493	55650	1.8	0.22	63472	195022
CREON	STENO INTEST	4	0.0074	538.02	134.5	41	93	56.79	18.23	428	1135	6.4	2.26	5654	13169
NORLESTRIN	NEOPL BONE	1	0.0019	530.33	0	42	154794	2.34	0.93	30533	55733	1.8	0.22	63237	192170
SODIUM HYALURONATE	OPHTHALMITIS	20	0.0377	529.86	317.91	43	24	206.43	129.74	42	52	28.84	18.12	345	517
LOTRISONE	SKIN STRIAE	17	0.0323	526.76	278.87	44	32	185.63	111.66	52	69	25.05	15.4	474	692
EMBOLEX	VASOSPASM	19	0.0376	505.85	292.86	45	29	196.33	121.75	46	59	27.63	17.47	375	543
MIVACURIUM CHLORIDE	CHOLINESTERASE DEC	16	0.0319	502.05	282.4	46	31	175.3	103.72	55	77	23.76	14.59	529	753
PENBUTOLOL SULFATE	ADAMS STOKES SYND	1	0.002	500.08	0	47	154795	2.33	0.93	30638	55766	1.81	0.23	62560	187489
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	452.71	407.21	9	7	256.19	230.43	5	5
DOXACURIUM CHLORIDE	PARALYSIS FLACCID	16	0.033	484.79	272.69	49	33	173.11	102.42	56	79	23.48	14.13	539	792
PHYSOSTIGMINE	CHOLINESTERASE DEC	1	0.0021	482.23	0	50	154796	2.33	0.93	30681	56029	1.81	0.21	62662	197242

Table G.10a: Top fifty drug and adverse event pairs selected by RR using the lower bound (RR_{025}) of the 95% confidence interval estimate of RR. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
PERDIEM	STENO ESOPH	65	0.0241	2694.87	2072.98	3	1	791.17	615.57	3	3	95.5	74.3	29	35
IOPHENDYLATE	ARACHNOIDITIS	293	0.185	1583.73	1405.36	6	2	1206.22	1073.85	1	1	349.02	311.32	2	2
HEMOFIL	HIV TEST POS	306	0.1983	1543.39	1371.9	9	3	1194.52	1065.95	2	2	358.53	320.87	1	1
TROLAMINE	OTITIS EXT	14	0.0074	1895.68	947.84	4	4	210.56	119.84	38	62	21.32	12.13	642	1004
INPERSOL W/DEXTROSE	PERITONITIS	387	0.542	714.05	643.93	20	5	644.88	582.98	4	4	323.51	293.66	3	3
HYDROCOR'NE-NEOM-POLY B	PAIN EAR	37	0.0413	895.62	629.35	13	6	370.77	264.93	12	14	53.2	37.52	94	131
ETIDOCAINE	TRISMUS	82	0.1051	780.57	618.74	17	7	502.1	401.82	7	8	108.42	87.88	21	26
BERACTANT	HEM LUNG	71	0.089	797.58	617.84	16	8	481.99	379.26	8	9	95.73	75.47	28	33
PANALBA K.M.	DISCOLOR TOOTH	289	0.4245	680.75	603.02	26	9	598.72	532.64	5	5	267.66	238.76	4	4
BSS	OPHTHALMITIS	28	0.0315	889.14	571.59	14	10	310.66	237.21	17	18	41.27	28.45	164	224
DIANEAL	PERITONITIS	150	0.2276	658.94	553.51	28	11	524.7	445.7	6	6	170.1	144.86	7	9
HEMOFIL M	HEPATITIS HBSAG	6	0.0038	1577.25	525.75	8	12	92.71	37.64	168	404	9.37	3.93	2826	5759
PANCRELIPASE	STENO INTEST COLON	34	0.0468	727.21	491.94	18	13	322.71	226.92	15	19	48.51	34.27	117	156
DIETHYLSTILBESTROL	ANOMALY CONGEN UG	60	0.0927	647.58	485.69	30	14	397.15	305.72	10	10	80.63	62.31	44	50
COLFOSCERIL PALMITATE	HEM LUNG	41	0.0602	681.25	481.86	25	15	345.34	251.08	14	17	57.63	42.28	84	106
CHLOROXINE	HAIR DISCOLOR	26	0.0368	707.23	462.42	21	16	272.13	181.59	23	32	37.77	25.45	203	277
KOATE	HIV SYND	7	0.0044	1578.59	451.03	7	17	107.86	47.23	125	277	10.77	4.83	2231	4217
CHOLINE	OPHTHALMITIS	58	0.0976	594.08	450.68	35	18	371.54	284.7	11	11	77	58.9	47	57
DURANEST W/EPINEPHRINE	TRISMUS	46	0.0732	628.21	450.67	33	19	349.13	258.63	13	15	63.54	47.75	66	81
BSS PLUS	OPHTHALMITIS	32	0.0467	685.52	449.87	24	20	303.77	211.3	20	22	45.6	31.44	129	183
OPCON A	MYDRIASIS	351	0.717	489.53	439.33	48	21	452.71	407.21	9	7	256.19	230.43	5	5
METIPRANOLOL	IRITIS	22	0.0353	623.42	368.39	34	22	233.41	150.03	32	42	32.09	20.9	284	401
SODIUM HYALURONATE	IRITIS	21	0.0381	551.61	341.47	39	23	216.18	137.36	36	49	30.56	19.68	312	441
SODIUM HYALURONATE	OPHTHALMITIS	20	0.0377	529.86	317.91	43	24	206.43	129.74	42	52	28.84	18.12	345	517
DEMECLOCYCLINE	DISCOLOR TOOTH	257	0.7424	346.17	304.41	91	25	320.84	283.33	16	12	184.3	163.13	6	6
CORTISPORIN	PAIN EAR	132	0.3701	356.64	297.2	81	26	307.86	258.55	19	16	128.91	109.11	15	17
IMMUNE GLOBULIN, HUMAN	HEPATITIS C	218	0.6455	337.74	294.36	94	27	309.61	270.47	18	13	167.41	147.07	8	8
BSS PLUS	CORNEAL OPACITY	52	0.1293	402.19	293.91	65	28	276.51	208.57	21	24	66.71	50.1	61	75
EMBOLEX	VASOSPASM	19	0.0376	505.85	292.86	45	29	196.33	121.75	46	59	27.63	17.47	375	543
DINOPROSTONE	UTER SPASM	56	0.1467	381.61	286.21	72	30	272.47	207.8	22	25	70.17	53.31	54	71

Table G.10b: Top fifty drug and adverse event pairs selected by RR using the lower bound (RR_{025}) of the 95% confidence interval estimate of RR. The ranks assigned to the drug and event combinations based on RR, RR_{025}, λ (EBGM), λ_{025}, ϕ and ϕ_{025} are designated RK1, RK11 RK2, RK22, RK3 and RK33 respectively – continued.

Drug	Event	Count	Expected	RR	RR_{025}	RK1	RK11	EBGM	λ_{025}	RK2	RK22	ϕ	ϕ_{025}	RK3	RK33
MIVACURIUM CHLORIDE	CHOLINESTERASE DEC	16	0.0319	502.05	282.4	46	31	175.3	103.72	55	77	23.76	14.59	529	753
LOTRISONE	SKIN STRIAE	17	0.0323	526.76	278.87	44	32	185.63	111.66	52	69	25.05	15.4	474	692
DOXACURIUM CHLORIDE	PARALYSIS FLACCID	16	0.033	484.79	272.69	49	33	173.11	102.42	56	79	23.48	14.13	539	792
MIVACURIUM CHLORIDE	PARALYSIS FLACCID	104	0.3295	315.65	257.98	108	34	267.85	219.9	24	20	105.76	87.46	22	27
MYSTECLIN F	ANOMALY TOOTH	28	0.0698	401.12	257.86	66	35	217.27	147.29	35	44	38.89	27.02	185	252
PHENIRAMINE	MYDRIASIS	26	0.067	387.95	253.66	70	36	206.08	137.48	43	48	36.26	24.73	218	295
DOXYLAMINE	HANGOVER	44	0.1253	351.27	247.48	88	37	238.87	175.58	30	33	56.39	41.71	86	108
BENZOCAINE	METHEMOGLOBIN	29	0.0769	376.93	246.96	74	38	213.19	145.49	37	45	40.05	27.79	174	239
MYSTECLIN F	DISCOLOR TOOTH	134	0.4577	292.78	244.72	130	39	259.43	218.26	25	21	120.37	102.14	18	18
OXYTOCIN	HEM POSTPARTUM	103	0.3457	297.96	242.99	124	40	254.6	208.79	26	23	103.26	85.04	24	29
CALCIPOTRIENE	PSORIASIS	47	0.142	331	239.45	99	41	233.87	173.75	31	34	59.21	44.52	80	95
BSS	KERATITIS	74	0.2425	305.12	239.15	118	42	245.5	194.18	28	28	82.7	65.74	40	46
METHYSERGIDE	FIBRO RETROPERIT	87	0.2897	300.36	238.21	122	43	249.61	201.14	27	26	92.35	75.08	32	34
PANCRELIPASE	STENO INTEST	16	0.0378	423.38	238.15	61	44	164.43	97.32	61	86	23.26	13.97	547	805
DESOXIMETASONE	SKIN STRIAE	10	0.0169	590.6	236.24	36	45	129.98	66.08	91	153	15.18	8.12	1189	1944
CHOLINE	CORNEAL OPACITY	77	0.2602	295.96	230.62	128	46	241.31	191.75	29	29	84.46	67.31	37	42
NONOXYNOL	BALANITIS	51	0.1662	306.82	228.61	116	47	226.43	170.41	33	35	62.6	47.67	72	82
NORGESTREL	CARCINOMA CERVIX SIT	6	0.0089	677.3	225.77	27	48	85.66	34.83	197	457	9.37	4.09	2828	5415
SODIUM HYALURONATE	CORNEAL OPACITY	30	0.0938	319.8	213.2	103	49	196.04	134.65	47	51	40.4	27.89	171	236
APROTININ BOVINE	THROM CORONARY	13	0.0288	452.15	208.68	57	50	146.91	81.71	74	110	19.19	10.76	785	1243

References

[1] Agency for Healthcare Research and Quality. Reducing and preventing adverse drug events to decrease hospital cost. Research in Action, Issue 1, [online], Available at: `http://www.ahrq.gov/qual/aderia/aderia.htm`, 2001.

[2] I. Ahmed, F. Haramburu, A. Fourrier-Réglat, F. Thiessard, C. Kreft-Jais, G. Miremont-Salamé, B. Bégaud, and P. Tubert-Bitter. Bayesian pharmacovigilance signal detection methods revisited in multiple comparison setting. *Statistics in Medicine*, 28(13):1774–1792, 2009.

[3] I. Ahmed, F. Thiessard, G. Miremont-Salamé, B. Bégaud, and P. Tubert-Bitter. Pharmacovigilance data mining with methods based on false discovery rate: A comparative simulation study. *Clinical Pharmacology and Therapeutics*, 88(4):492–498, 2010.

[4] A. Alvarez-Requejo, A. Carvajal, B. Bérgaud, Y. Moride, T. Vega, and L. H. Martín Arias. Under-reporting of adverse drug reactions: Estimate based on a spontaneous reporting scheme and a sentinel system. *European Journal of Clinical Pharmacology*, 54(6):483–488, 1998.

[5] W. K. Amery. Why there is a need for pharmacovigilance. *Pharmacoepidemilogy and Drug Safety*, 8(1):61–64, 1999.

[6] A. Bate and S. J. W. Evans. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiology and Drug Safety*, 18(6):427–436, 2009.

[7] A. Bate, M. Lindquist, I. R. Edwards, S. Olsson, R. Orre, A. Lansner, and R. M. De Freitas. A bayesian neural network method for adverse drug reaction signal generation. *European Journal of Clinical Pharmacology*, 54(4):315–321, 1998.

[8] D. W. Bates, N. Spell, D. J. Cullen, E. Burdick, N. Laird, L. A. Petersen, S. D. Small, B. J. Sweitzer, and L. L. Leape. The costs of adverse drug events in hospitalized patients. *Journal of the American Medical Association*, 277(4):307–311, 1997.

[9] H. J. M. Beijer and C. J. de Blaey. Hospitalisations caused by adverse drug reactions (ADR): a meta-analysis of observational studies. *Pharmacy World and Science*, 24(2):46–54, 2002.

[10] B. P. Carlin and T. A. Louis. *Bayesian Methods for Data Analysis*. Chapman and Hall, Boca Raton, 3 edition, 2009.

[11] Carrasco-Garrido, P. and Ana de Andrés, L. and Barrera, V. H. and Ángel de Miguel, G. and Jiménez-García, R. Trends of adverse drug reactions related hospitalizations in spain (2001-2006). *BioMed Central*, 10(287):1–7, 2010.

[12] O. Caster, G. N. Norén, D. Madigan, and A. Bate. Large-Scale Regression-Based Pattern Discovery: The Example of Screening the WHO Global Drug Safety Database. *Statistical Analysis and Data Mining*, 3(4):197–208, 2010.

[13] Census Bureau. Population estimates. Technical report, US Census Bureau, Populaton Division, 4600 Silver Hill Road, Washington, DC 20233, 2010.

[14] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[15] D. C. Classen, S. L. Pestotnik, R. S. Evans, J. F. Lloyd, and J. P. Burke. Adverse drug events in hospitalized patients: Excess length of stay, extra costs, and attributable mortality. *Journal of the American Medical Association*, 277(4):301–306, 1997.

[16] E. C. Davies, C. F. Green, S. Taylor, P. R. Williamson, D. R. Mottram, and M. Pirmohamed. Adverse drug reactions in hospital in-patients: A prospective analysis of 3695 patient-episodes. *PLoS ONE*, 4(2):1–7, 2009.

[17] S. Davis, B. King, and J. M. Raine. Spontaneous reporting - UK. In R. D. Mann and E. B. Andrews, editors, *Pharmacovigilance*, chapter 15, pages 199–215. John Wiley and Sons Ltd., Chichester, 2007.

[18] Drugsite Trust. More about drug side effects. [online], Available at: `http://www.drugs.com/sfx/`, 2010.

[19] W. DuMouchel. Bayesian data mining in large frequency tables, with an application to FDA spontaneous reporting system. *The American Statistician*, 53(3):177–190, 1999.

[20] W. DuMouchel. Statistical issues in the analysis of spontaneous report databases. [online], Available at: `http://www.iom.edu/~/media/files/activity%20files/research/drugforum/dumouchel.pdf`, 2007.

[21] W. DuMouchel and D. Pregibon. Empirical Bayes screening for multi-item associations. In *KDD '01 Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 67–76. San Francisco, CA, USA, 2001.

[22] I. R. Edwards and J. K. Aronson. Adverse drug reactions: definitions, diagnosis, and management. *Lancet*, 356(9237):1255–1259, 2000.

[23] S. J. W. Evans. Pharmacovigilance: a science or fielding emergencies. *Statistics in Medicine*, 19(23):3199–3209, 2000.

[24] S. J. W. Evans. Statistics: Analysis and presentation of safety data. In J. Talbot and Waller P., editors, *Stephens's Detection of New Adverse Drug Reactions*, chapter 6, pages 301–328. John Wiley and Sons Ltd., Chichester, 5 edition, 2004.

[25] S. J. W. Evans, P. C. Waller, and S. Davis. Use of proportional ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*, 10(6):483–486, 2001.

[26] Food and Drugs Administration. About FDA product approval. [online], Available at: `http://www.fda.gov/NewsEvents/ProductsApprovals/ucm106288`, 2009.

[27] Food and Drugs Administration. Managing drug safety issues. [online], Available at: `http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm1078`, 2009.

[28] Food and Drugs Administration. The Adverse Event Reporting System (AERS). [online], Available at: `http://www.fda.gov/Drugs/InformationOnDrugs/ucm135151.htm`, 2009.

[29] Food and Drugs Administration. Latest Quarterly Data Files - Asc_nts.doc/README.doc. [online], Available at: `http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInfo`, 2010.

[30] Food and Drugs Administration. The Adverse Event Reporting System (AERS) Statistics. [online], Available at: `http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInfo`, 2010.

[31] Food and Drugs Administration. Drug safety communications. [online], Available at: `http://www.fda.gov/Drugs/DrugSafety/ucm199082.htm.`, 2011.

[32] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[33] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Beyasian Data Analysis*. Chapman and Hall, London, 2 edition, 2004.

[34] P. Ghosh and A. Dewanji. Analysis of spontaneous adverse drug reaction (ADR) reports using supplementary information. *Statistics in Medicine*, 30(16):2040–2055, 2011.

[35] M. Goettler, S. Schneeweiss, and J. Hasford. Adverse drug reaction monitoring – cost and benefit considerations Part II: Cost and preventability of adverse drug reactions leading to hospital admission. *Pharmacoepidemiology and Drug Safety*, 6(Suppl.3):S79 – S90, 1997.

[36] A. L. Gould. Practical pharmacovigilance analysis strategies. *Pharmacoepidemiology and Drug Safety*, 12(7):559Ũ574, 2003.

[37] Government Accountability Office. FDA drug review: Postapproval risks, 1976-85. Technical report, US Government Accountability Office (GAO), Program Evaluation and Methodology Division, Washington, DC, 1990.

[38] Government Accountability Office. Adverse drug events: Substantial problem but magnitude uncertain. Technical report, US Government Accountability Office (GAO), Health, Education and Human Services Division, Washington, DC, 2000.

[39] D. G. Grahame-Smith and J. K. Aronson. *Oxford textbook of clinical pharmacology and drug therapy*. Oxford University Press, Oxford, 1984.

[40] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[41] S. D. Guikema and J. P. Goffelt. A flexible count data regression model for risk analysis. *Risk Analysis*, 28(1):213–223, 2008.

[42] K. Y. Hartigan-Go and J. Q. Wong. Inclusion of therapeutic failures as adverse drug reactions. In J. K. Aronson, editor, *Side effects of drugs, Annual 23 – A world-wide yearly survey of new data and trends in adverse drug reactions*. Essay, Elsevier, Amsterdam, 2000.

[43] J. B. Kadane, G. Shmueli, T. P. Minka, S. Borle, and P. Boatwright. Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Analysis*, 1(2):363–374, 2006.

[44] R. King, B. J. T. Morgan, O. Gimenez, and S. P. Brooks. *Bayesian Analysis for Population Ecology*. Chapman and Hall, London, 2010.

[45] J. Lazarou, B. H. Pomeranz, and P. N. Corey. Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-analysis of Prospective Studies. *Journal of the American Medical Association*, 279(15):1200–1205, 1998.

[46] D. Madigan, P. Ryan, S. Simpson, and I. Zorych. Bayesian methods in pharmacovigilance. In J. M. Bernado, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and West M., editors, *Bayesian Statistics*, volume 9, pages 1–17. Oxford University Press, 2010.

[47] J. Marin and C. P. Robert. *Bayesian core – A Practical Approach to Computational Bayesian Statistics*. Springer, New York, 2007.

[48] Medicines and Healthcare Products Regulatory Agency. How we regulate. [online], Available at: `http://www.mhra.gov.uk/Howweregulate/index.htm`, 2005.

[49] Medicines and Healthcare Products Regulatory Agency. Periodic safety update reports. [online], Available at: `http://www.mhra.gov.uk/Howweregulate/Medicines/Licensingofmedicines/Informationforlicenceapplicants/PeriodicSafetyUpdateReports/CON2030706`, 2010.

[50] Medicines and Healthcare Products Regulatory Agency. Pharmacovigilance - how we monitor the safety of medicines. [online], Available at: `http://www.mhra.gov.uk/Safetyinformation/Howwemonitorthesafetyofproducts/Medicines/Pharmacovigilance/index.htm`, 2010.

[51] Merck and Company, Inc. Types of adverse drug reactions. [online], Available at: `http://www.merck.com/mmhe/sec02/ch015/ch015b.html`, 2007.

[52] N. Moore, C. Kreft-Jais, F. Haramburu, Noblet C., M. Andrejak, M. Ollagnier, and B. Béaud. Reports of hypoglycaemia associated with the use of ACE inhibitors and other drugs: a case/non case-study in the French pharmacovigilance system database. *British Journal of Clinical Pharmacology*, 44(5):513–518, November 1997.

[53] T. J. Moore, M. R. Cohen, and C. D. Furberg. Serious adverse drug events reported to the food and drugs administration, 1998-2005. *Journal of the American Medical Association*, 167(16):1752–1759, 2007.

[54] T. J. Moore, B. M. Psaty, and C. D. Furberg. Time to act on drug safety. *Journal of the American Medical Association*, 279(19):1571–1573, 1998.

[55] N. Muehlberger, S. Schneeweiss, and J. Hasford. Adverse drug reaction monitoring – cost and benefit considerations Part I: Frequency of adverse drug reactions causing hospital admissions. *Pharmacoepidemiology and Drug Safety*, 6(Suppl.3):S71 – S77, 1997.

[56] P. Müller, G. Parmigiani, C. Robert, and J. Rousseau. Optimal sample size for multiple testing: The case of gene expression microarrays. *Journal of American Statistical Association*, 99(468):990–1001, 2004. Theory and Methods.

[57] National Center for Health Statistics. Health, United States 2010; with special feature on death and dying. Technical report, US National Center for Health Statistics (NCHS), Hyattsville, MD, 2011.

[58] G. N. Norén, A. Bate, R. Orre, and I. R. Edwards. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Statistics in Medicine*, 25(21):3740–3757, 2006.

[59] G. N. Norén, J. Hopstadius, and A. Bate. Shrinkage observed-to-Expected ratios for robust and transparent large-scalepattern discovery. *Statistical Methods in Medical Research*, 22(1):57–65, November 2011.

[60] I. Ntzoufras. *Bayesian modeling using WinBUGS*. John Wiley and Sons Ltd., New Jersey, 2009.

[61] R. Orre, A. Lansner, A. Bate, and M. Lindquist. Bayesian neural networks with confidence estimations applied to data mining. *Computational Statistics and Data Analysis*, 34(4):473Ũ493, 2000.

[62] PDR.net. Full precribing infomation. [online], Available at: `http://www.pdr.net`, 2013.

[63] W. J. Pichler. Adverse drug reactions. [Online], Available at: `http://www.uptodate.com/contents/drug-allergy-classification-and-clinical-features?`, 2013.

[64] M. Pirmohamed, A. M. Breckenridge, N. R. Kitteringham, and B. K. Park. Adverse drug reactions. *British Medical Journal*, 316(7140):1295–1298, 1998.

[65] M. Pirmohamed, S. James, S. Meakin, C. Green, A. K. Scott, T. J. Walley, K. Farrar, B. K. Park, and A. M. Breckenridge. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18820 patients. *British Medical Journal*, 329(7456):15–19, 2004.

[66] M. Pirmohamed and B. K. Park. Adverse drug reactions: back to the future. *British Journal of Clinical Pharmacology*, 55(5):486–492, 2003.

[67] M. Pirmohamed and B. K. Park. Mechanisms of adverse drug reactions. In R. D. Mann and E. B. Andrews, editors, *Pharmacovigilance*, chapter 8, pages 85–103. John Wiley and Sons Ltd., Chichester, 2 edition, 2007.

[68] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, Available at: `http://www.R-project.org/`, 2011.

[69] M. D. Rawlins. Spontaneous reporting of adverse drug reactions I: The data. *British Journal of Clinical Pharmacology*, 26(1):1–5,7–11, 1988.

[70] M. D. Rawlins, G. N. Fracchia, and E. Rodriguez-Farré. EURO-ADR: Pharmacovigilance and research (A european perspective). *Pharmacoepidemiology and Drug Safety*, 1(5):261–268, 1992.

[71] M. D. Rawlins and J. W. Thompson. Mechanisms of adverse drug reactions. In D. M. Davies, editor, *Textbook of Adverse Drug Reactions*, chapter 3, pages 18–45. Oxford University Press, Oxford, 1991.

[72] C. P. Robert and G. Casella. *Introducing Monte Carlo methods with R*. Springer, New York, 2010.

[73] R. J. Royer. Mechanism of action of adverse drug reactions: An overview. *Pharmacoepidemilogy and Drug Safety*, 6(Suppl.3):S43–S50, 1997.

[74] RPSGB and BMJG. *British National Formulary*. Pharmaceutical Press, Royal Pharmaceutical Society of Great Britain and British Medical Journal Group, London, 65 edition, 2013.

[75] RxList Inc. RxList: The internet drug index. [online], Available at: `http://www.rxlist.com/script/main/hp.asp`, 2013.

[76] SAS Institute Inc. *SAS software*. Copyright©(2002-2010), SAS and all other SAS Institute Inc. product or service names are registered trademarks of SAS Institute Inc., Cary, NC, USA, 2010.

[77] S. J. Senn. *Statistical Issues in Drug Development*. John Wiley and Sons Ltd., Chichester, 2007.

[78] G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright. A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Applied Statistics*, 54(1):127–142, 2005.

[79] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of Royal Statistical Society B*, 64(4):583–639, 2002.

[80] Statisticat and LLC. *Bayesian Inference*. R package version 13.03.04, Available at: `http://cran.r-project.org/web/packages/LaplacesDemon/index.html`, 2013.

[81] W. P. Stephenson and M. Hauben. Data mining for signals in spontaneous reporting databases: proceed with caution. *Pharmacoepidemiology and Drug Safety*, 16(4):359–365, 2007.

[82] A. Szarfman, S. G. Machado, and R. T. O'Neill. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Safety*, 25(6):381–392, 2002.

[83] The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Medical dictionary for regulatory activities (MedDRA). [online], Available at: `http://www.ich.org/products/meddra.html`, 2013.

[84] P. Wiffen, M. Gill, J. Edwards, and A. Moore. Adverse drug reactions in hospital patients – a systematic review of prospective and retrospective studies. *Bandolier Extra*, pages 1–15, 2002.

[85] World Book. Health. In *World Book Encyclopedia*, volume H, pages 94–97. World Book International Inc., London, 1994.

[86] World Health Organisation. Global intelligence network for benefits and risk in medical products. [online], Avalable at: `http://www.who-umc.org/DynPage.aspx`, 2002.

[87] World Health Organisation. *Viewpoint Part I.* WHO Collaborating Centre for International Drug Monitoring, Uppsala Monitoring Centre (UMC), Stora Torget 3, S-753 20 Uppsala, Sweden, 2002.

[88] World Health Organisation. *Viewpoint Part II.* WHO Collaborating Centre for International Drug Monitoring, Uppsala Monitoring Centre (UMC), Stora Torget 3, S-753 20 Uppsala, Sweden, 2004.

[89] D. K. Wysowski and L. Swartz. Adverse drug event surveillance and drug withdrawals in the united states, 1969-2002. *Journal of the American Medical Association*, 165(12):1363–1369, 2005.