Measuring health inequalities in Malta:
Development and application of a small area-based deprivation index

Dorothy Galea Gauci

B.Psy (Hons)(Melit), MSc.(DLSHTM)

Submitted in fulfilment of the requirements for the Degree of
Doctor of Philosophy

School of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University of Glasgow

October 2023

Abstract

Background - Improvement in population health requires that health systems focus on social determinants as important risk factors. Despite some limitations, small area-based deprivation indices have been useful in the analysis of health inequalities. These indices can be linked to data sources through small area-based geographical information. The aim of this thesis was to develop and validate a small area-based deprivation index for Malta using census data and apply it to the analysis of health inequalities.

Methods - Data from two censuses were used to create the index. Nineteen area level items were considered. After Z-standardisation, principal component analysis was used for item reduction and a weighted additive score was constructed. Quintiles of deprivation from least to most deprived were produced. The index was applied to the analysis of deprivation gradients for mortality and cancer incidence using age-standardised rates. The Relative Index of Inequality (RII) was estimated. Multi-level logistic regression models were used to analyse the association between individual socioeconomic factors and the area deprivation measure on two self-reported health outcomes and mortality.

Results - The final index contained five items within the domains of education, employment and living conditions. Premature mortality rates declined between the two periods, however the linear gradient for premature mortality by deprivation quintile, increased. This was most notable for males in the most deprived areas, where the premature mortality rate in the most recent period increased. Results for old-age mortality were less consistent. The relationship between deprivation and cancer incidence was not consistent, with negative and positive gradients observed for different sites. Results for bronchus-lung cancer incidence presented the strongest deprivation gradient. In the multi-level models, a gradient was still seen with increasing deprivation for all health outcomes, though the odds were attenuated when including individual predictors. The strongest relationships were maintained for the most deprived quintile.

Discussion - The small area-based deprivation index presented deprivation gradients for mortality and cancer incidence. Even when including individual predictors of socioeconomic status, the area measure still contributed to the analysis of health inequalities. The findings were generally in line with what was found elsewhere. The index developed has wider implications for research, evidence-based policy making and practice in Malta.

Conclusion- Area-based indices assign a level of deprivation to an area and not individuals, and thus may be prone to ecological fallacy. Despite this, there is continued value in their use in health inequalities research.

Contents

List of tables

List of figures

Acknowledgements

Firstly, I would like to acknowledge the Ministry for Education, Sport, Youth, Research, and Innovation - the research presented in this thesis was part-financed by the ENDEAVOUR Scholarships Scheme (Group B).

I would like to thank my supervisors: Professor Alastair H Leyland and Professor Ruth Dundas for their expertise, constant guidance, and support over the course of my PhD. There were many obstacles along the way, and without their encouragement and genuine belief in the value of this research, I do not believe I would have managed to get this far. It has been a privilege to work with such highly respected academics.

I am indebted to the National Statistics Office and Directorate for Health Information and Research for providing the data that was used in this research. I would especially like to thank the National Statistics Office for their support, as my employer, along the duration of my research and my colleagues, for their constant encouragement, especially during the more difficult periods.

I am thankful for all the life lessons that I have received from my parents. Despite their humble background, they always instilled within me the value of education and were an example of living with generosity of spirit and social consciousnesseses. I am sure that their example has impacted the decisions I have taken along my academic and professional career, and ultimately influenced my interest in social epidemiology.

Finally, I must thank the two most important men in my life. Matthew, thank you for standing by me throughout this journey. It has been long and arduous and there have been many times that I have wanted to give up. You have been beyond patient and supportive. I am forever grateful for the many times that you carried the load of life's responsibilities to give me space to work. Benjamin, thank you for giving my life and work deeper meaning. Though you are too young to understand what this all means, I hope I make you proud.

Author's Declaration

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

List of abbreviations

ASIR – Age-Standardised Incidence Rate

ASMR – Age-Standardised Mortality Rate

ARP– At-risk-of-poverty rate

CI – Confidence Interval

DHIR – Directorate for Health Information and Research

EA – Enumeration Area

EFTA – European Free Trade Association

EHIS – European Health Interview Survey

ESP – European Standard Population

ESS – European Statistical System

EU – European Union

HBSC – Health Behaviour in School-Aged Children

ICC - Intra-class Correlation Coefficient

ICD-O – International Classification of Diseases for Oncology

ILO – International Labour Organisation

IMD – English Index of Multiple Deprivation

IPA – International Protection Agency

IRSD - Index of Relative Social Deprivation

ISCED – International Standard Classification of Education

ISCO – International Standard Classification of Occupation

LAU - Local Administrative Units

LFS – Labour Force Survey

MOR - Median Odds Ratio

MQL - Marginal Quasi-Likelihood

NUTS - Nomenclature of Territorial Units for Statistics

NZDep – New Zealand Deprivation Index

OECD – Organisation for Economic Co-operation and Development

PPI – Property Price Index

PQL - Penalised Quasi-Likelihood

PYLL – Potential Years of Life Lost

RII – Relative Index of Inequality

SD – Standard Deviation

SILC – Social Income and Living Conditions

SII – Slope Index of Inequality

SIMD – Scottish Index of Multiple Deprivation

UNECE – United Nation Economic Commission for Europe

UNESCO – United Nations Education, Culture and Science Organisation

UPA – Underprivileged Areas

VANDIX - Vancouver Area Neighbourhood Deprivation Index

WHO – World Health Organisation

ZCTA – Zip Code Tabulation Areas

Chapter 1: Introduction

1.1 Socioeconomic determinants of health

It has long been recognised by researchers that there are socioeconomic determinants of health. Since the 1980's and publications such as the Black report, researchers have presented a wealth of evidence showing that socioeconomic inequalities in health do exist and improvements in population health outcomes may require health planners, managers and policy makers to look beyond simply addressing clinical or lifestyle risk factors (Marmot, 2001). While policy makers and politicians may have been dismissive of the findings of the Black report when it was initially published, a recent resurgence and focus on health inequalities has put this topic squarely on the policy agenda, even at the European Union (EU) level (Marmot *et al.*, 2012).

Despite the fact that over thirty years have passed since the Black report was published, it is widely acknowledged that socioeconomic inequalities in health still persist, even in developed countries (Mackenbach, 2012, 2014) and these disparities have been found to have a substantial economic impact (Mackenbach, Meerding and Kunst, 2011). Research has shown that there are marked differences between high and low socioeconomic groups for a variety of health outcomes, including all-cause and cause specific mortality, morbidity of common chronic conditions, maternal and newborn outcomes, and measures of self-rated health. The degree of variation between the socioeconomic groups also differs between countries (Mackenbach *et al.*, 2003, 2008; Blumenshine *et al.*, 2010; Bleich *et al.*, 2012; Huisman *et al.*, 2013; E. Wang *et al.*, 2020). Studies amongst school-aged children also show that gaps in health outcomes and healthy behaviours by social affluence are present across the life course (Elgar *et al.*, 2015).

The recent push to address health inequalities, has helped to highlight the need for more research in this field, if health systems are to be geared towards addressing current socioeconomic gaps in health (World Health Organisation, 2008). However, measuring and understanding socioeconomic differences in health for regular monitoring, may not be easy. Much of the research in health inequalities has made use of more readily available self-reported data from cross-sectional health surveys which allow for comparison across a larger number of countries. Such data however only offers a snapshot of health status, suffer from issues of temporality and interpretation of findings, and comparisons may not be easy when considering possible differences in methodologies. Furthermore, surveys may suffer from low response, response bias and may be expensive and time consuming to conduct on a regular basis. Surveys may also underrepresent small subsets of the population or exclude certain hidden populations all together. Some large follow up studies such as the Whitehall study amongst British civil servants (Marmot *et al.*, 1991) have helped to address the issue of temporality seen in cross-sectional surveys by following up a large cohort, however this methodology proves time consuming, expensive and not cost-effective for regular monitoring. Large country comparative studies using registry and administrative data sources, alternatives to cross-sectional surveys, tend to be geographically biased towards Nordic countries and Western Europe, where there is a well-established history in the field of health inequalities research. In other areas such as Southern Europe, such country level data is not always available and may be restricted to major cities or urban areas (Mackenbach *et al.*, 2008; Huisman *et al.*, 2013). While countries may have several health registers and administrative sources, they may not routinely collect any data on individual level socioeconomic characteristics of the population being considered. The lack of a cohesive health information system, poor quality data, and a limited ability to disaggregate data that is available by socioeconomic characteristics, may limit countries ability to develop and sustain a health inequality monitoring system (Hosseinpoor, Bergen and Schlotheuber, 2015). The adoption of area-based measures may enhance the measurement and monitoring of health inequalities by increasing the opportunities for the use of health data sources available within health systems (Hosseinpoor and Bergen, 2016).

1.2 Small area-based deprivation indices

Small area-based deprivation indices have been used for several years, especially in the United Kingdom, to understand the differences in socioeconomic status across geographical areas and to see how small area-based deprivation relates to certain outcomes, including health. Small areas used to create these indices generally group individuals into geographically defined small output units. These are usually grouped based on known boundaries such as post-code, census tracts/enumeration areas or any other geographic boundaries that can be used to combine persons into small areas that are expected to capture homogenous groups. The index outputs can be linked to any data source where the same geographical breakdown is available, thus facilitating the analysis of any outcome measure by area level deprivation. While area-based indices assign a level of deprivation to an area as a whole and not individuals, and thus may be prone to the ecological fallacy, they have proven useful tools in the study of health inequalities, especially when data on individual level socioeconomic characteristics are not available in the data being analysed. This may be because socioeconomic characteristics tend to cluster in areas (Chyn and Katz, 2021). The environment of the area itself may also play an important role in the relationship found between area-based deprivation measures and health. These include exposure to pollution, access to open and green spaces, walkability, access to healthy foods, and the level of social cohesion and safety (Diez Roux, 2018).

Essentially, being deprived means being denied something that is considered necessary or essential to live and work fully and meaningfully within society. There are many different domains of deprivation under the umbrella of social deprivation. The traditional income-based measure of material deprivation focuses on the impact limited financial resources have on social participation and is the measure most linked to the concept of poverty. Other measures of deprivation may relate to employment deprivation or education deprivation, focusing on being deprived of employment or education opportunities. A concept that is related to social deprivation, is social exclusion, which looks at group factors which exclude persons from participating in society, rather than income or poverty related issues, such as disability, gender, ethnicity, or age (Chandola and Conibere, 2015). Deprivation is relative, in that, to classify someone as deprived it must be relative to the standard level of resources, opportunities and social conditions available to the majority of members within the society in which deprivation is being measured (Townsend, 1987). Deprivation is also not a dichotomous measure, in that society is not split into those who are deprived and those who are not deprived. Deprivation is measured along a gradient and its impact on health can also be seen to move along this gradient.

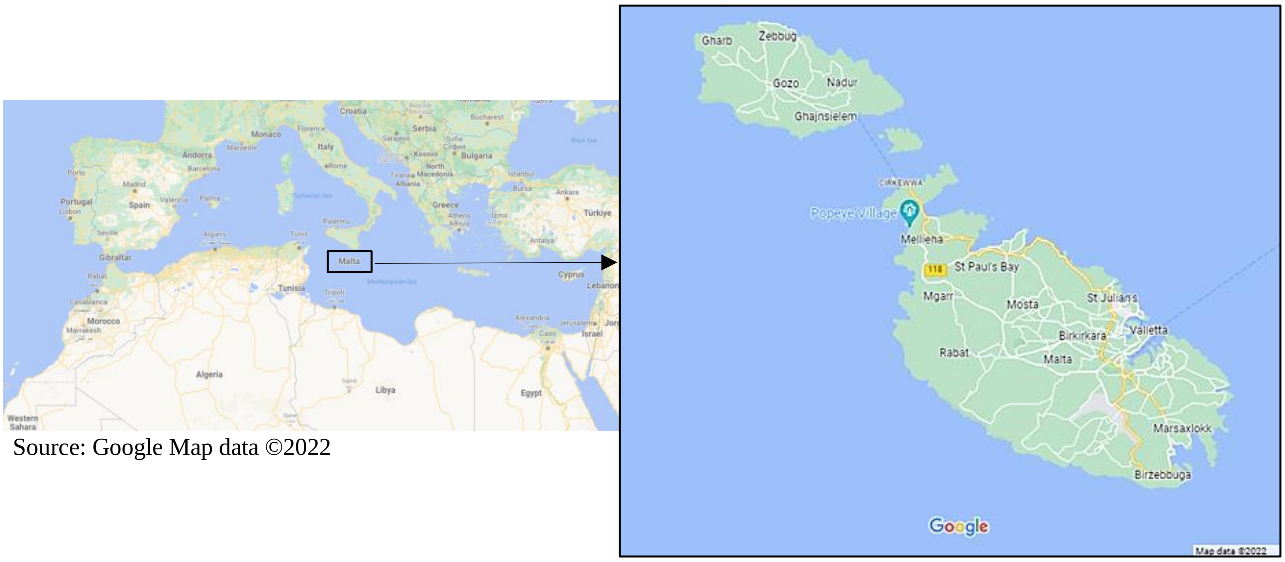
Many small area-based deprivation indices are developed with data derived from the enumeration of persons and dwellings through the census, which is usually conducted decennially. The advantage of using the census is that it considers the entire enumerated population and is not a sample. This is usually strengthened by the fact that in most countries, the census is governed through some kind of legal act obliging all residents to participate. Furthermore, the core questions in the census generally remain constant over time and in some cases, across countries. This is especially the case within the European Union (EU), where legally binding regulations ensure that all EU Member States implement a core set of questions in their national level census (European Parliament, 2008). This allows for reasonable comparison within countries, over time, and between countries. One of the biggest limitations of using the census, however, is that it is conducted every ten years, therefore information emerging from it may become outdated as the inter-census period moves forward.

While a traditional census involves the direct enumeration of persons and dwellings, usually through door-to-door canvassing by enumerators, more countries are moving towards register-based or combined approaches. The register-based census developed by Nordic countries in the 1970s, involves no direct collection of data with the enumerated population, but enumeration of residents is done using various individual level population registers. The combined approach uses elements of both the traditional method and the registered method by combining the administrative data with a limited collection of some parts of the data through field enumeration. The reasons for moving away from the traditional approach may vary from country to country, however the popularity of these methods must be acknowledged – based on information gathered in 2018 by the United Nation Economic Commission for Europe (UNECE), 69% of EU and European Free Trade Association (EFTA) countries planned to use either the fully registered or combined approach for the 2021 census (UNECE, 2019). Taking into consideration these developments, a census can therefore be defined as the national level enumeration of all individuals deemed resident in the specific country territory, irrespective of whether enumeration is conducted through traditional, register or combined methods.

When discussing small area-based deprivation indices one must acknowledge the emergence of indices of multiple deprivation which use multiple data sources. Such indices include, amongst others, the English Index of Multiple Deprivation (IMD) and the Scottish Index of Multiple Deprivation (SIMD) (Ministry of Housing Communities and Local Government, 2019; The Scottish Government, 2020). Multiple deprivation indices use the same concept of developing a deprivation score at the small area level, however deprivation is measured across several domains, using several data sources. In the case of both the English and Scottish indices, for example, it is over seven distinct domains – income, employment, education, health, crime, housing, and access to services. Most of the data used for the indicators is sourced from administrative databases such as benefit and employment records. Census data is used for a minority of indicators when alternative data from administrative sources is not available. For the purposes of this thesis, such indices will not be considered. This is primarily because several limitations in administrative sources available in Malta do not allow for the development of a similar index. Administrative sources in Malta tend to be developed as standalone registers, with varying degrees of coverage, quality, and structure. Location data within administrative sources in Malta can also be problematic. Administrative data are not available at the enumeration area level, as these areas are developed exclusively for the census and are not standard areas that can be mapped easily to other data sources. The use of post-codes is also limited. The post-code system was updated in 2008 and many administrative sources either historically did not include this field entirely or have a large amount of missing information as addresses were not updated, or individuals did not supply a post-code when entering the system. While address data may be available in administrative sources, the quality and structure across sources varies considerably as no national address register exists. Currently, this makes it difficult to combine such administrative data together to develop a small area-based deprivation index, therefore this thesis will focus on small area-based deprivation indices developed using the census.

1.3 The Maltese Islands

The Maltese Islands, referred to as Malta, is an archipelago located in the middle of the Mediterranean Sea, sandwiched between Europe and North Africa. There are two inhabited islands within the archipelago – Malta, the main island, and Gozo, the smaller sister island. The two islands are connected through a regular ferry service, no airport is presently available on the sister island (figure 1.1).

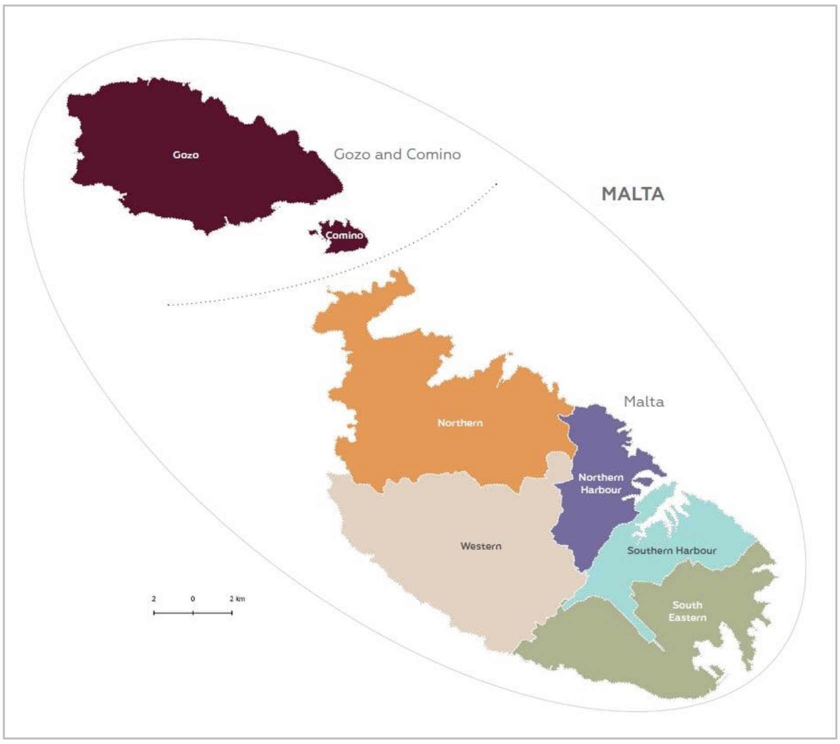


Source: Google Map data ©2022

Figure 1.1 Map of the Maltese Islands

Malta is part of the Commonwealth and obtained independence from the United Kingdom in 1964. The social, education and health systems in Malta are heavily influenced by the colonial history of the island with a similar education system, social security system and national health service as found in the United Kingdom. Malta joined the EU in 2004 and is the smallest country in the bloc, both with respect to land mass and population. Malta is the most densely populated EU Member State. At the end of 2022, the total population of Malta stood at 542,051, while the population density for a country area of approximately 315km^2 was 1,720 inhabitants per km^2 (National Statistics Office, 2023j).

The two inhabited islands in the archipelago are divided into 68 official localities, 14 in Gozo and the rest in Malta. The localities denote the official boundaries set up for the purposes of local government administration as outlined by the local government act (Government of Malta, 1993). The population size of the localities as at the end of 2022, ranged from 35,419 to 202. The average size of the localities was slightly over 7,900 (National Statistics Office, 2023d). The National Statistics Office (NSO) divided Malta into six broad districts of residence for the purposes of statistical spatial analysis – Gozo is considered as a district on its own, while the island of Malta is further divided into five districts – North Harbour, South Harbour, North, West, and South-East. These districts were first created for the spatial analysis of data from the 1967 census. The grouping of the localities into these specific regions was based on several factors, including, population densities per unit area, urban and rural characteristics, the concentration of industrial and other economic activity in the port areas, and the small size of the islands, in general. The original groupings of the two harbour districts were the Inner and Outer Harbours (Census Office Malta, 1968). These two districts were updated into the North Harbour and South Harbour for population data published in 2000 and have remained grouped in this way since then. Nine of the 14 localities previously grouped into the Inner Harbour region were combined with 4 localities previously in the Outer Harbour to create the Southern Harbour. Eight of the 12 localities previously grouped as the Outer Harbour were combined with the remaining five localities previously in the Inner Harbour to create the Northern Harbour. No information was provided in the first report where these new districts were used explaining the reason for the regrouping (National Statistics Office, 2001). The North Harbour is the largest district, in terms of population size, with 31.1% of the total population as at end 2022 living in this area (National Statistics Office, 2023d). Figure 1.2 maps the 6 districts in their current format.



Source: National Statistics Office ©2022

Figure 1.2: The six districts of the Maltese Islands

The 68 localities and 6 districts have been adopted as the official spatial breakdowns for Malta by Eurostat, the European Commission statistical body, for comparative spatial dissemination as part of the system of Local Administrative Units (LAU) within the broader nomenclature of territorial units for statistics (NUTS) classification. The districts correspond to LAU1 within the classification, while the localities correspond to LAU2 (National Statistics Office, 2023c). Since 2017, the NSO has published an annual report on regional statistics which presents spatial analysis of statistics available across the NSO using mainly the 6 districts or the 68 localities, if data is available at this level (National Statistics Office, 2017). So, even though the original six districts were created nearly 60 years ago, they are still being used today for spatial analysis of official statistics.

The population of Malta has grown considerably since 2013, from 428,156 to 542,051 in 2022, an increase of 26.7% (National Statistics Office, 2024). The change in government following the 2013 election led to a change in economic policies which encouraged migration into the country, primarily to fill gaps in the labour market with foreign workers, as the economy grew following the economic downturn in 2010 (Grech, 2016a). As a result, the share of the foreign population increased rapidly over a ten year period, from 6.8% at the end of 2013, to 25.3% by the end of 2022 (National Statistics Office, 2024). Net migration has increased annually since 2013, with only a slowdown in 2020, brought on by the COVID-19 pandemic. In 2022, however, the trend continued with a net migration of 21,798 persons, the highest recorded net migration in the country since statistics have been published. The population in 2022 grew by 4.2% compared to the previous year, with 83.1% of the net migrants being non-EU citizens. Males made up 65.8% of these net migrants. On the other hand, the natural increase for 2022 was the lowest recorded in the past 15 years, at just 79 (National Statistics Office, 2023j). Since 2002, there has also been a sizeable number of irregular migrants who have sought asylum in Malta due, in part, to regional instability in their home countries and the location of Malta in the Mediterranean. In 2019 alone, the International Protection Agency (IPA) received 4,090 applications for asylum (National Statistics Office, 2019). These shifts in the demographic trends over the past 10 years has led to changes in the population structure. Figure 1.3 shows the population pyramid for the population as at end 2011, while figure 1.4 shows the pyramid for the population as at end 2022. The share of the population aged less than 15 has decreased from 15% in 2011 to 12.7% in 2022, while the share aged 75 and over has increased from 7% to 8.2%. Due to the large increases in migrants in recent years, the share of males in the population has also overtaken the share of females, especially in the working age population. In 2011, 49.7% of the population was male, by 2022 this increased to 52.5%.

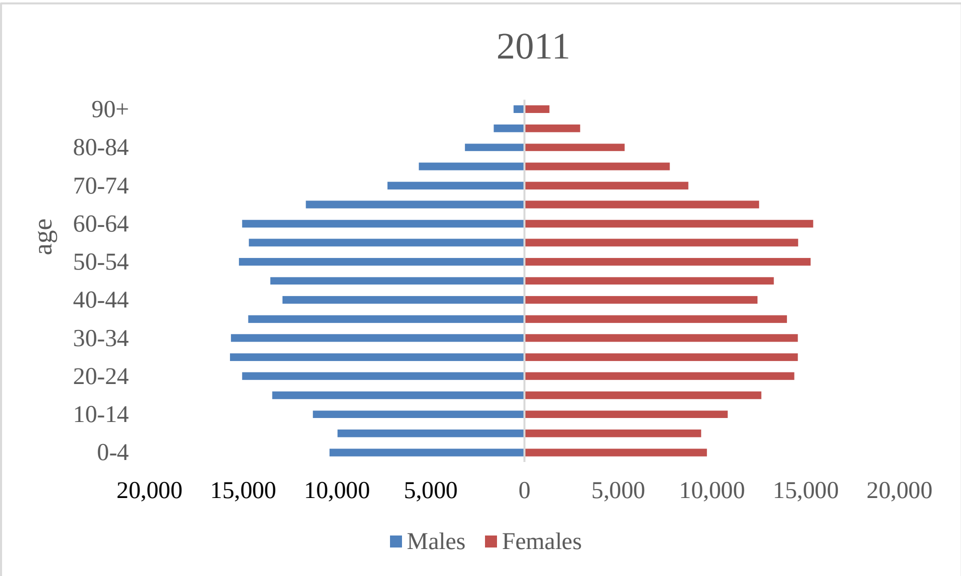


Figure 1.3: Population pyramid: 2011

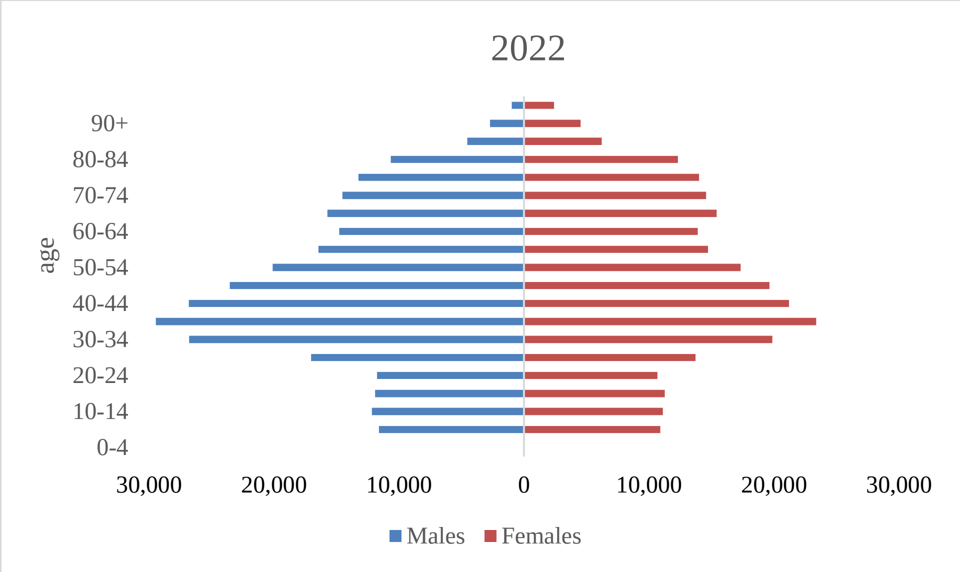


Figure 1.4: Population pyramid: 2022

In 2022, the total and youth unemployment rates for Malta were lower than the EU average. The total unemployment rate in 2022 for the EU 27 countries stood at 6.8% of the population aged 15-74, while for Malta it was 2.9%. The youth unemployment rate, measured as the proportion of unemployed among the population aged 15 to 29, stood at 11.3% within the EU and 5.9 in Malta (Eurostat, 2023f). The share of people at-risk-of-poverty was also below the average for the EU, at 20.1% compared to 21.6% (Eurostat, 2023e). Data from the Labour Force Survey (LFS) shows that during quarter four of 2022, 8.8% of the estimated gainfully occupied population were engaged in elementary occupations (National Statistics Office, 2023g). These statistics are not disaggregated by citizenship, however employment data published by the national employment agency, Jobsplus, show considerable differences in the type of occupations held by non-EU workers compared to the total gainfully occupied population. Of the non-EU citizens gainfully occupied at the end of 2022, 30% were engaged in an elementary occupation (Jobsplus, 2023). Based on the 2021 census, when excluding citizens of the United Kingdom, 59% of the foreign population were citizens of non-EU countries (National Statistics Office, 2023b). In, 2022, Malta had the eighth highest proportion of early-school leavers among those aged 18-24 within the EU. While the EU average stood at 9.6% in 2022, the proportion for Malta was 10.1%. The current target set for the EU is 9% by 2030 (Eurostat, 2023a). Results from the 2011 census show that 6% of the population aged 10 and over were illiterate and this increased to 14% in those aged 70 – 79 and 20.5% in those aged 80 – 89 (National Statistics Office, 2014b). As at end 2021, Malta also had the lowest total fertility rate in the EU at 1.13, while the EU average stood at 1.53 (Eurostat, 2023c). The current phenomenon of population aging is expected to become an even more pressing issue in the future if the current trend in fertility continues. The elderly are an important sub-group, not only because of lower educational attainment and illiteracy, but also because of restricted financial means as the economic sustainability of the present pension system becomes increasingly threatened (Musu, 2015). While these indicators provide only a snapshot of the current socioeconomic status of the country, the recent and rapid demographic shifts are expected to lead to considerable change in the socioeconomic characteristics of the islands in the future.

1.4 Measurement of deprivation in Malta

As part of the EU, Malta is obliged under the European Statistical System (ESS) to conduct the EU- Social Income Living Condition (SILC) survey. This cross-sectional survey collects data from private households on income, poverty, social exclusion and living conditions, and is used to produce indicators measuring material and social deprivation and risk of poverty or the at-risk-of-poverty rate (ARP). The survey was first launched in 2003 and is now an established annual mechanism within the EU to collect comparable data. The questionnaire also includes health related questions, specifically self-perceived health, and self-reported health limitations, as well as questions on unmet need for healthcare (European Commission, 2015)

The NSO in Malta has published indicators arising from the EU-SILC since 2005. The two main indicators related to deprivation and poverty arising from the survey are estimated based on thresholds. The ARP indicator sets the at-risk-of-poverty threshold at 60 per cent of the national equivalised income. Any person falling below this threshold is classified as at-risk-of-poverty, while anyone above the line would not be deemed at risk. In 2022, 16.7% of the population living in private households, equivalent to 85,797 persons, were deemed at-risk-of-poverty. The percentage of the population at-risk-of-poverty is highest among those aged 65 and over, at 33.3%, while there is also a clear geographical gradient with the Northern Harbour having the highest share of persons at-risk-of- poverty at 22.5% and the South-East having the lowest share at 8.5%. It must be noted that these estimates are not adjusted by the age and sex structure of the districts (National Statistics Office, 2020). The indicator on material and social deprivation is estimated based on 13 deprivation items. Seven of these relate to the household, and include, amongst others, a household's ability to afford a meal with meat/chicken/fish or vegetarian equivalent every second day; afford keeping their home appropriately warm; have access to a car/van for personal use; and face unexpected expenses. The remaining six relate to the individual and include, amongst others, whether persons in the household can replace worn-out clothes with new ones; participate in regular leisure activities; have an internet connection; and get together with friends/family for a drink/meal at least once a month. Persons lacking at least five items out of the 13 material and social deprivation items are considered materially and socially deprived, while those lacking at least seven items are considered severely materially and socially deprived. In 2022, 14.6% of the population experienced material deprivation with 4.9% being severely deprived. As for the at-risk-of-poverty rate, the percentage of those materially deprived was highest in those aged 65+ at 18.4%. This indicator was not disseminated by geographical location (National Statistics Office, 2023f).

The EU-SILC survey is now a long-standing tool for the measurement of income, social, and living conditions in Malta, and is used by many policy makers and researchers to assess the status of poverty and deprivation in Malta. While surely a valuable measure because of its long time series and comparability across the EU, it is not without limitations. Firstly, the indicators on poverty and deprivation focus mainly on monetary aspects of deprivation, either explicitly on income or ability to afford specific items, replace items or face unexpected expenses. Secondly, the indicators outputted from the survey are based on a threshold, persons above the line are deemed not to be deprived or at-risk-of-poverty. The lack of a ranked measure of deprivation does not allow for an understanding of a possible gradient. Finally, the EU-SILC is a sample survey, and may be impacted by sampling and non-sampling biases. Marginalised or hidden populations may be excluded from the sampling frame while non-response may be higher in persons who are more deprived (Goodman and Gatward, 2008). Estimates based on samples may also be limited by small counts. In fact, estimates for Malta from the EU-SILC are not disseminated if based on less than 20 reporting households or the non-response on an item exceeds 50%. Estimates based on 20 to 49 households or with a non-response of between 20 and 49% are flagged as to be used with caution. This may limit the survey from disseminating more detailed breakdowns for the main indicators, in fact estimates by locality and broad citizenship are not currently published (National Statistics Office, 2023f).

1.5 Socioeconomic determinants of health in Malta

The first research which attempted to outline health and social inequities in Malta was published in 1990. The article is descriptive and does not present any specific analysis of data, but merely aims to introduce the topic and highlight any sub-groups that may be of significance. The author suggests that the chronically unemployed, single mothers and the elderly warrant attention with respect to health inequity, however, Malta's small size contributes to a measure of social homogeneity, and regional disparities found in large countries may not be present. The author does however conclude that the monitoring of equity needs to be formalised in the country (Agius, 1990).

In 2014, the Office for the Commissioner of Mental Health, conducted a national survey on health literacy amongst the household population aged 18 and over, and attempted to analyse, amongst others, health literacy levels within different groups considered vulnerable. Findings from the study showed that at least 50% of persons within these groups were classified as having a health literacy that was inadequate/problematic. These included persons with very low or lower middle self-assessed social status, residents of Gozo, persons aged 76 years and over, widows, and persons who found it very difficult to pay for medication should it have been necessary (Office of the Commissioner for Mental Health, 2014).

A 2017 descriptive study aimed to specifically analyse the social determinants of health in Malta by analysing aggregate data available from surveys conducted amongst the population 16 and over, and mortality register data. The analysis used highest level of education achieved when analysing survey data and district of residence for the mortality statistics. This analysis found gradients in health outcomes for age-standardised mortality by district of residence, and self-perceived health, prevalence of certain chronic conditions and unmet need for health care by education level (Deguara, England and Azzopardi Muscat, 2017). While showing that gradients do exist, the analysis was only based on univariate relationships and did not adjust for competing risk factors in the survey data analysis. Another study, also published in 2017, attempted to analyse socioeconomic predictors of hypertension prevalence using data from a health examination survey conducted amongst the population aged 18 to 70. After adjusting for co-variates, the study found no relationship between education level or employment status, but did find spatial differences, with higher prevalence amongst persons living in the districts of Gozo, North Harbour and the West when compared to the Southern Harbour district

(Cuschieri *et al.*, 2017). In a paper published in 2022, the same researchers attempted to analyse spatial patterns in health inequalities using data from the same health examination survey, but this time focusing on prevalence of obesity and diabetes. The findings confirmed the results published in 2017, that prevalence of certain non-communicable conditions was related to district of residence (Cuschieri, Calleja and Mamo, 2022).

Since 2001, Malta has participated in the World Health Organization (WHO) Health Behaviour in School-Aged Children (HBSC) survey. The survey is conducted every five years and collects self-reported data on the physical, emotional, and psychological aspects of health, and the influences of the family, schools and peers on young people aged 11, 13 and 15 years. The survey is conducted within schools, in a classroom setting. Findings of the 2017/2018 report for Malta show higher daily consumption of sugared soft drinks, less time engaging in physical activity, a smaller share reporting excellent health, lower life satisfaction, higher reported multiple health complaints and a higher share reporting feeling low amongst boys and girls within families with lower affluence compared to those in families with higher affluence. The family affluence scale was used to measure affluence by asking the respondents to indicate the availability of six material assets in the household, such as a car, computer and dishwasher (World Health Organisation, 2020).

In 2022, the National Statistics Office (NSO) published its first analysis of social indicators combined with health indicators available from the EU-SILC survey. The analysis found that those who were at-risk-of-poverty or social exclusion were more likely to report fair/bad general health, chronic illness, being limited due to a health problem, and feeling downhearted and depressed for the years 2019, 2020 and 2021 (National Statistics Office, 2023e). The analysis was univariate, however, and the findings were not adjusted for competing risk factors.

In 2018, in response to the World Health Organisation (WHO) Review of Social Determinants and the Health Divide in the WHO European Region (Marmot *et al.*, 2012), the Ministry for Health established the Social Determinants of Health Unit within the Office of the Superintendence of Public Health. The project involves the collaboration of cross ministerial partners and civil society groups and was set-up through funding from the European Social Fund. The project launched a national survey in 2018, focusing specifically on health inequalities, however the publication of the findings of the project are still pending at this time (Social Determinantes of Health Unit, n.d.).

1.6 The census of population and housing in Malta

The decennial census of population and housing has a long-standing history in Malta. Since 1842, the census has been conducted approximately every ten years and aims to collect a stock of all persons, households, and dwellings. Since 1948, the census has been conducted within the provisions of a national legal framework – the Census Act – which governs the implementation of all aspects related to the census including the processing and publication of data. Since 2011, as a Member State of the EU, Malta also conducts its census within the parameters established by the EU regulation which outlines common rules for the implementation and provision of census data, every ten years, across the EU Member States (National Statistics Office, 2023a).

Malta has so far employed a traditional mode for data collection meaning that individuals are enumerated through direct collection of data provided by them through a survey questionnaire. Collection is conducted through the creation of enumeration areas and a team of enumerators is engaged to conduct canvassing and enumeration of dwellings. While some questions may vary between census rounds depending on national and international requirements, the questionnaire collects information on the demographic, social, economic, and household profile of the population, as well as information on dwellings. This thesis will make specific reference to the 2005 and 2011 rounds of the census conducted in Malta. More details will be provided in section 3.2.

1.7 Aims and objectives

While there has been an increased interest and recognition of the need for health inequalities research in Malta, most analysis has been restricted to cross-sectional surveys which only provide snap shots at intermittent time points and may be limited by sampling and non-sampling bias. These sample surveys may also be impacted by small counts due to sample sizes within a small population like Malta, therefore analysis of detailed patterns within the data may be restricted due to lower reliability of estimates. The use of surveys also limits the type of health outcomes that can be collected, for example, without linking survey data to health registers, surveys do not allow for the socioeconomic patterning of outcomes such as mortality or disease incidence. While Malta has several well-established country level health registries and administrative health sources, these tend to lack data on the more traditional socioeconomic indicators such as education, income, and employment status. While spatial analysis of health outcomes is possible when some location data is available, most analysis in this regard has been conducted using the established six geographical districts which were developed by the NSO. These were originally conceived as part of the spatial analysis of the 1967 census and therefore are not expected to consider the major demographic and social shifts that have occurred since then. Apart from that, they are now considered too large, covering large portions of the population and are more likely heterogeneous due to population growth and internal migration. This makes them very limited for spatial analysis in general and in health inequalities research, specifically, as they may not be able to clearly identify or explain differences in health outcomes experienced within the population living in different areas.

Within the context of a small country, where resources are limited and there is a push to reduce data collection burden from surveys, the routine and comprehensive study of health inequalities needs to look at alternatives to bridge information gaps. A small area-based deprivation index could serve as a valuable tool to bridge this gap and facilitate the regular analysis of health inequalities. This is because such an index provides the flexibility for researchers to use available health data, which is routinely updated, provided some information on location is stored within the register. The availability of the census allows for the development of a small area-based index that considers the entire population and can be updated with each new round in the census cycle. The census is a well-established exercise, which is legally mandated and conducted every ten years and captures information on the entire resident population through full enumeration.

This research therefore aims to develop and validate a small area-based deprivation index for Malta using census data and apply it to the analysis of health inequalities. The specific objectives are outlined below:

1. To develop an area-based deprivation index for Malta using census data.
2. To test the performance of the index in health inequalities analysis at two different spatial levels – enumeration area and locality.
3. To analyse the relationship between small area deprivation and all-cause mortality.
4. To analyse the relationship between small area deprivation and cancer incidence.
5. To analyse the relationship between individual level socioeconomic predictors, small area deprivation and selected health outcomes.

1.8 Overview of thesis

The remainder of this thesis is divided into six chapters. Chapter 2 provides an overview of the literature related to small area-based deprivation indices developed using census data. This chapter primarily focuses on the methods used by other researchers to develop, test, and validate their indices. This review served as the methodological basis on which the index in this research was developed as described in Chapter 3, therefore only a brief section has been included regarding the wider use of these indices in health inequalities research and analysis.

Chapter 3 provides a detailed explanation of the development and validation of a small area-based deprivation index for Malta using census data – this includes a description of the data sources and methods used as well as the results from the analysis. The Chapter concludes with a discussion of the results and main strengths and limitations.

Chapter 4 and Chapter 5 relate to the application of the indices described in Chapter 3 to health inequalities analysis using administrative data. Chapter 4 presents the application of the index to age-standardised mortality rates and is split into two parts. The first part presents a comparative analysis of the relationship between deprivation and mortality when using the index estimated at two different small area geographical levels – enumeration area and locality. The second part presents the application of the locality index to the analysis of all-cause mortality by deprivation, within the five years around the census. Chapter 4 than presents similar analysis but with a focus on cancer incidence. In this Chapter, analysis is presented using age-standardised incidence rates for the top six cancers amongst the adult population in Malta. In both Chapter 4 and Chapter 5, a summary measure of the linear association between deprivation and the age-standardised rates was estimated to quantify the relative socioeconomic gradient. Both Chapters conclude with a discussion of the results and main strengths and limitations.

Chapter 6 is the final analysis chapter and presents multi-level analysis of the association between individual and area measures of socioeconomic status on two selected self-reported health outcomes and mortality. Using multi-level logistic regression models, the analysis presented in this Chapter primarily focused on assessing whether the area-based measure of deprivation developed in this thesis still contributed to the analysis of health inequalities when individual level socioeconomic characteristics were also included. The Chapter concludes with a discussion of the results and main strengths and limitations.

Chapter 7 is the final chapter and presents a general discussion of the work presented in this thesis. Apart from discussing the main findings emerging from the analysis presented in the previous chapters, this Chapter discusses the main strengths and limitations as well as presents the implications of this thesis to policy, practice, and research.

Chapter 2: Review of census-based deprivation indices

2.1 Introduction

The main objective of this research was the development, validation, and application of a small area-based deprivation index for Malta using census data, with specific focus on health inequalities analysis. In line with this objective, this chapter presents a detailed review of the literature related to census-based deprivation indices which have been developed, validated, and applied by other researchers. Indices reviewed are those developed for stand-alone research or indices developed with the intention to be updated regularly, in some cases by statistical offices or other authorities, for public dissemination.

2.2 Search strategy and inclusion and exclusion criteria

The first literature search was conducted in May 2017 and updated in December 2020. The search was conducted in EBSCOhost, Embase, Web of Science and PubMed. No restriction on year of publication was set. Search tests were conducted to develop the most appropriate combination of spelling variations and synonyms for keywords to ensure an exhaustive search of the research within the field. Studies were initially screened based on title and abstract. The remaining articles were assessed for inclusion based on review of the full text. Some additional articles or reports were included which were not identified through the search, mainly grey literature, such as methodological reports published by national statistical authorities. Table 2.1 presents the search terms used and the inclusion and exclusion criteria.

Table 2.1: Search terms and inclusion and exclusion criteria for literature review

Search terms				
"deprivation"	AND	"area" OR "communit*" OR "municipalit*" OR "district" OR "neighborhood" OR "neighbourhood"	AND	"index" OR "indic*" OR "measure" OR "score"

Inclusion criteria

- Study described the development of a census small area-based deprivation index.
- Study described the analysis of health inequalities using a census small area-based deprivation index.

Exclusion criteria

- Article could not be found in English.
- Full article was not available.
- Index not created with census data, for example, sample surveys, or in full or in part, based on administrative data sources.
- Article referring to the creation of more than one index (multiple deprivation index) and not a single composite index of deprivation.
- Article referred to individual level measures of deprivation or single area item (not a composite index).
- Index is a replication of an already existing index with no country specific changes.

Figure 2.1 presents a flow diagram of the review process with the final number of articles included for review.

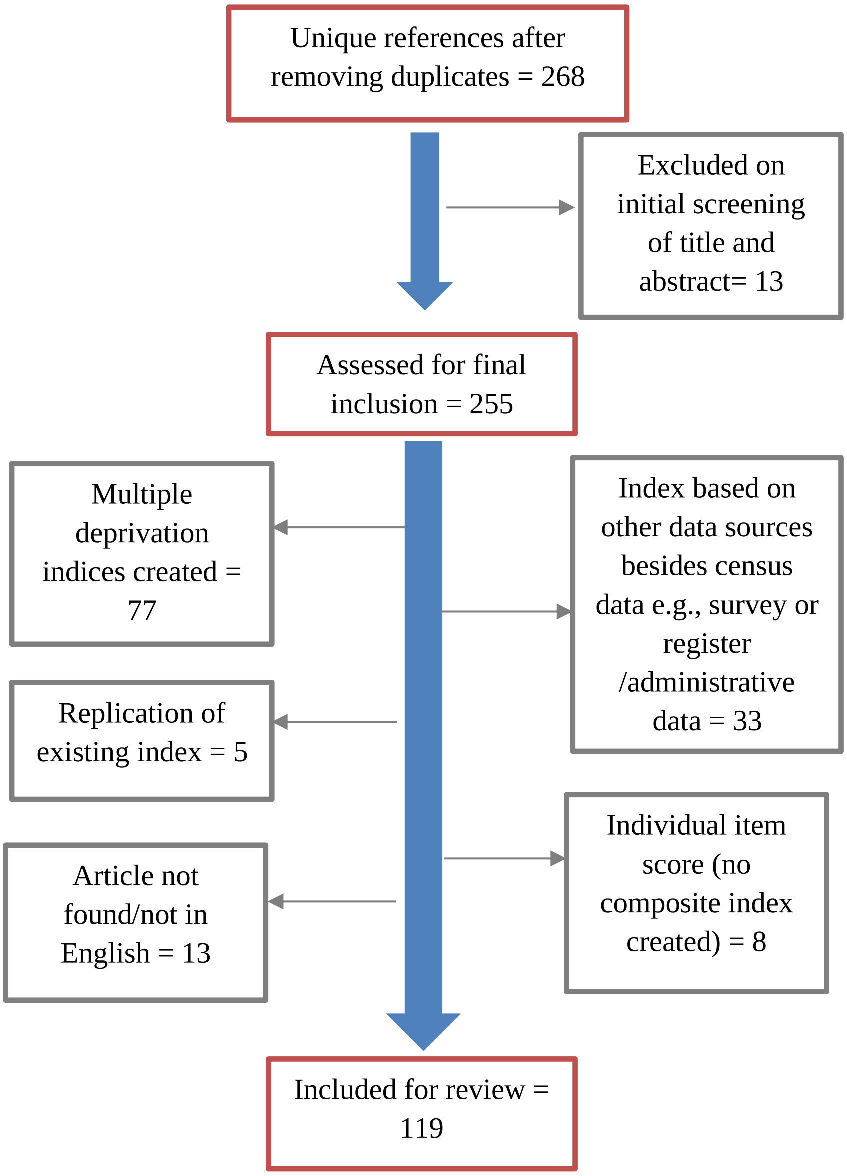


Figure 2.1: Flow diagram describing literature review selection process

The remainder of this chapter will be divided as follows: Section 2.3 will provide an overview of the census small area-based deprivation indices reviewed in this chapter; Section 2.4 will describe the steps and methods used to develop these indices; Section 2.5 will describe the validation of the indices, including predictive validity by testing for a health gradient; while section 2.6 will present a brief overview of the wider application of the indices to health inequalities analysis. Finally, section 2.8 will discuss and summarise this chapter.

2.3 Overview of the census small area-based deprivation indices

In total, twenty-six (26) unique indices, published between 1981 to 2020, were included for review (Table 2.2). While the initial index may have been updated since its original creation, for the purposes of this review, reference will be made to the first instance the index was developed and tested, where applicable. The indices were produced in nineteen (19) countries at national or regional level. Four (4) of the indices have been updated and published by national authorities, at least once – the Carstairs Score, updated and disseminated by the Medical Research Council/Chief Scientist Office Social and Public Health Unit, University of Glasgow; the Index of Relative Social Deprivation (IRSD) updated and disseminated by the Australian Bureau of Statistics; the New Zealand Index of Deprivation (NZDep) updated and disseminated by the University of Otago Department of Public Health; and the Townsend Index updated and disseminated by the Office of National Statistics, United Kingdom (UK).

Table 2.2: Summary of indices included in review

Index	Year[1]	Country (region)
Jarman underprivileged area index (UPA)	1981 (1971)	England and Wales
Townsend Index	1987 (1981)	England (Northeast)
Carstairs score	1991 (1981)	Scotland
Index of Relative Social Deprivation (IRSD)	1990 (1986)	Australia
Socioeconomic Factor Index (SEFI)	1996 (1986)	Canada (Manitoba)
New Zealand Index of Deprivation (NZDep)	1997 (1991)	New Zealand
Singh Area Deprivation Index	2003 (1990)	United States
Tello et al	2005 (1991)	Italy (Verona)
Messer et al	2006 (2000)	United States
Stimpson et al	2007 (1990)	United States
Winkleby et al	2007 (1995)	Sweden
Vancouver Area Neighbourhood Deprivation Index (VANDIX)	2007 (2001)	Canada (Vancouver)
Havard et al	2008 (1999)	France (Strasbourg)
Sanchez-Cantalejo et al	2008 (2001)	Spain
Lian et al	2011 (2000)	United States
Choi et al	2011 (2005)	South Korea (Busan)
Torres-Cintrón et al	2012 (2000)	Puerto Rico
Panczak et al	2012 (2000)	Switzerland
Rossen	2013 (2000)	United States
Bender et al	2015 (1999)	Denmark
Multi-criteria deprivation index for the city of Quito (MDIQ)	2015 (2010)	Ecuador (Quito)
Weng et al	2016 (2010)	China (Shenzen)
Palmetto Small-Area Deprivation Index (SADI)	2016 (2000)	United States (South Carolina)
Aungkulanon et al	2017 (2010)	Thailand
Powell-Wiley et al	2020 (2000)	United States
Wang et al	2020 (2010)	China

[1]Year refers to year of publication and (year) refers to the year of the census data used to create index

The rest of this chapter will summarise how the indices were developed, tested, and validated as well as their application in the measurement of health inequalities. Where a name for the index was created by the author, this will be used to reference the work.

2.4 Development of the index

2.4.1 Defining the small area

Defining the small area to be used to develop a deprivation index is a crucial step, as the underlying assumption when creating such an index is that the small area measure of deprivation can act as a proxy for individual level measures. This is one of the main criticisms of using such indices to measure health inequalities – the ecological bias or fallacy of assigning a group level score to an individual. Selection of the area used to assign a score therefore should ideally be a small geographical area that can be assumed to be homogenous with respect to socioeconomic status of the individuals living within it (Soobader *et al.*, 2001). Working with small geographic areas that are assumed to be homogenous should reduce the classification error at the individual level and thus limit ecological bias of analysis of individual outcomes when using area measures of deprivation. Apart from this, when selecting the small area, it is ideal that a link can be established between the geographical breakdown used to develop the index and outcomes intended for analysis by deprivation. If the small area is not easily linkable to other sources because of the small area selected, then this limits the wider use of the index. The wider usability of the index may not have been a concern for researchers who developed an index for standalone research, as opposed to researchers who developed a tool for wider use, such as one created by a national statistical institute.

Of the indices being reviewed, in many cases, the selection of the small area was a practical one, in that the area selected was the smallest area breakdown that was available for extraction of census data at the time of the index creation. Notwithstanding this, once an index is created, it can be produced at smaller or larger geographic areas if the index items can be extracted at this level from the census. Due to the decennial nature of the census exercise, advances may have been made in spatial analysis since the index creation, therefore the area available at the time of the creation of the index may have been updated to a smaller area or possibly a spatial grid, if census outputs are now being geo-coded. The indices considered in this review generally made use of small areas already implemented for spatial outputting of statistics, mainly census data or other official statistics, or areas defined by administrative boundaries.

Ten of the indices were developed and outputted using an established census small area. The number of the areas and size in terms of population counts varied considerably. Census tracts in the United States, on average, contain 4,000 residents (Messer *et al.*, 2006; Stimpson *et al.*, 2007; Lian *et al.*, 2011; Rossen, 2014). This size is similar to the census districts in Denmark (Bender *et al.*, 2015). On the other hand, census blocks in France contained on average 2,300 residents (Havard *et al.*, 2008), census dissemination areas in Canada contained approximately 400 to 700 residents (Bell and Hayes, 2012), and in Italy, the average size of census blocks was 162 (Tello *et al.*, 2005). Information on the average census block size was not provided for Ecuador (Cabrera-Barona *et al.*, 2015).

While these indices were created using the smallest output level available at the time to extract census data, the linking of the index to external data sources was sometimes complex. In four cases, linking of the index to external data was only possible after the involvement of the census office. This either involved a restricted access request for the linking of national survey data to the census tract data used to build the index (Stimpson *et al.*, 2007; Rossen, 2014; Powell-Wiley *et al.*, 2019), or assignment of an anonymised census block number by the census office to external data supplied by the researcher (Tello *et al.*, 2005). In two cases, linking was possible because researchers had individual residential address data available in the external data source, and this could be mapped to the census block or tract map which was publicly available (Havard *et al.*, 2008; Lian *et al.*, 2014). In Denmark, researchers with ethical approval to access administrative registers linked individual records to census data using a unique 10 digit number given to all Danish residents (Bender *et al.*, 2015). In Ecuador, outcome measures analysed using the deprivation index were based on questions in the census itself or estimated using spatial distances from a centroid for each census block, therefore the index was not linked to external data (Cabrera-Barona *et al.*, 2015). For the remaining two indices, no information was provided on how the final index was linked to the external data analysed by the researcher (Messer *et al.*, 2006; Bell *et al.*, 2007).

Several indices were developed using standardised small output areas that were developed externally to the census. This facilitated data linkage to other sources since these areas were not created just for use within the census, which would limit their utility as discussed above. These can be grouped into two categories, those based on postcodes or zip code areas, and small areas developed by statistical offices for spatial analysis. Post code sectors were used in the development of the Carstairs score in Scotland (Carstairs and Morris, 1991) and the Palmetto Small-Area Deprivation Index in the United States (Lopez-De Fede *et al.*, 2016). The post code sectors used in Scotland had an average population size of 4,200, while in the United States, the average size for the zip code tabulation areas (ZCTA) was 10,800. Indices developed in Australia (Australian Bureau of Statistics, 1998), New Zealand (Salmond, Crampton and Sutton, 1998) and Sweden (Winkleby, Sundquist and Cubbin, 2007) made use of small output areas developed by the national statistical authorities. In Australia, collection districts are the smallest area at which the index is available, these districts contain approximately 250 dwellings. While the index is available at this level, the authors advise that these scores are only used to estimate weighted aggregate values at higher levels due to possible instability of the index scores in areas with very small population sizes. In New Zealand, the index was constructed based on meshblocks which have a median population size of 90. In some cases, very small meshblocks were combined to cover, as far as possible, a population size of 100. For the index developed in Sweden, small area market statistics (SAMS) were used. These areas have been developed based on boundaries defined by homogenous types of buildings. The average size ranges from 1,000 to 2,000.

In some cases, while data from the census was available at a smaller census output level, such as census tract/block or enumeration area, the final index was produced using a larger spatial area. This decision was made so that analysis could be conducted on outcome data that was only available at the larger area level. Researchers who developed the Socioeconomic Factor Index (SEFI) obtained data at the census enumeration area level but aggregated this data to the municipality level reducing 1,825 enumeration areas into 255 municipalities for a population of approximately 1 million (Frohlich and Mustard, 1996). Singh developed an index at the census tract level but used tract scores to produce an average score at the county level. Initial analysis was conducted on 59,525 census tracts with an average size of 4,000, while the final index was produced for 3,097 counties by averaging out the census tract data to the higher level counties (Singh, 2003).

The use of administrative boundaries, while in most cases was the most practical option to facilitate analysis of external data, was also deemed appropriate because administrative areas are used by policy makers and service planners. For example, when developing the Townsend Index, while local authority ward was selected as the area unit of analysis because, at the time, health data in England and Wales could only be matched to census data at this level, the assigning of a deprivation score to wards was considered useful for policy makers and local planners. Townsend did acknowledge that using local administrative areas may have several disadvantages, including that boundaries for these areas may have been created arbitrarily and therefore may not distinguish homogenous neighbourhoods (Townsend, Phillimore and Beattie, 1988). Similarly, Singh noted, that while census tracts are possibly more homogenous, these are not stable, compared to counties within which health and social policies were developed and implemented, making the selection of counties, apart from practical for analysis, also more relevant for those using the index (Singh, 2003). Another disadvantage of using areas based on administrative boundaries, is that they may vary considerably in size since these geographical areas were not created to encompass a certain number of persons within them. In fact, at the time the Townsend Index was developed, wards ranged from a population of 500 to 15,000 (Townsend, Phillimore and Beattie, 1988). An index developed in South Korea made use of towns, the smallest local administrative unit, however the average population size for towns in the period the index was created was 84,066, ranging from 26,825 to 238,120 (Choi *et al.*, 2011). Researchers in Thailand tried to address the issue of varying sizes of administrative areas by combining districts into what they called "super districts", however this resulted in areas with a median size of 189,067, ranging from 100,970 to 492,490 (Aungkulanon *et al.*, 2017). Indices developed in Spain (Sánchez-Cantalejo, Ocana-Riola and Fernández-Ajuria, 2008), Puerto Rico (Torres-Cintrón *et al.*, 2012) and China (Weng *et al.*, 2016) were also developed using municipalities or districts, however no information on size was provided. The outcome data analysed by the researchers were available at this level, therefore it can be presumed that this geographical breakdown was selected based on this factor.

Finally, researchers developing an index in Switzerland created unique neighbourhoods specifically for the construction of the index. Having access to geographic coordinates for residential buildings based on data available from the statistical office, census data was geocoded, and neighbourhood boundaries were created based on road network connectivity and a target number of households within neighbourhoods. The average size of the neighbourhoods was 52.7 households, which was equal to an average of 120 individuals per area (Panczak *et al.*, 2012).

As can be seen, the concept of a "small area" is very relative in each context within which the indices were developed. In some cases, small areas referred to breakdowns including as little as 100 persons. In other cases, the small area available covered average population sizes in the tens of thousands. Very small output areas do also face some challenges due to small counts, as there may be confidentiality issues in accessing the data and issues with deprivation scores estimated based on extreme item values due to small population counts. In some cases, data suppression due to confidentiality required exclusion of small areas which contained a total population below a minimum threshold. For example, in France, census blocks with population sizes below 250 had to be excluded due to confidentiality restrictions (Havard *et al.*, 2008), in Australia areas with less than 10 people were excluded (Australian Bureau of Statistics, 1998), while researchers in Sweden excluded areas where there were fewer than 50 people when creating their index (Winkleby, Sundquist and Cubbin, 2007).

2.4.2 Defining the domains and items that measure deprivation

Selection of individual items that are intended to be used to create a composite index on deprivation should encompass the domains that are expected to measure deprivation, relative to the country or context within which the index is being developed. In turn, it is expected that these domains measuring deprivation would directly or indirectly impact health outcomes, if applied in health inequalities analysis (Allik *et al.*, 2020). Indeed, domains traditionally included in measures of deprivation cover a broad set of categories theoretically considered to encompass material and social deprivation, and many researchers have used the conceptual underpinnings of deprivation outlined by authors such as Townsend (Townsend, 1987), as well as what was applied in other indices, to base the selection of domains and items considered for their own index. Ultimately, however, the selection of domains and more importantly, items, is restricted by data availability within the census. While some questions are generally universal across censuses, such as those related to education and employment, other questions are collected by certain countries and not others, such as income.

Since indices considered here were those developed at the area level, items operationalised for potential inclusion in the index were related to the experience of individuals or households within the area of analysis. This was, for most of the items, the proportion of the population or households within the small area experiencing a specific condition. In a small number of cases, items which were potentially included were not estimated as a proportion, but as a mean or median value for the area. In general, the items were estimated to measure the share of the population within the category considered most disadvantaged for each item, but there were cases where items referred to the most advantaged category.

This section will summarise the domains and items considered by researchers as potential measures for deprivation when developing their indices, regardless of whether the items and domains were included in the final score. Item reduction and selection, where applicable, will be discussed in section 2.4.4. The next sections will discuss each domain separately.

2.4.2.1 Employment

All the twenty-six (26) indices reviewed in this chapter considered items related to the domain of employment for inclusion in their index. It must be noted that some researchers classified employment items under the domain of income, and this was mainly because occupation category was considered as a proxy measure for income when questions specifically related to income were not available. In this section, items related to occupation will be classified under employment and only items directly collecting information on income will be referenced within that domain.

The employment items included as measures of deprivation generally fall under two categories - employment status and occupation classification. By far the most common item was unemployment, with twenty-three (23) of the researchers including an item which measured unemployment for potential inclusion in their index. In nearly all cases this was estimated as a share of the working age population, either with a minimum age threshold such as 15 and over, or with both a minimum and maximum, such as 15 to 64. There was only one case where the researchers calculated unemployment as a proportion of the entire population in the small area and this was not adjusted to take into account possible differences in age structure across areas due to differing shares of the non-working population (Weng *et al.*, 2016). While the majority of the indices estimated the unemployment item for the total working population, there were instances where only male unemployment was taken into consideration (Carstairs and Morris, 1991; Choi *et al.*, 2011; Rossen, 2014). In a small number of cases, separate items were created for male and female unemployment (Australian Bureau of Statistics, 1998; Havard *et al.*, 2008).

Items related to occupation category were considered for inclusion by fifteen (15) of the researchers when creating their index. Occupation category for the employed population was considered as a proxy for social class, with most of the items estimating the proportion of the working population in the area engaged in occupations considered low skilled, elementary occupations or manual labour. In a small number of cases, the category used to calculate the item was the highest occupation classification, such as the proportion engaged in management or professional roles (Singh, 2003; Messer *et al.*, 2006; Stimpson *et al.*, 2007; Torres-Cintrón *et al.*, 2012).

Less common items considered were the proportion of those gainfully occupied who were self-employed (Havard *et al.*, 2008), proportion working part-time (Lopez-De Fede *et al.*, 2016), female labour force participation (Frohlich and Mustard, 1996; Bell *et al.*, 2007; Havard *et al.*, 2008), proportion working for no pay (Cabrera-Barona *et al.*, 2015) and the proportion who lost the ability to work due to illness, disability and other reasons (Z. Wang *et al.*, 2020).

Eleven (11) of the indices considered only one item related to employment, either measuring employment status or occupation category. The remainder considered more than one item from both categories of employment for possible inclusion in the index. In a small number of cases, redundancy was included with the same items created using different age cut offs to estimate unemployment, or different occupation categories for the items related to occupation. This was, in the most part, an intentional inclusion of redundancy, so that the different items could be tested to decide which was the best measure of deprivation for inclusion in the final index.

2.4.2.2 Education

Education was considered for inclusion by all but three of the index creators (Jarman, 1983; Townsend, Phillimore and Beattie, 1988; Carstairs and Morris, 1991). The items considered were generally measures of educational attainment and were constructed based on a threshold that was considered to discriminate low or high educational attainment. In most cases this was related to the level of education attained, certifications obtained, or duration of years in formal education. Items were either intended to estimate the share with low education attainment, such as the proportion of the population with no formal schooling, or high education attainment, such as the proportion of the population with a tertiary level degree. The threshold set varied across the research contexts, as would be expected, since the level of attainment within a country considered to be a measure of deprivation is relative to the average educational attainment within the country. Since level of educational attainment is ongoing for those still of school age, items related to education were restricted to the population above an age threshold generally corresponding to the age when compulsory schooling is complete. In one instance the education item had no explicit age threshold but referred to the level of education attained by the head of household which indirectly set an age limit (Panczak *et al.*, 2012). In only two cases were items related to education attainment estimated using the entire population in the denominator, however there was no adjustment conducted in relation to variation in age structure to take into account the differing proportion of children within areas (Tello *et al.*, 2005; Cabrera-Barona *et al.*, 2015; Weng *et al.*, 2016). Items were not restricted to a specific gender category, as was seen in some cases for the employment items but were measured among the total population within the restricted age category of interest.

Less common items related to education considered by some of the researchers were illiteracy (Sánchez-Cantalejo, Ocana-Riola and Fernández-Ajuria, 2008; Weng *et al.*, 2016; Aungkulanon *et al.*, 2017; E. Wang *et al.*, 2020), language proficiency (Australian Bureau of Statistics, 1998; Singh, 2003; Torres-Cintrón *et al.*, 2012) and early school leavers (Australian Bureau of Statistics, 1998; Lopez-De Fede *et al.*, 2016).

Nine of the indices considered just one item on education, while the rest considered more than one, in many cases, the items were similar conceptually, but used different thresholds to define low level of education or different age groups. This was, in the most part, an intentional inclusion of redundancy, so that the different items could be tested to decide which item best measured deprivation for inclusion in the final index.

2.4.2.3 Income

Items specifically related to income were considered by fourteen (14) of the indices. The majority of the items measured the proportion of persons experiencing a household or family income below a certain threshold considered to constitute a level of poverty in the reference year, though in a few instances the median value was used rather than a proportion (Singh, 2003; Bell *et al.*, 2007; Stimpson *et al.*, 2007; Havard *et al.*, 2008; Rossen, 2014; Powell-Wiley *et al.*, 2019). As opposed to the items on education and employment, since household income was considered, the items were measured as a proportion of the total population in the area with no restriction on age group or gender.

Apart from items directly measuring household income, another common item related to this domain was the proportion of persons receiving social benefits (Salmond, Crampton and Sutton, 1998; Messer *et al.*, 2006; Winkleby, Sundquist and Cubbin, 2007; Lian *et al.*, 2011; Rossen, 2014; Powell-Wiley *et al.*, 2019). Five of the indices considered just one item on income, while the rest of the fourteen included more than one item from the income domain for consideration.

2.4.2.4 Living conditions

Items related to living conditions were considered by all but four of the indices (Winkleby, Sundquist and Cubbin, 2007; Sánchez-Cantalejo, Ocana-Riola and Fernández-Ajuria, 2008; Rossen, 2014; Bender *et al.*, 2015). Items generally focused on the proportion of persons with or without access to specific items, with no access to a car, being the most common item (Townsend, Phillimore and Beattie, 1988; Carstairs and Morris, 1991; Australian Bureau of Statistics, 1998; Salmond, Crampton and Sutton, 1998; Singh, 2003; Havard *et al.*, 2008; Choi *et al.*, 2011; Lian *et al.*, 2011; Torres-Cintrón *et al.*, 2012; Lopez-De Fede *et al.*, 2016; Aungkulanon *et al.*, 2017). Less common items were no access to the internet (Australian Bureau of Statistics, 1998), a telephone (Singh, 2003; Stimpson *et al.*, 2007; Torres-Cintrón *et al.*, 2012; Aungkulanon *et al.*, 2017; Powell-Wiley *et al.*, 2019) and basic facilities such as running water, connection to a sewage system, kitchen facilities, toilets and refrigerators (Singh, 2003; Stimpson *et al.*, 2007; Cabrera-Barona *et al.*, 2015; Weng *et al.*, 2016; Aungkulanon *et al.*, 2017; Z. Wang *et al.*, 2020).

Items related to over-crowding were also very common (Jarman, 1983; Townsend, Phillimore and Beattie, 1988; Carstairs and Morris, 1991; Australian Bureau of Statistics, 1998; Salmond, Crampton and Sutton, 1998; Singh, 2003; Messer *et al.*, 2006; Havard *et al.*, 2008; Panczak *et al.*, 2012; Cabrera-Barona *et al.*, 2015; Aungkulanon *et al.*, 2017). Less common items related to the proportion of persons living in dwellings that were rented (Salmond, Crampton and Sutton, 1998; Singh, 2003; Tello *et al.*, 2005; Bell *et al.*, 2007; Havard *et al.*, 2008; Choi *et al.*, 2011; Aungkulanon *et al.*, 2017) and the mean/median value of the dwelling or mean/median amount of rent paid (Australian Bureau of Statistics, 1998; Singh, 2003; Stimpson *et al.*, 2007; Panczak *et al.*, 2012; Powell-Wiley *et al.*, 2019).

As with the items related to income, these were measured as a proportion of the total population in the area and no age or gender restrictions were conducted. Seven of the indices considered just one item related to living conditions, while the rest considered more than one item.

2.4.2.5 Household type

Sixteen (16) of the indices considered items related to household types. These items have generally focused on specific types of households such as single parent families (Jarman, 1983; Frohlich and Mustard, 1996; Australian Bureau of Statistics, 1998; Salmond, Crampton and Sutton, 1998; Singh, 2003; Tello *et al.*, 2005; Messer *et al.*, 2006; Havard *et al.*, 2008; Choi *et al.*, 2011; Lian *et al.*, 2011; Rossen, 2014; Powell-Wiley *et al.*, 2020), proportion of persons who are separated, divorced or widowed (Australian Bureau of Statistics, 1998; Salmond, Crampton and Sutton, 1998; Tello *et al.*, 2005; Choi *et al.*, 2011; Aungkulanon *et al.*, 2017) and the proportion of certain population groups living alone (Jarman, 1983; Choi *et al.*, 2011; Weng *et al.*, 2016; Aungkulanon *et al.*, 2017). Despite the popularity of items of household type in deprivation indices, their inclusion has been criticized, since it is assumed that it is the measurement of conditions within an area and not the kinds of people in the area that make the area deprived (Carstairs and Morris, 1991).

2.4.2.6 Other domains

Other domains which have been less commonly included as measures of deprivation relate to items which measure demographic features of the population, such as the proportion of children and the elderly (Jarman, 1983; Choi *et al.*, 2011), items related to migration and ethnicity (Jarman, 1983; Havard *et al.*, 2008; Lian *et al.*, 2011; Aungkulanon *et al.*, 2017), items related to urban/rural differences (Aungkulanon *et al.*, 2017; Z. Wang *et al.*, 2020) and items related to health (Australian Bureau of Statistics, 1998; Cabrera-Barona *et al.*, 2015). As with household type, the inclusion of items related to demographic and migration characteristics of the population has been criticised. The inclusion of items related to urban/rural differences and health as a component of deprivation restricts the use of such an index to measure differences in deprivation for outcomes related to these items as they would have been partially adjusted for in the deprivation index itself.

2.4.3 Data treatment

While the items discussed in the previous section were estimated as either proportions of the population/households within the small area or mean/median values, depending on the item being considered, in most cases, these raw estimates were not used when running selection techniques to be discussed in section 2.4.4 and the eventual computation of the final scores discussed in section 2.4.5. Nearly all researchers applied some form of data treatment on the initial item values prior to their use, this mainly related to standardisation. Only seven (7) of the indices did not refer to any form of data standardisation prior to analysis or computation (Singh, 2003; Tello *et al.*, 2005; Messer *et al.*, 2006; Stimpson *et al.*, 2007; Panczak *et al.*, 2012; Bender *et al.*, 2015; Cabrera-Barona *et al.*, 2015). The reason that standardised scores were used was to ensure that variables were on the same scale, with similar means and standard deviations. This ensured that items with higher proportions were not given more weight in the analysis or index computation.

The standardisation of the raw items into Z-scores was by far the most common method applied. In sixteen (16) of the indices this method of standardisation was used. Z-score standardisation is conducted by subtracting the mean from each value of an item and dividing it by the standard deviation. The resulting Z-standardised values have a mean of 0 and can be positive or negative. In a few cases some or all of the items were normalised apart from being Z-standardised (Townsend, Phillimore and Beattie, 1988; Frohlich and Mustard, 1996; Rossen, 2014; Weng *et al.*, 2016), while in one (1) case only normalisation was conducted (Sánchez-Cantalejo, Ocana-Riola and Fernández-Ajuria, 2008). Normalisation of items intended to measure deprivation does however seem counter-intuitive, since normalising the data would remove the impact of outliers, and it is likely that such outliers would differentiate areas that are most or least deprived. In one index, indirect age standardisation was conducted on the items that were restricted to specific age groups and gender, while two items related to households were adjusted by household size and composition using a standard equivalence scale (Salmond, Crampton and Sutton, 1998).

Apart from standardisation of the items for scaling purposes, in a few cases, reverse coding was conducted to ensure that raw items that measured higher advantage in areas were reversed (Bell *et al.*, 2007; Torres-Cintrón *et al.*, 2012; Powell-Wiley *et al.*, 2020). In most cases, the items selected for possible inclusion in the indices were intended to reflect higher disadvantage, therefore including a mix of items measuring higher disadvantage or advantage in the same index would distort the deprivation score. Reverse coding such items would ensure that when combined, higher values would refer to higher disadvantage and thus positive and negative items would not cancel each other out.

2.4.4 Selection of items

Section 2.4.2 outlined the domains and items that researchers have considered as measures of deprivation when creating their indices. In some cases, researchers have only developed a list of items and used those for their index, while others have developed an initial list and applied a selection process to produce a final list of items to be included in their index. The methods used to select the final items can be grouped into three categories – selection of items based on a theoretical framework usually based on the literature and other indices created previously, selection based on expert opinion collected through a consultation process, and selection through factor reduction and other data analysis techniques. It must be noted that in some instances factor analysis was used by researchers to construct the index but not as a data reduction technique. These cases will be discussed in section 2.4.5 in relation to computation of the final score. One must acknowledge that even if researchers applied a technique to select items from a predetermined list, that list was developed based on existing knowledge or research, therefore in all indices there is always an aspect of the item selection that is being based on some preconceived theoretical or conceptual understanding of what defines deprivation. In three of the indices, it was not clear how the final items were selected, therefore they are being excluded from this section (Stimpson *et al.*, 2007; Rossen, 2014; Bender *et al.*, 2015).

2.4.4.1 Theoretical framework

In total, seven (7) of the indices were developed using a list of items selected based on a theoretical framework of deprivation driven by the literature and previous studies. The Townsend Index was developed based on the conceptual framework developed by Townsend himself and the items selected were based on their perceived representation of material deprivation (Townsend, Phillimore and Beattie, 1988). A recent index developed in China also relied on the conceptual framework developed by Townsend to select the items for their index (Z. Wang *et al.*, 2020).

In some cases while a general review of the literature was conducted to define the items, there was an attempt to include national or regional literature so as to select items relative to the context within which the index was being created and which were shown to be associated to deprivation (Salmond, Crampton and Sutton, 1998; Tello *et al.*, 2005). In some cases, however, the context of research used was not specified therefore it was not clear if any of the research was specific to the country in which the index was being developed (Panczak *et al.*, 2012; Aungkulanon *et al.*, 2017). In the case of the Carstairs score, items were included in the index based on previous work which referred to a list used by a public department to identify areas for priority treatment with respect to planning (Carstairs and Morris, 1991).

2.4.4.2 Expert opinion

Expert opinion was used in the selection of items for four (4) of the indices. The first index to use this approach was the Jarman Underprivileged Areas (UPA). This index was developed specifically to identify underprivileged areas for the purposes of health planning, therefore the experts selected were general practitioners who were invited to respond to a survey. In this survey, the general practitioners were asked to assign a score to each item depending on how much they felt that item increased demand on their services (Jarman, 1983). The Vancouver Area Neighbourhood Deprivation Index (VANDIX) applied a similar approach. A survey was circulated amongst medical health officers in British Columbia and each of the respondents had to comment on 21 indicators which they believed contributed to poor health outcomes in the province. The researchers eventually only selected items which were rated as strongly agree and agree by the majority of the respondents (Bell *et al.*, 2007). While in the case of the Jarman UPA, 1,802 questionnaires were analysed, for the VANDIX, the selection of items was based on only 10 questionnaires out of 27 circulated.

Weng used a combination of data analysis as well as expert opinions to reduce the items from an original list of 35 indicators. First, Pearson's correlation was conducted to assess redundancy. Items that were highly correlated (correlation coefficient >=0.75) to each other were discarded. The remaining items were than reviewed by a panel of 20 experts who had to score the items on a scale of fitness in measuring deprivation. The panel of experts consisted of five sociologists, five economists, five social medicine scientists, and five social geographers (Weng *et al.*, 2016). A similar approach was implemented by Cabrera-Barona. Variances Inflation Factors (VIF) were calculated for 12 items to identify multi-collinearities between them. The items were then provided to a panel of 32 experts who were members of public and private institutions working in the fields of medicine, geography and territorial planning, environmental and social sciences (Cabrera-Barona *et al.*, 2015)

The main criticism of the use of expert opinion in selecting items has been that their opinions may be biased, especially if the number of responses is low and the pool of experts is restricted to a specific group of individuals, as was the case for both the creation of the Jarman UPA and the VANDIX.

2.4.4.3 Factor reduction and other data analysis techniques

Factor analysis as a form of data reduction was the most common analytical technique applied, and was used in the creation of nine (9) of the indices (Australian Bureau of Statistics, 1998; Singh, 2003; Messer *et al.*, 2006; Havard *et al.*, 2008; Sánchez-Cantalejo, Ocana-Riola and Fernández-Ajuria, 2008; Choi *et al.*, 2011; Lian *et al.*, 2011; Torres-Cintrón *et al.*, 2012; Powell-Wiley *et al.*, 2019). The process included an initial review of the literature to identify a list of items, which varied from as much as 52 (Havard *et al.*, 2008) to 14 items (Torres-Cintrón *et al.*, 2012). The aim of the analysis was to extract the items that most strongly correlated with the first component, or principal component, based on factor loadings. The items loading on the principal component were assumed to represent those measuring deprivation in the data being analysed. This reduction technique was generally conducted through an iterative process with a minimum loading score set onto which items would be retained. The threshold set by the researchers varied between 0.25 - >0.5. Factor analysis provides an estimate of the variance explained by the emerging components. The first principal component accounts for the largest proportion of variance in the original dataset, with each subsequent component explaining less of the variance. The variances of the principal components used to produce the indices range from below 50% (Salmond, Crampton and Sutton, 1998; Lian et al., 2011), to between 50 and 70% (Messer et al., 2006; Havard et al., 2008; Panczak et al., 2012; Torres-Cintrón et al., 2012).

Item selection was conducted for two (2) of the indices based on the individual correlation of the items to selected health outcomes (Frohlich and Mustard, 1996; Lopez-De Fede *et al.*, 2016). In the case of the Socioeconomic Factor Index (SEFI), individual items were first correlated to a prototype health index created by the researchers. Items that had a correlation value of 0.10 or greater were then included in a stepwise multilinear regression analysis with the same health index as the outcome measure. The items that were significant in the regression analysis were retained (Frohlich and Mustard, 1996). The Palmetto Small-Area Deprivation index applied a similar approach. Correlations were estimated between the individual items and five estimates of prevalence of chronic conditions. The mean correlation was estimated across the five conditions and the item with the highest mean from each domain was selected. Additional items were only included if the domain was not already represented in the items included (Lopez-De Fede *et al.*, 2016).

2.4.5 Computation of the final deprivation score

The number of items that made up the final indices varied considerably, the average number of items across the indices was 8, however this ranged from 3 to 19. Refer to appendix A for a detailed list of all items included in each index. The domains of employment, education and living conditions were the most common domains included across all the final indices.

For nearly all the indices, the deprivation score was computed through an additive procedure, with the only difference across indices being whether the items were weighted by some factor prior to the addition or left unweighted. Fourteen (14) of the indices weighted the individual items. In nine (9) of these cases the weights used emerged from a PCA analysis of the items, with factor loadings for each item being used as the weight (Australian Bureau of Statistics, 1998; Salmond, Crampton and Sutton, 1998; Singh, 2003; Messer *et al.*, 2006; Winkleby, Sundquist and Cubbin, 2007; Havard *et al.*, 2008; Lian *et al.*, 2011; Torres-Cintrón *et al.*, 2012; Z. Wang *et al.*, 2020). In one case, regression coefficients emerging from a regression analysis used to select the items were used as weights (Frohlich and Mustard, 1996). For the indices in which items were selected through expert opinion, the response levels assigned by the respondents for each item were used to estimate weights assigned to the items before summation (Jarman, 1983; Bell *et al.*, 2007; Cabrera-Barona *et al.*, 2015; Weng *et al.*, 2016). For seven (7) of the indices, unweighted summation of the items was conducted. In one instance, the individual items were first ranked and grouped into quartiles and the quartiles were summed rather than the items themselves (Bender *et al.*, 2015). Less common methods of construction of the final deprivation score were using the component matrix from the PCA and summing the factor loadings themselves instead of the items (Tello *et al.*, 2005; Sánchez-Cantalejo, Ocana-Riola and Fernández-Ajuria, 2008; Panczak *et al.*, 2012; Aungkulanon *et al.*, 2017) and estimating the mean value of the items within an area (Rossen, 2014).

Once the deprivation score was calculated, a few researchers applied standardisation methods to the score (Singh, 2003; Messer *et al.*, 2006; Cabrera-Barona *et al.*, 2015).

Categorisation of the final score was more common, though the categories varied across indices with some using percentiles (Tello *et al.*, 2005), quartiles (Messer *et al.*, 2006; Lian *et al.*, 2011; Bender *et al.*, 2015), quintiles (Australian Bureau of Statistics, 1998; Havard *et al.*, 2008; Choi *et al.*, 2011; Torres-Cintrón *et al.*, 2012; Aungkulanon *et al.*, 2017) and deciles (Z. Wang *et al.*, 2020). In one case, the score was categorised based on a number of standard deviations from the mean (Winkleby, Sundquist and Cubbin, 2007).

2.4.6 Sensitivity analysis

Sensitivity analysis was rarely conducted by the researchers when developing their indices. In many cases, the sensitivity analysis addressed the possible impact of varying area size on the index outputs. The developers of the New Zealand Index of Deprivation (NZDep) produced an alternative index based on regrouping of the small areas into larger areas including at least 200 persons instead of 100. The results of this analysis found that the weights using areas of 200 persons were very similar to those using 100 persons, so the alternative index was not retained (Salmond, Crampton and Sutton, 1998). Similarly, in Switzerland, the index based on neighbourhoods was compared to analysis based on larger areas (Panczak *et al.*, 2012). Singh tested the impact of using three different geographical levels on the factor loadings emerging from the PCA of the 17 items included in his index. This was conducted at the census tract, zip code and county levels and factor loadings were generally similar, with the variance and reliability coefficients almost identical (Singh, 2003).

In two indices, sensitivity analysis focused on the impact of excluding areas because of the impact of their size on the final index. Tello assessed the stability of the index by excluding census blocks of fewer than 30 inhabitants. Results when excluding these blocks were only marginally modified so they were retained as is and not aggregated to larger areas (Tello *et al.*, 2005). Wang reproduced the index excluding counties at the two extremes of the population range, therefore excluding both the smallest and largest areas. The index using all counties was correlated with the index using the restricted number of counties. The correlation coefficient was 0.999 (Z. Wang *et al.*, 2020).

Sensitivity analysis was conducted for two indices to assess the impact of varying the weighting method used to construct the index. Weng tested the consistency of the index when using four different methods to assign weights and plotted the analysis using the four indices against each other using a binary regression. The analysis found that the different weighting methods presented identical patterns of area deprivation (Weng *et al.*, 2016). Wang, on the other hand compared the weighted final index against an index constructed with no weights. The correlation of the indices yielded a corelation coefficient of 0.989 (Z. Wang *et al.*, 2020). Finally, sensitivity analysis was conducted by the NZDep to assess the impact of using unstandardised items to construct the index. The difference in the weights emerging from the analysis when using the unstandardised values was slight and the indices were highly correlated (r=0.997), however the standardised values were maintained since some very deprived small areas were affected non-negligibly (Salmond, Crampton and Sutton, 1998).

2.5 Validation

There are generally no standard methods available to validate small area-based deprivation indices. One of the strongest measures of validity of a deprivation measure is its ability to describe differences in health outcomes by deprivation, since it is expected that health outcomes behave along a deprivation gradient (Pampalon *et al.*, 2014; Allik *et al.*, 2020). Indeed, apart from this predictive validity, which will be discussed later in this section, other forms of validation were not commonly applied by the researchers when developing their index.

Internal consistency measured through the Cronbach's alpha, is a common tool to measure the reliability of a composite measure. Measured on a scale of 0 to 1, it quantifies the level of agreement between the items and is said to determine whether they collectively measure the same characteristic. The closer the value is to 1, the higher the internal consistency. In three of the indices, the Cronbach's alpha was reported as a measure of validity when creating the index. In all three cases the measure was above 0.9, indicating high internal consistency (Singh, 2003; Havard *et al.*, 2008; Lian *et al.*, 2011). Similar to the internal consistency measured through the Cronbach's alpha, researchers who developed the VANDIX estimated the weighted Kappa to quantify the level of discernment between respondents to their survey. The Kappa ranges from 0 to 1, with values closer to 1 indicating strong agreement. The weighted Kappa score was 0.16 suggesting a low level of consistency between the respondents (Bell *et al.*, 2007).

In a small number of cases, validation of the index was conducted by comparing the index with other external indicators that are believed to measure deprivation. Wang compared the deprivation index developed in their research against the official list of poverty-stricken counties and found a correspondence between the classification of areas of 70.6% (Z. Wang *et al.*, 2020). Panczak used data on income and financial deprivation as measured through the Swiss Household Panel Survey to estimate the equivalised household income for approximately 3,700 households and compared this estimate to the index decile assigned to the households. Results showed that mean household income increased with every unit increase in advantaged socioeconomic position based on the index (Panczak *et al.*, 2012). For two indices, validity was tested by recreating the index scores using the Townsend and Carstairs index with the census data used to create their index and comparing the results. Havard produced both the Townsend and Carstairs indices and correlated these with the index created in their research, yielding high correlation coefficients for both (Townsend, 0.97; $p<0.0$, Carstairs, 0.96; $p<0.01$)(Havard *et al.*, 2008). On the other hand Lopez-De Fede recreated the Townsend index using the census data used to create their index and compared the predictive validity of both indices (Lopez-De Fede *et al.*, 2016).

Finally, face validity, while probably the simplest form of validity, was only mentioned in one index. This may be because it is the most subjective and not easily measured. In the case of the Index of Relative Social Deprivation (IRSD), the researchers acknowledged the use of face validity based on local subjective knowledge. In each state, several of the small areas were selected to cover the range of deprivation values and they were independently assessed and graded based on local knowledge of the areas (Australian Bureau of Statistics, 1998).

Predictive validity through the application of the index to health outcomes was conducted for all but four of the indices (Jarman, 1983; Australian Bureau of Statistics, 1998; Sánchez-Cantalejo, Ocana-Riola and Fernández-Ajuria, 2008; Z. Wang *et al.*, 2020). The large share of predictive validity using health outcomes was in part since many of the indices were developed specifically for the patterning of health outcomes by deprivation. The types of outcomes varied between self-reported survey data and administrative data. While in most cases the analysis looked at the relationship between area deprivation and the health outcomes only, some of the analysis used multi-level models to combine individual predictors with the area deprivation measures constructed for the analysis.

By far the most common health outcome used to assess the predictive validity of the index was mortality. Standardised mortality rates were the preferred outcome measure. Analysis was either related to all cause or cause specific mortality and in some cases addressed specific age groups to focus on differences in premature and old age mortality (Townsend, Phillimore and Beattie, 1988; Carstairs and Morris, 1991; Frohlich and Mustard, 1996; Salmond, Crampton and Sutton, 1998; Singh, 2003; Choi *et al.*, 2011; Panczak *et al.*, 2012; Torres-Cintrón *et al.*, 2012; Aungkulanon *et al.*, 2017). Less common outcomes analysed were related to hospital service use based on admission and discharge data (Frohlich and Mustard, 1996; Salmond, Crampton and Sutton, 1998; Tello *et al.*, 2005), disease specific incidence and attack rates (Salmond, Crampton and Sutton, 1998; Winkleby, Sundquist and Cubbin, 2007; Havard *et al.*, 2008; Weng *et al.*, 2016), birth outcomes (Townsend, Phillimore and Beattie, 1988; Messer *et al.,* 2006) and self-reported health status. For three of the indices, analysis was conducted using multi-level models to assess the impact of area and individual level measures of socioeconomic status on colorectal cancer survival (Lian *et al.*, 2011), prevalence of obesity in children and adolescents (Rossen, 2014) and participation in population based screening (Bender *et al.*, 2015).

2.6 Applications in health inequalities analysis

Apart from the application to the analysis of health inequalities discussed as part of the predictive validity of the indices, several of the small area-based deprivation indices discussed here have been widely used for the analysis of a variety of health outcomes measured through surveys and administrative data sources. Analysis has been conducted to assess both the impact of area deprivation itself on the outcomes of interest, as well as through multi-level models combined with individual measures of socioeconomic status. In some cases, the indices presented have been replicated in other countries for area deprivation analysis. This is especially for the more long-standing and well-known indices such as the Carstairs score and Townsend Index.

The Carstairs score has been applied to the analysis of the impact of area deprivation on a variety of health outcomes in the United Kingdom, since it was developed. These include, but are not limited to, health related quality of life among older people measured through interviews (Breeze *et al.*, 2005), individual-level biomarkers (Chaparro *et al.*, 2018), frequency of musculoskeletal pain measured through a population survey (Neasham *et al.*, 2001), still birth and neonatal mortality due to congenital anomalies (Urwin *et al.*, 1998), prevalence of type 2 diabetes, insulin resistance and fasting blood glucose (Andersen *et al.*, 2008) and pre-term birth (Gray *et al.*, 2008). Apart from these examples, the Carstairs score has been replicated in other countries for use in the analysis of area deprivation on health inequalities (Park *et al.,* 2010). The Townsend Index has also been used in health inequalities analysis for a variety of health outcomes including coronary heart disease mortality (Janghorbani, Jones and Nelder, 2006), incidence of meningococcal disease (Fone *et al.*, 2003), hospital admissions (Luben *et al.*, 2019) and the effect of the provision of high-fluoride toothpaste on dental caries in children (Ellwood *et al.*, 2004). Similar to the Carstairs score, the Townsend Index has been adopted by other researchers in different countries for use in health inequalities analysis (Jeong *et al.,* 2006; Cubbin *et al.*, 2008; Young *et al.*, 2012)

In the United States, the Singh Area Deprivation Index has been used widely in the analysis of health inequalities since it was developed. These include, but are not limited to, the relationship between area deprivation and Medicare reimbursement for inpatient intra-coronary stent discharges (Tuliani *et al.*, 2017), survival in children with acute lymphoblastic leukaemia (Schraw *et al.*, 2020), ethnic disparities in hospital admissions from COVID-19 (Ingraham *et al.*, 2020), opioid use and drug poisoning mortality (Kurani *et al.*, 2020) and discharge destination after elective hip replacement (Mehta *et al.*, 2020). Similarly in Canada, the VANDIX has been used for the patterning of health outcomes by deprivation (Walker, Schuurman and Hameed, 2014; Amram *et al.*, 2015; Auluck *et al.*, 2016).

In New Zealand the NZDep, which is now regularly updated each census round, has been adopted as a measure of deprivation for regular health inequalities monitoring. These include, analysis of life expectancy (Tobias and Cheung, 2003), risks associated with poor water quality (Hales *et al.*, 2003), measured waist to body ratio in children (Egli *et al.*, 2020) and risk of diabetes complications in young people (Wijayaratna *et al.*, 2021). Similarly, in Australia, where the IRSD is regularly updated and disseminated by the national statistical authorities, there are a number of examples of the application of the index in health inequalities analysis (Cass *et al.*, 2001; Amir and Donath, 2008; Adams *et al.*, 2009; Stavrou *et al.*, 2009; Butler *et al.*, 2010).

2.7 Discussion

This chapter has presented a detailed review of the literature related to small area-based census deprivation indices, with specific focus on steps taken to develop the indices and their application in health inequalities analysis. These indices are not new, with the first indices being developed as far back as the 1980s. The methods to develop such indices are generally quite established. Defining the small area to be used to develop a deprivation index is a crucial step, as the underlying assumption when creating such an index is that the small area-based measure of deprivation can act as a proxy for individual level measures. In most cases this has been a practical decision as researchers have been restricted by the spatial level available for the dissemination of census data.

Defining the items and domains that measure deprivation are firstly based on a conceptual framework and previous research. When building a small area-based deprivation index using the census, researchers are also restricted by what questions are available in the census. The most common domains used to define deprivation were employment, education and living conditions. While some researchers included items in their index only based on a theoretical framework, others reduced an initial list of items based on either expert opinion or an analytical procedure. Computation of final deprivation scores was generally done through the addition of the items, which were more often standardised prior to the computation. While summation was the most common method to compute the final score, the differential weighting of the individual items was not always applied by researchers, with some assuming all items had the same weight.

Sensitivity analysis of the final deprivation index was rarely conducted by the researchers, with the most common type of analysis conducted done to test the impact of using different area sizes on the index scores There are generally no standard methods available to validate small area-based deprivation indices and apart from predictive validity of the index, other forms of validation were not very common. These were generally restricted to an analysis of internal consistency of the final items using Cronbach's alpha, validation of the index scores compared to other external data that were assumed to be measures of deprivation and face validity based on subjective knowledge. The most common health outcomes used to assess the predictive validity of the indices were mortality estimates, hospital service use, disease incidence and attack rates, and birth outcomes. Finally, apart from the examples presented related to predictive validity, there are many examples of the wide use of census small area-based deprivation indices in health inequalities analysis

Chapter 3: Development and validation of a small area-based deprivation index for Malta

3.1 Introduction

As outlined in Chapter 2, there are several examples of small area-based deprivation indices which have been developed by researchers using census data. These indices are not new, with the earliest versions having been developed nearly 40 years ago. Most of these indices have been created following a consistent set of steps which focus broadly on defining the small area, selecting items that measure deprivation relative to the context where the index was created and the computation of a deprivation score using the selected items. Other steps, such as sensitivity analysis and validation of the final index, were less common. While there is no standard framework for the development of such indices, and some choices taken by researchers have been based on convenience or tradition, there are a number of key stages that can be followed (Allik *et al.*, 2020).

Based on the literature presented in Chapter 2, and the specific contextual factors related to Malta and the Maltese census, which were discussed in Chapter 1, this chapter will present the development and validation of a small area-based deprivation index for Malta. As discussed in Chapter 2, predictive validity is the most common type of validity applied by researchers when validating small area-based deprivation indices. In the context of health inequalities analysis, predictive validity of a small area-based deprivation index would be its ability to predict a health gradient. This type of validation will be discussed in detail in Chapter 4, Chapter 5, and Chapter 6, where examples of the application of the index to health inequalities research will be presented. The rest of the chapter will be divided as follows: Section 3.2 will describe the data source used. Section 3.3 will describe the methods used and results will be presented in section 3.4. Finally, section 3.5 will discuss the findings presented in this chapter.

3.2 Data source

3.2.1 Spatial area and population coverage

As described in Chapter 1, the census of population and housing is conducted in Malta every ten years and aims to enumerate all persons resident in the country as at midnight on census night. The two most recent census data files were used in this study. While the census is usually conducted every ten years, the last two available census data files refer to reference years 2005 and 2011. The 2015 census date was shifted to 2011 to align the census date to the reference period outlined in the European Union (EU) regulation which was developed to harmonise the census exercise across all EU Member States. Both censuses were conducted in November. Persons enumerated in the census were all individuals considered usually resident on midnight of census night. 'Usually resident' is defined as a person living in their usual residence for at least 12 months, or with the intention to live in their usual residence for at least 12 months (National Statistics Office, 2014b).

The smallest area level available in the census is the enumeration area (EA). Enumeration areas applied within the census in Malta are developed for operational purposes only, so that the enumeration exercise can be conducted within small areas assigned to each enumerator. The EAs are therefore not intended to be used as outputs for statistical publication. EAs are also not fixed. Each census round, new EAs are devised with the purpose of dividing localities into small areas containing, on average, 150 households. The number of parts a locality is broken up into will therefore depend on the size of the population and the density of dwellings within localities at the time of the census. Given these limitations, an index produced at the EA level cannot be linked to other external data sources since the EAs cannot be mapped to area level data available in these sources. The only fixed small area breakdown with clear boundaries available are the 68 official localities described in Chapter 1. For this reason, while the selection of items for the index was conducted at the EA level and an EA index was produced, a locality index was also developed, and this was the index ultimately intended for wider use.

While the census covers the entire population, that is persons resident within both private households and institutions, the institutionalized population was excluded from the data used to create the index. There are two reasons for this – firstly, questions on dwelling and living conditions are not included in the enumeration of the institutionalised population.

Secondly, institutions are enumerated separately from private households and are assigned to dummy EAs which refer to institution type rather than a specific EA related to geographical location. Apart from the exclusion of the institutionalised population, persons living in private households who were assigned to dummy EAs were also excluded. These were records where data fields were imputed during post-census enumeration through administrative sources and were assigned to a dummy EA code even if locality of residence was available. Additionally, two persons classified as homeless in the 2011 census were also assigned to a dummy EA and were excluded. Table 3.1 presents a breakdown of the population within each census round and the final count included in this analysis. For both census years, less than 5% of the population were ultimately excluded from the analysis.

Table.3.1: Summary of census population included in analysis: 2005 and 2011

	2005		**2011**	
	N	**%**	**N**	**%**
Total enumerated population	404,962	100	417,432	100
Of which				
Institutionalised population	6,347	1.6	8,649	2.1
Private residents assigned to dummy EAs	8,959	2.2	8,882	2.1
Final population count included in analysis	**389,656**	**96.2**	**399,901**	**95.8**

EAs: Enumeration Areas; N: Count

3.2.2 Census questionnaire, domains, and items

The two questionnaires for both censuses (2005 and 2011) were reviewed and the questions in both, compared. Since the aim was to develop an index that could be updated with each round of the census, only questions present in each census were considered. In total, eight person-level questions and two dwelling-level questions were not retained in the 2011 census from the 2005 census. Conversely, two person-level questions and one dwelling-level question introduced in the 2011 census were also excluded. Based on the review of the literature described in Chapter 2, the questions in the census were operationalised into 19 items, grouped across six domains, which were considered to measure some aspect of deprivation in Malta. The items were estimated as a proportion of the population, households or dwellings within the geographical small area, either the EA or locality. Table 3.2 presents the items, with an explanation of how each was computed. Basic services cover access to a kitchen/kitchenette, bath/shower, toilet, and water supply. Appliances considered in the items related to lack of access to appliances within a dwelling were fridge/fridge-freezer, cooker and washing machine.

Table 3.2: Items considered for inclusion in the deprivation index

Domain	Item	Computation
Migration	Proportion without Maltese citizenship	Number of persons without Maltese citizenship/Total population
	Proportion 5+ not living in the same house 1 year from census day	Number of persons aged 5 and over not in same household 1 year from Census night/ Total population aged 5 and over
Employment	Proportion of males of working age who are unemployed	Number of unemployed males aged 15 – 64/ Total males aged 15 – 64
	Proportion of females of working aged who are unemployed	Number of unemployed females aged 15 – 64/ Total females aged 15 – 64
	Proportion of total working age population who are unemployed	Number of unemployed aged 15 – 64/Total population aged 15 – 64
	Proportion of employed population in elementary occupations	Number of employed persons aged 15 – 64 in elementary occupations/Total employed population aged 15-64
Education	Proportion of population aged 10 and over who are illiterate	Number of persons aged 10 and over who cannot read/write a simple sentence/ Total population aged 10 and over
	Proportion of population aged 15 and over who have no qualifications	Number of persons aged 15 and over who have no qualifications/Total population aged 15 and over
	Proportion of population aged 18 to 24 who are early school leavers	Number of persons 18 – 24 who obtained a secondary level of education or lower and were not students at the time of the census/Total population aged 18 - 24
Household type	Proportion of elderly living alone	Number of elderly (65+) living alone/ Total elderly population (65+)
	Proportion of private households which are single parent households with one or more dependent child	Single parent households with one or more dependent child/Total number of households
Living conditions	Proportion of population living in over-crowded private dwellings	Number of persons living within over-crowded occupied private dwellings/Total population in private occupied dwellings
	Proportion of population living in private occupied dwellings without at least one basic service	Number of persons living in private occupied dwellings without at least one basic service/ Total population living in private occupied dwellings

Domain	Item	Computation
	Proportion of population living in private occupied dwellings without at least one appliance	Number of persons living in private occupied dwellings without at least one appliance/ Total population living in private occupied dwellings
	Proportion of population living in private occupied dwellings without at least two appliances	Number of persons living in private occupied dwellings without at least two appliances / Total population living in private occupied dwellings
	Proportion of population living in private occupied dwellings without all three appliances	Number of persons living in private occupied dwellings without at least three appliances/ Total population living in private occupied dwellings
	Proportion of private occupied dwellings not owner occupied	Number of private dwellings not owner occupied/Total number of occupied private dwellings
	Proportion of persons living in private occupied dwellings needing serious repair/dilapidated	Number of persons living in dwellings needing serious repair/dilapidated/ Total number of private occupied dwellings
Dwelling status	Proportion of all dwellings that are vacant	Number of vacant dwellings/Total dwelling stock

To remain consistent with the census methodology and census outputs as disseminated by the National Statistics Office (NSO) (National Statistics Office, 2014a), the following definitions were retained when computing the items listed above:

Number of rooms: The total number of rooms in a dwelling include normal bedrooms, dining rooms, living rooms, habitable rooms in cellar/basement, kitchens, and study rooms. But exclude kitchenettes, verandas, corridors, washrooms, bathrooms, box rooms, garages, halls, and rooms used solely for business purposes.

Unemployed: Based on the International Labour Organisation (ILO) definition. A person is classified as unemployed if during the reference week they were: i) 'without work,' that is not in wage employment or self-employment; and ii) 'currently available for work,' that is, available for wage employment or self-employment within two weeks after the Census; and iii) 'seeking work,' that is, had taken specific steps to seek wage employment or self-employment during the four weeks prior to the census.

Elementary occupations: Classified according to the International Standard Classification of Occupations (ISCO). Covers the subdivisions 'cleaners and helpers;' 'agricultural, forestry and fishery labourers;' 'labourers in mining, construction, manufacturing, and transport;' 'food preparation assistants;' 'street and related sales and service workers;' 'refuse workers' and 'other elementary workers.'

Dependent child: This is defined as either i) All persons below 16, or ii) Persons aged 16 to 24, living in a household of which at least one of their parents is a member and who are economically inactive.

Dwelling: A building that is used entirely or primarily as a residence, including any associated structures, such as garages, and all permanent fixtures customarily installed in residences. A household resides in a dwelling, whenever the latter is occupied. Otherwise, a dwelling is said to be unoccupied or vacant. While for the 2011 census the vacant dwellings that are seasonally vacant could be excluded from the dwelling stock at the EA level, this was not possible for the 2005 data and therefore the whole dwelling stock was retained.

Owner occupied dwelling: Dwellings where at least one occupant of the dwelling owns parts or the whole of the dwelling.

In addition to the standard definitions used in the Census as outlined above, the definition used for a crowded dwelling was specifically a dwelling where the number of persons exceeded the number of rooms based on the definition of a room discussed above. Finally, since the indicator 'early school leavers' is regularly estimated within the Labour Force Survey (LFS) and aims to monitor the proportion of youth not continuing education or training after formal schooling (Eurostat, 2023a), this definition was used for the item related to early school leavers considered in this analysis.

3.2.3 Data request

Census data are available through the NSO and several reports were published online for both the 2005 and 2011 census rounds (National Statistics Office, 2007, 2009, 2014b). Since data at the EA level are not disseminated, a specific request was required to obtain the data at this level of breakdown. Prior to requesting the data, a meeting was held with staff at the NSO to discuss the data requirements. While initially the plan was to request a microdata file which was intended to be used to produce the relevant aggregate indicators, following discussions, it was agreed that such a request may lead to suppression of records due to issues of confidentiality. Given the small population size of EAs, micro-data may become identifiable when broken down across several variables and thus suppression of records, deemed identifiable, would be required. Since the index was intended to cover the entire enumerated population, it was agreed that the aggregate items as proportions by EA and locality would be requested rather than a micro-data file. Thus, the NSO provided proportions for the above items separately by EA and locality. The NSO does not require ethical approval to process any of its data requests as it processes all requests within the legal parameters stipulated by the Malta Statistics Authority Act (Government of Malta, 2008) and for census data specifically, requests are also processed within the parameters of the Census Act (Government of Malta, 1948). Since no micro-data was requested, the NSO also did not require that the researcher enter a legal contract to receive and process this data. Following the formal request, the aggregate data tables containing proportions for 2011 and 2005 at the EA and locality level were deemed to be unidentifiable by the NSO and were provided to the researcher in March 2022.

The table below summarises the data provided and used for the analysis which will be discussed:

Table 3.3: Summary of data source used for the development of the indices

Census reference years	2011 and 2005
Number of items	19
Numerical format of items provided	Proportions within spatial area
Spatial area breakdown	EA and locality
Number of aggregate data tables	4

3.3 Methods

3.3.1 Selection of items and computation of index

Initial analysis was conducted for all 19 items to assess the variation of the item values across EAs. The mean, median and range of values was estimated. Items included in the index would expect to have enough ability to discriminate between the EAs based on the range of values between areas. Spearman correlations were also estimated to assess the linear relationship between the 19 items, for each census year separately. Since it is assumed that all items measure some component of deprivation in Malta, the 19 items were expected to be correlated with each other. However highly correlated items may indicate that the items were measuring the same aspect of deprivation and could be redundant. To avoid redundancy of items in the final index, highly correlated items were reviewed for possible exclusion.

Principal component analysis (PCA) was the method selected to develop the index. PCA is a technique that involves summarising many correlated variables into a set of new uncorrelated components, each of which is a linear combination of the original variables. If the original variables are highly correlated, much of the variation can be summarised by a reduced set of components, hence enabling more parsimonious analysis by reducing possible redundancy. The first principal component accounts for the largest proportion of variance in the original dataset, with each subsequent component explaining less of the variance. As applied in other indices described in Chapter 2, the principal component was assumed to be the component containing the items that measure deprivation in Malta, therefore the aim was to select the first principal component to create the index. PCA was chosen due to the extensive application of this method in the creation of other census-based deprivation indices as outlined in Chapter 2. PCA was also deemed to be the most appropriate approach because as a data reduction technique it allows for the selection of items to be partly data driven. Finally, PCA offered the ability to assess the relative weight of each item on the principal component through factor loadings. This allowed for an assessment of the relative weights of each item for the two respective census years. The 19 items were standardised into Z-scores prior to running the PCA. The standardisation was used to prevent variables with larger proportions, or larger ranges, from having a disproportionate influence on the index.

The PCA was conducted using varimax rotation, rotating the factors minimizes the complexity of the factor loadings to make the structure simpler to interpret. Varimax is an orthogonal rotation method that minimizes the number of variables that have high loadings on each factor. When only one factor is extracted, no rotation is conducted and the unrotated loadings are presented. The first round of PCA was conducted with all 19 items, including those with negative correlations and those that were strongly correlated. An iterative process was applied with exclusion of items prior to running of the next PCA. The process was repeated until all items that remained loaded only on the principal component. The threshold for an item to be considered as loading on a component was set for 0.3 in absolute value, this limit is an accepted standard level in the literature (Tavakol and Wetzel, 2020).

Since data were available for two census years, two separate PCA's were conducted with the 2005 and 2011 data, in parallel, with the aim of comparing the outputs of the two analyses when deciding upon the items to be excluded with each iteration. More weight was given to the results of the 2011 PCA in the decision process, since the data were more recent and thus were assumed to be more relevant to the data patterns that reflect the current scenario within the country. Finally, since the 2021 census was ongoing at the time of this analysis and the questionnaire for this census was publicly available, it was also reviewed and taken into consideration when making the decisions on the items to be excluded based on whether the item could ultimately be included in the index if updated with the 2021 data. The PCA was conducted with the data at the EA level since the goal was to use the smallest area-level data available to select the items for the index. Once the final items were selected, a second PCA was conducted at the locality level to extract factor loadings specifically for this dataset.

Apart from the factor loadings, the variance explained (%) by the principal component, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were also estimated for each PCA. The latter two measures indicate whether the data is suitable for factor analysis. A KMO value ranges from 0 to 1, the closer the value is to 1, the higher the sampling adequacy. A Bartlett's test of sphericity compares the correlation matrix of the items to the identity matrix. A p-value below the threshold indicates that the correlation matrix diverges significantly from the identity matrix and thus is suitable for data reduction analysis.

The final index was computed by weighting the Z-scores at each spatial level with the respective factor loadings and summing up the items into a composite score for each EA and locality in the two census years. Since the Z-scores of the items were used, the resulting index produced a range of numbers that were positive to negative. Higher positive scores indicated increasing deprivation, while higher negative scores indicated decreasing deprivation. The index scores were grouped into quintiles of deprivation. To do this, the data was first ranked by score, then by population size for the two geographic levels. Quintiles were produced by segmenting the ranked scores based on a proportionate distribution of the population. This was done to ensure that the population counts within the quintiles were distributed as evenly as possible. Quintile 1 represented the least deprived areas and quintile 5 represented the most deprived areas.

Given that the EA is a smaller geographic area, producing an index at the locality level may lead to pockets of deprivation in localities with multiple EAs being diluted when considering locality deprivation. To assess the extent of this dilution, the locality level quintile was assigned to each EA within the locality and this data was cross tabulated for each year. With this information, it was possible to estimate the misclassification of EA quintile when aggregated to the locality level quintile based on the deprivation index estimated at the locality level.

The PCA analysis and computation of the index was conducted in SPSS version 20. Maps were produced using QGIS version 3.22.5. Maps were produced using the official Maltese base map designating the local authority boundaries which is available online from the Planning Authority geoportal.

3.3.2 Sensitivity analysis

Before finalising the index, sensitivity analysis was conducted to assess the impact of selected changes in the analytical process on the index output. This focused on primarily two changes, the impact of removing very small EAs when running the PCA and the impact of excluding items from the index following the resulting index from the first PCA.

As each variable at the EA level was expressed as a proportion, areas with small denominators may have been deemed unreliable as they are more likely to have volatile estimates and proportions of 0 or 1. Due to the small size of the population of Malta and the possibility of smaller EAs in specific localities, it was not desirable to exclude any EAs in the analysis to avoid excluding substantial portions of specific areas, especially those that are less densely populated. However, since it was possible that proportions from small areas could impact the PCA results, sensitivity analysis was conducted to assess the impact of excluding the EAs within the lowest 5% of the EA population size range, ranked highest to lowest, for the 2011 census. This excluded 51 EAs with a population of 174 or less, reducing the number of EAs from 1022 to 971. The removal of these EAs excluded 5,122 persons from the analysis, 1.3% of the population included for analysis from the 2011 census. The PCA for 2011 was replicated to determine if the same items load onto the principal component as extracted in the first run, and to assess the factor loadings on the items in the final principal component when compared to analysis with the full set of EAs. Table 3.4 shows that not all localities had small EAs, with only 13 localities having EAs excluded based on the threshold set above. The table presents the total population and number of EAs in these localities in the full analysis and the sensitivity analysis, respectively.

Table 3.4: Summary of localities and number of small EAs excluded in the sensitivity analysis using 2011 data

Locality	**Population size first run**	**No. EAs first run**	**Population size sensitivity analysis**	**No. EAs sensitivity analysis**
San Pawl il-Baħar	15,038	63	13,091	48
Tas-Sliema	12,717	50	12,311	46
Iż-Żejtun	10,977	23	10,809	22
Ir-Rabat	10,682	26	10,636	25
Marsaskala	10,628	32	10,297	30
Il-Ħamrun	8,729	24	8,557	23
Il-Mellieħa	8,249	29	7,906	22
San Ġiljan	7,587	26	7,294	24
L-Imsida	7,265	24	6,950	22
Tal-Pieta	3,687	10	3,517	9
Iż-Żebbuġ	1,682	14	1,075	4
Ix-Xgħajra	1,507	5	1,387	4
Il-Munxar	1,020	6	816	2
Total	**99,768**	**332**	**94,646**	**281**

EA: Enumeration area

Sensitivity analysis was also conducted to assess the impact of the exclusion of items from the computation of the index. Initial face validity assessment of the preliminary index at the locality level was conducted to assess whether the deprivation quintile assigned to the localities was consistent over the two time points and as expected based on subjective knowledge. This face validity provided an indication of whether there may be issues with items included in the index and their subsequent influence on the deprivation score. In such cases, sensitivity analysis was conducted by re-running the PCA without the item and extracting new factor loadings to compute an updated deprivation score, as required. More details are provided in section 3.4.2 regarding this part of the sensitivity analysis, including the justification for the item tested for exclusion and the outcome of the exclusion on the index.

3.3.3 Validation of the final index

While validation of any composite measure is important to check that the measure is capturing what it is intends to measure, there are generally no standard methods available to validate small area-based deprivation indices. One of the strongest measures of validity of a deprivation measure is its ability to describe differences in health outcomes by deprivation, since it is expected that health outcomes behave along a deprivation gradient (Allik *et al.*, 2020). This type of predictive validity will be presented in Chapter 4 through Chapter 6, where the index will be applied to health inequalities analysis. An attempt has been made to conduct additional validation, as far as is possible.

Internal consistency measured through the Cronbach's alpha, is one of the most common tools to measure the reliability of a composite measure. Measured on a scale of 0 to 1, it quantifies the level of agreement between the items and is said to determine whether they collectively measure the same characteristic. The closer the value is to 1, the higher the internal consistency.

Content validity is the agreement that the items included in the index capture the general concept of deprivation. Since the notion of deprivation is seen to be relative, there is no one clear definition of what it means to be deprived as it is an area-based measure reflects the specific reality at the time of data capture. This type of validation was conducted by comparing the domains and items included in the final index to the indices reviewed in Chapter 2.

Finally, criterion validity tests how accurately a score measures the outcome it was designed to measure. This is usually tested by correlating it to other external measures of deprivation. Since no measures of deprivation are available at the locality level, a weighted average quintile was estimated for the six districts for the locality index based on the 2011 census. The weight applied to each locality was based on the size of the locality with respect to the total population of the district. This average quintile score for 2011 was compared to the social, income and living condition indicators available for 2011 at the district level, as published in the annual NSO regional statistics publication (National Statistics Office, 2017).

3.4 Results

3.4.1 Preliminary index

Table 3.5 presents the unstandardised mean, median and ranges for the proportions (presented as percentages) for each item in the respective census years estimated for the EAs. These have been ranked highest to lowest. For each census round, the top five proportions within the EAs are for no qualifications, vacant dwellings, early school leavers, occupied dwellings that are not owner occupied and elderly living alone. Mean proportions for items between 2005 and 2011 have reduced for no qualifications, early school leavers, elementary occupations, appliance poverty (1 or 2 appliances), crowded housing, persons living in dwellings needing serious repair/dilapidated and persons living in dwellings without basic services. On the other hand, the mean proportions for persons without Maltese citizenship, vacant dwellings and single-parent households have increased.

Table 3.5: Mean, median and range for the 19 unstandardised items at EA level: 2005 and 2011

Items	2005						2011				
	Mean	SD	Median	Min	Max		Mean	SD	Median	Min	Max
% population (+15) with no qualifications	56.7	15.0	58.8	14.7	90.7		37.0	11.9	37.6	7.1	63.8
% dwellings that are vacant	23.8	16.2	20.5	0.0	90.8		30.7	17.8	26.6	1.9	94.2
% population 18-24 who are early school leavers	43.8	17.5	42.9	0.0	90.0		27.5	16.3	25.0	0.0	100.0
% occupied dwellings not owner occupied	24.9	18.3	20.7	0.0	91.4		24.4	17.3	20.3	0.0	92.5
% elderly (65+) living alone	22.0	10.0	22.0	0.0	75.0		22.1	9.9	21.6	0.0	100.0
% population in elementary occupations	12.1	7.9	11.2	0.0	48.5		9.8	5.6	9.2	0.0	34.5
% population (10+) illiterate	6.9	4.6	6.0	0.0	25.9		6.0	4.1	5.2	0.0	25.4
% unemployed males of working age	5.2	3.5	4.4	0.0	23.6		5.6	4.1	4.7	0.0	33.3
% population without Maltese citizenship	3.1	4.1	1.5	0.0	34.1		5.4	7.2	2.7	0.0	67.9
% population living in dwellings without at least one appliance	7.6	3.8	6.9	0.0	31.4		5.0	3.3	4.3	0.0	25.7
% population (5+) who moved 1 year from Census day	4.7	3.6	3.7	0.0	29.8		4.9	4.1	3.7	0.0	41.6
% population living in over-crowded dwellings	5.9	4.9	4.7	0.0	31.8		4.6	4.3	3.6	0.0	27.3
% unemployed of working age	4.1	2.3	3.7	0.0	17.8		4.4	2.8	3.8	0.0	27.3
% single parent households with dependent children	2.5	1.9	2.2	0.0	12.5		3.6	2.4	3.1	0.0	18.5
% unemployed females of working age	3.0	2.0	2.7	0.0	20.0		3.1	2.5	2.7	0.0	28.6
% population living in dwellings needing serious repair/dilapidated	4.1	3.6	3.2	0.0	28.1		1.6	1.9	1.0	0.0	13.5
% population living in dwellings without at least two appliances	0.9	1.0	0.6	0.0	7.2		0.4	0.7	0.2	0.0	5.4
% population living in dwellings without at least one basic service	1.6	1.9	1.0	0.0	13.7		0.4	0.7	0.0	0.0	7.5
% population living in dwellings without all three appliances	0.2	0.4	0.0	0.0	4.4		0.1	0.2	0.0	0.0	1.5

SD: Standard Deviation

The remainder of the analysis will refer to the Z-standardised items and not the unstandardised proportions. For both 2005 and 2011 the items proportion without Maltese citizenship, housing stability, vacant properties and single parent households were negatively correlated with more than one other item. While the items unemployment amongst males and unemployment total were highly correlated with each other (Spearman's rho 0.897 in 2005 and 0.874 in 2011); as were illiteracy and elementary occupations (Spearman's rho 0.850 in 2005 and 0.795 in 2011) (see Appendix B).

Table 3.6 presents the results of the six iterations of PCA analysis of the EA Z-scores which were conducted to extract the final items on the principal component in 2005 and 2011. Below is an overview of the six iterations with justification for the exclusion of items at each step:

- **Iteration 1**

  All 19 items were included in the first PCA. This iteration extracted 4 factors for the 2011 data and 5 factors for 2005. The variance explained by the principal component was 33.9% for 2011 and 35.6% for 2005. Results for the KMO and Bartlett's test showed that the data was suitable for factor analysis. Five of the items did not load on the principal component for both the 2005 and 2011 PCA – three of the items were those which were negatively correlated with other items as described previously – % population without Maltese citizenship (MT_CITIZEN), % population (5+) who moved 1 year from Census Day (HOUSE_STABILITY), and % dwellings that are vacant (VACANT). Apart from these, % population living in dwellings without all three appliances (APP_POV_3) and % elderly (65+) living alone (ELDERLY_ALONE) did not load on the principal component. These five items were excluded from the second iteration of the PCA.

- **Iteration 2**

  The second iteration was conducted on 14 items. This PCA extracted 3 factors for both 2005 and 2011. The variance explained by the principal component was 44.2% for 2011 and 46.4% for 2005. Results for the KMO and Bartlett's test showed that the data was suitable for factor analysis. Four of the items did not load on the principal component for both the 2005 and 2011 PCA – % unemployed females of working age (UNEMPLOYED_F), % population living in dwellings without at least one appliance (APP_POV_1), % population living in dwellings without at least two appliances (APP_POV_2) and % population living in dwellings needing serious repair/dilapidated (STATUS_REPAIR). The item % single parent households with dependent children (SINGLE_PARENTS) did not load on the principal component for 2011. It did load on the principal component for 2005, however with a negative factor loading. Since the item % single parent households with dependent children (SINGLE_PARENTS) exhibited negative correlations with other items in the initial correlation matrix as well, it was excluded in the third iteration. Apart from this item, % unemployed females of working age (UNEMPLOYED_F) was also excluded since two other items measuring unemployment were in the analysis. To still retain some of the items on dwelling status at this stage of the process, only % population living in dwellings without at least two appliances (APP_POV_2) was excluded in the third iteration as it did not load on the principal component in the first iteration for 2011 and % population living in dwellings without at least one appliance (APP_POV_1) and % population living in dwellings needing serious repair/dilapidated (STATUS_REPAIR) loaded for both years in the first iteration.

- **Iteration 3**

  The third iteration was conducted on 11 items. This PCA extracted 3 factors for 2005 and 2 factors for 2011. The variance explained by the principal component was 51.6% for 2011 and 55.3% for 2005. Results for the KMO and Bartlett's test showed that the data was suitable for factor analysis. This iteration produced opposing principal components for the two years. The items % unemployed males of working age (UNEMPLOYED_M), % occupied dwellings not owner occupied (RENTAL), % population in elementary occupations (ELEM_OCC), % population living in dwellings without at least one appliance (APP_POV_1) and % population living in dwellings needing serious repair/dilapidated (STATUS_REPAIR) loaded on the principal component for 2011 but loaded on the second component for 2005 (not shown). The principal component for 2005 included the items % population (+15) with no qualifications (NO_QUAL) and % population (10+) illiterate (ILLITERATE) which loaded on the second component for 2011 (not shown). These two items were highly correlated in both years, and therefore to reduce redundancy, one of them was excluded for the fourth iteration. Since the % population (+15) with no qualifications (NO_QUAL) item had a higher loading, % population (10+) illiterate (ILLITERATE) was excluded.

- **Iteration 4**

  The fourth iteration was conducted on 10 items and produced 2 factors for each census year. The variance explained by the principal component was 51.3% for 2011 and 54.5% for 2005. Results for the KMO and Bartlett's tests showed that the data was suitable for factor analysis. The items % population in elementary occupations (ELEM_OCC) and % population (+15) with no qualifications (NO_QUAL) did not load on the principal component for the 2011 data, while the items % population living in dwellings without at least one appliance (APP_POV_1), % population living in dwellings without at least one basic service (BASIC_SERVICE) and % population living in dwellings needing serious repair/dilapidated (STATUS_REPAIR) did not load on the 2005 data. Since the factor loadings for % population in elementary occupations (ELEM_OCC) and % population (+15) with no qualifications (NO_QUAL) were high for the 2005 data, they were retained. So as not to exclude all the items that did not load on 2005 principal component related to dwellings and living conditions, only % population living in dwellings without at least one appliance (APP_POV_1) and % population living in dwellings without at least one basic service (BASIC_SERVICE) were excluded as they cannot be computed with data from the 2021 census based on the published questionnaire.

- **Iteration 5**

  The fifth iteration was conducted on 8 items and the analyses produced just the principal component for both census years. The variance explained by the principal component was 55.3% for 2011 and 58.4% for 2005. Results for the KMO and Bartlett's test showed that the data was suitable for factor analysis. All items loaded on the 2005 principal component, while the items for % population in elementary occupations (ELEM_OCC) and % population (+15) with no qualifications (NO_QUAL) did not load on the 2011 principal component. Since the item % population 18-24 who are early school leavers (ESL) is also an item within the domain of education and loaded consistently for both years in each iteration, the item % population (+15) with no qualifications (NO_QUAL) was excluded for the sixth iteration since its performance was less consistent across PCAs.

- **Iteration 6**

  The sixth iteration was conducted on 7 items, and it was the final iteration since all these items loaded on the principal component for both 2005 and 2011. The variance explained by the principal component was equal for both years at 57.8%. Results for the KMO and Bartlett's test showed that the data was suitable for factor analysis. The items were % unemployed of working age (UNEMPLOYED_T), % unemployed males of working age (UNEMPLOYED_M), % population 18-24 who are early school leavers (ESL), % population living in over-crowded dwellings (CROWDED), % occupied dwellings not owner occupied (RENTAL), % population in elementary occupations (ELEM_OCC) and % population living in dwellings needing serious repair/dilapidated (STATUS_REPAIR). The items % population 18-24 who are early school leavers (ESL) and % population living in over-crowded dwellings (CROWDED) were the most consistent across the 6 iterations as they loaded on the principal component each time, for both years. For both years, items related to employment had the highest factor loadings, while the item related to status of repair had the lowest factor loadings.

Table 3.6: Results of the PCA iterations using EA Z-standardised items: 2005 and 2011

Zscore of item (EA)	Rotated loadings (Iteration 1)		Rotated loadings (Iteration 2)		Rotated loadings (Iteration 3)		Rotated loadings (Iteration 4)		Loadings (Iteration 5)		Loadings (Iteration 6)	
	2011	**2005**	**2011**	**2005**	**2011**	**2005**	**2011**	**2005**	**2011**	**2005**	**2011**	**2005**
UNEMPLOYED_T	0.842	0.474	:	0.473	0.910	0.720	0.921	0.812	0.930	0.816	0.832	0.841
UNEMPLOYED_M	0.768	0.561	0.403	0.484	0.883	:	0.893	0.813	0.902	0.827	0.847	0.852
ESL	0.734	0.716	0.704	0.709	0.542	0.762	0.507	0.808	0.523	0.808	0.805	0.780
CROWDED	0.678	0.688	0.569	0.470	0.491	0.418	0.426	0.649	0.447	0.771	0.707	0.780
RENTAL	0.650	0.655	0.332	:	0.760	:	0.714	0.475	0.706	0.728	0.785	0.763
UNEMPLOYED_F	0.644	:	:	:	E	E	E	E	E	E	E	E
ELEM_OCC	0.608	0.713	0.859	0.807	0.313	:	:	0.713	:	0.722	0.728	0.681
SINGLE_PARENTS	0.537	:	:	-0.322	E	E	E	E	E	E	E	E
MT_CITIZEN	:	:	E	E	E	E	E	E	E	E	E	E
HOUSE_STABILITY	:	:	E	E	E	E	E	E	E	E	E	E
NO_QUAL	0.344	0.829	0.827	0.868	:	0.846	:	0.768	:	0.829	E	E
VACANT	:	:	E	E	E	E	E	E	E	E	E	E
ILLITERATE	0.482	0.838	0.817	0.771	:	0.820	E	E	E	E	E	E
ELDERLY_ALONE	:	:	E	E	E	E	E	E	E	E	E	E
APP_POV_1	0.333	0.556	:	:	0.558	:	0.531	:	E	E	E	E
BASIC_SERVICE	:	0.685	:	0.365	0.485	0.347	0.478	:	E	E	E	E
APP_POV_3	:	:	E	E	E	E	E	E	E	E	E	E
APP_POV_2	:	0.308	:	:	E	E	E	E	E	E	E	E
STATUS_REPAIR	0.348	0.673	:	:	0.428	:	0.376	:	0.340	0.581	0.583	0.590
Variance (%)	33.9	35.6	44.2	46.4	51.6	55.3	51.3	54.5	55.3	58.4	57.8	57.8
No. factors	4	5	3	3	2	3	2	2	1	1	1	1
KMO sampling adequacy	0.796	0.799	0.788	0.796	0.879	0.887	0.858	0.863	0.824	0.831	0.815	0.831
Bartlett's Sig.	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

: - Item did not load on primary component; E – item excluded in PCA

The final iteration of the PCA had two items related to unemployment, both with high factor loadings. In both years, these items were highly correlated. To avoid redundancy, one of the items was excluded from the final index computation. While the item % unemployed males of working age (UNEMPLOYED_M) had a slightly higher loading compared to % total unemployed of working age (UNEMPLOYED_T) in 2011, it was decided to retain the overall unemployment measure since the total unemployment rate is a standard indicator produced for national monitoring. Furthermore, total unemployment is a more relevant indicator as more women enter or re-enter the job market and indeed the use of male unemployment rather than total unemployment has been criticised in other indices, such as Carstairs (Allik *et al.*, 2016).

Before excluding the item % unemployed males of working age (UNEMPLOYED_M) from the index computation, the PCA was run for both years with the two unemployment items separately. Table 3.7 presents the factor loadings for the PCA with male and total unemployment items separately for each census year compared to the output of the sixth iteration which contained both items. The variance explained for the principal component when using the male unemployment item was marginally higher when compared to the variance when using total unemployment, however this was only a difference of 0.5%, therefore it was expected not to dramatically impact the index performance. Given the possible relevance, in the future, of retaining a total unemployment item and the marginal differences in variance and loadings, it was decided to retain this item in the final index.

Table 3.7: Factor loadings for PCA using total unemployment and male unemployment items separately: 2005 and 2011

Z-Score of items (EA)	**2005** Loadings iteration 6	Loadings (Total)	Loadings (Male)	**2011** Loadings iteration 6	Loadings (Total)	Loadings (Male)
% population living in over-crowded dwellings	0.780	0.806	0.807	0.707	0.750	0.745
% population 18-24 who are early school leavers	0.780	0.799	0.795	0.805	0.821	0.822
% occupied dwellings not owner occupied	0.763	0.781	0.785	0.785	0.796	0.790
% unemployed of working age	0.841	0.769	-	0.832	0.756	-
% unemployed males of working age	0.852	-	0.782	0.847	-	0.773
% population in elementary occupations	0.681	0.706	0.707	0.728	0.758	0.767
% population living in dwellings needing serious repair/dilapidated	0.590	0.630	0.629	0.583	0.618	0.620
Variance (%)	57.8	56.4	56.8	57.8	56.6	57.1

EA: Enumeration Area

A final PCA was conducted using the Z-scores for the six items at the locality level to extract factor loadings to compute the index at this geographical breakdown. Loadings and variance explained by the principal components for 2005 and 2011 using the locality level data are presented in table 3.8. Variance and factor loadings were greater at the locality level than the EA level.

Table 3.8: Results for PCA using items at the locality level: 2005 and 2011

Z-Score of items (locality)	**2005** Factor loadings	**2011** Factor loadings
% population living in over-crowded dwellings	0.940	0.925
% population 18-24 who are early school leavers	0.892	0.924
% occupied dwellings not owner occupied	0.835	0.885
% unemployed of working age	0.840	0.905
% population in elementary occupations	0.801	0.825
% population living in dwellings needing serious repair/dilapidated	0.752	0.754
Variance (%)	71.4	76.0

The deprivation score was computed for each EA and locality by weighting the Z-scores with the factor loadings than summing them as per the example below:

LOC_Dep_2005$_i$ = ∑((Z-score$_i$(CROWDED)* 0.940), (Z-score$_i$ (ESL)*0.892), (Z-score$_i$ (RENTAL)* 0.835), (Z-score$_i$ (UNEMPLOYED)* 0.840), (Z-score$_i$ (ELEM_OCC)* 0.801), (Z-score$_i$ (STATUS_REPAIR)*0.752)

The 2005 scores ranged from 13.21 to -5.90 at the EA level and 13.76 to -7.84 at the locality level, while the 2011 scores ranged from 15.15 to -5.89 at the EA level and 15.91 to -6.31 at the locality level. Figure 3.1 shows the scores for both indices by locality. Apart from a few localities, scores for the 2005 and 2011 index remained relatively consistent in terms of direction.

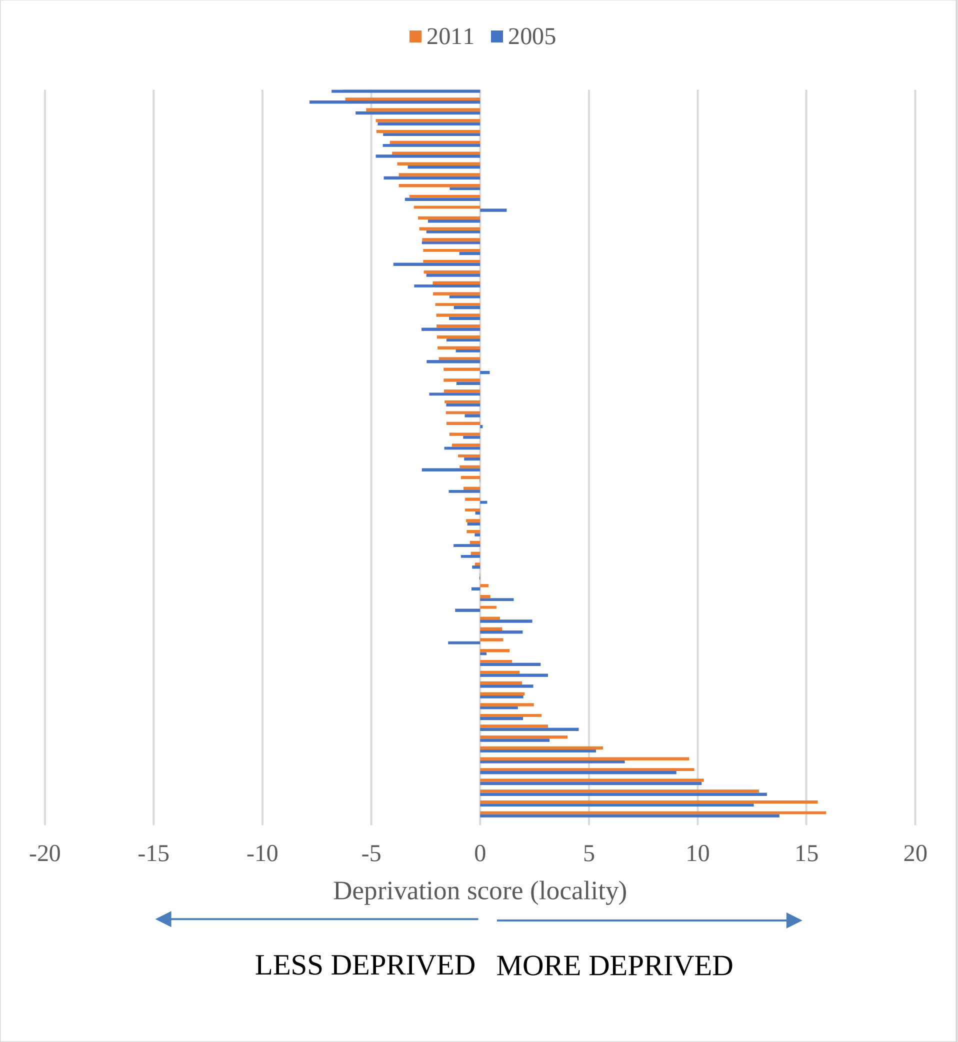


Figure 3.1: Deprivation scores for the 68 localities from preliminary index: 2005 and 2011

Figures 3.2 and 3.3 show the deprivation quintile for each locality mapped to the local administrative base map for 2005 and 2011, respectively. The base map outlining the localities can be found in Appendix C. Since the EAs are not based on fixed geographical boundaries, the same maps could not be produced for the EA level index. A preliminary assessment of the maps shows that the localities classified as most deprived were clustered within the Southern Harbour district. None of the most deprived localities were found in Gozo, the smaller sister island.

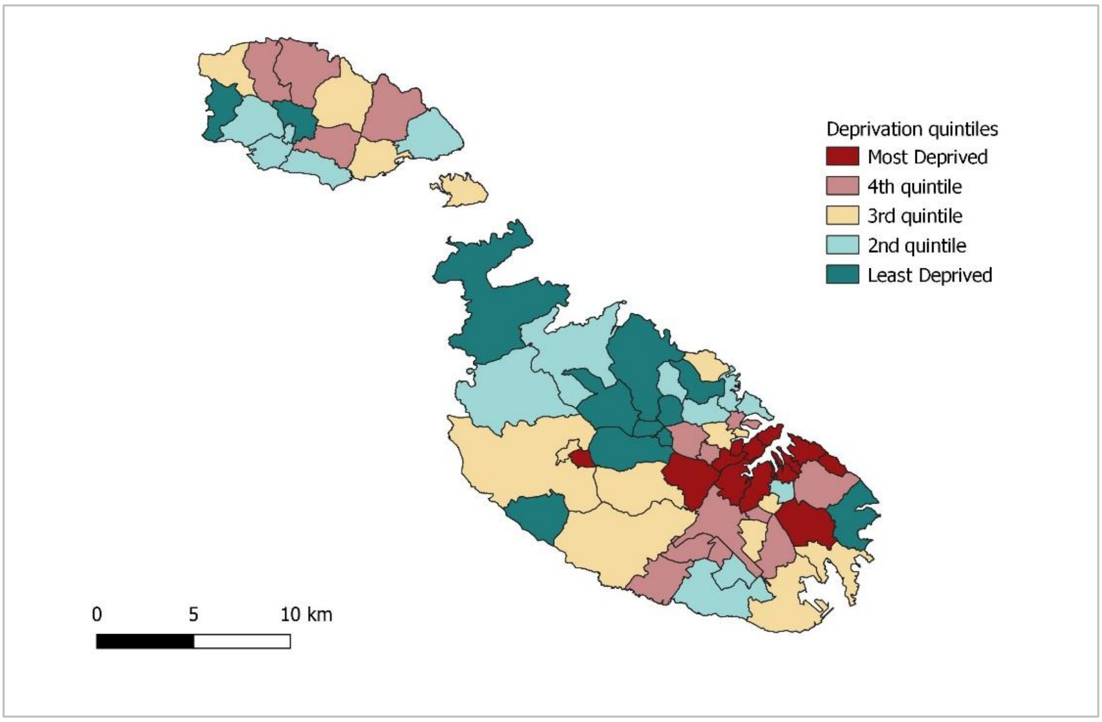


Figure 3.2: Preliminary deprivation index quintiles for the 68 localities: 2005

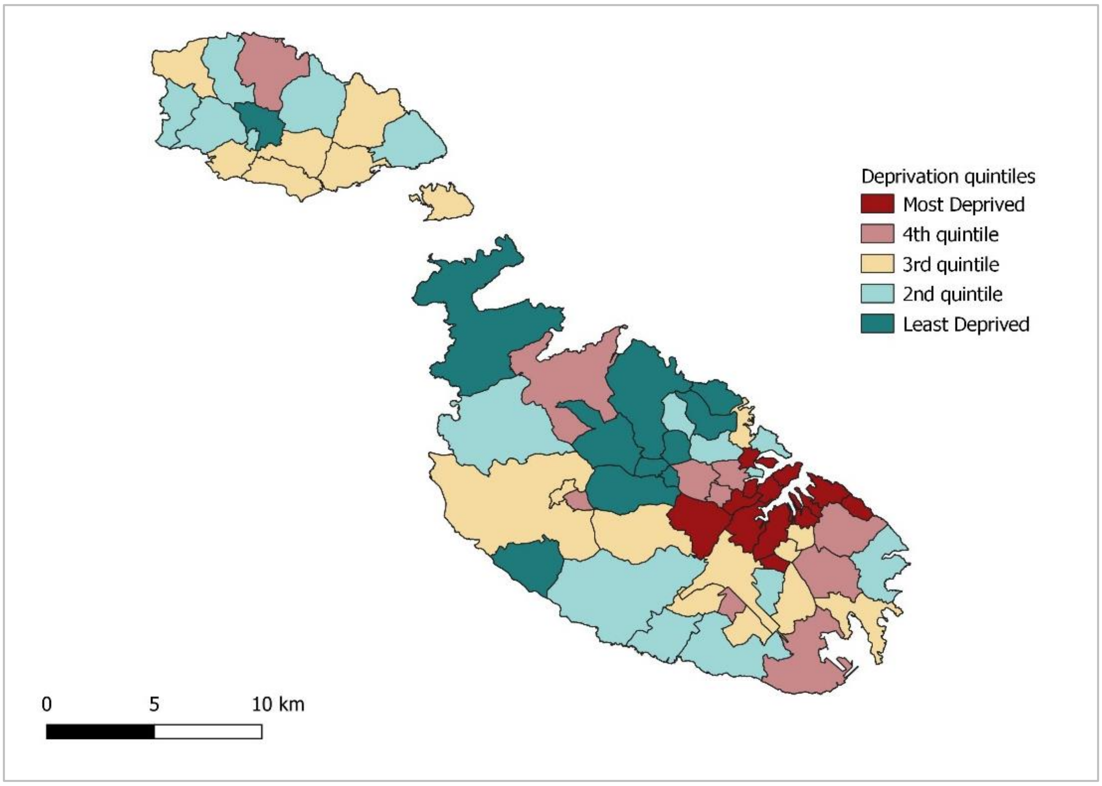


Figure 3.3: Preliminary deprivation index quintiles for the 68 localities: 2011

3.4.2 Sensitivity analysis

3.4.2.1 Exclusion of EAs with small population sizes

Sensitivity analysis was conducted to assess the impact on the PCA of excluding the EAs within the lower 5% of the EA population size range for the 2011 census. This was done to determine if excluding small areas resulted in the same items loading onto the principal component in the final run, and whether there were large differences in the factor loadings when compared to the analysis with the full set of EAs.

Table 3.9 shows the results of the PCA using the reduced number of EAs as well as the result of the final PCA used to compute the index at the EA level. The reduced EA PCA analysis produced the same principal component after the fourth iteration, with the fifth iteration being conducted excluding the male unemployment item to compare to the final iteration of the previous analysis. Factor loadings on the final reduced PCA (iteration 5) were generally like those in the full analysis, with only the item for total unemployment having a relatively higher loading in the reduced PCA. The variance in the reduced PCA was slightly higher than that of the full PCA analysis. However, given that the reduced and full analysis had the same items load on the principal component and the rank of the weights of the loadings in both were the same, this was not expected to impact the deprivation score considerably. It was decided to keep the full EA PCA analysis when estimating the EA level deprivation score so as not to exclude any EAs, especially since such an exclusion would only impact some localities as discussed previously.

Table 3.9: PCA results using 2011 data conducted excluding small EAs

Z-score of item (EA)	Rotated loadings (reduced EAs)	Rotated loadings (reduced EAs)	Rotated loadings (reduced EAs)	Loadings (reduced EAs)	Loadings (reduced EAs)	Loadings (final full analysis)
% population (+15) with no qualifications	0.872	0.524	:	E	E	E
% population (10+) illiterate	0.839	0.596	:	E	E	E
% population in elementary occupations	0.838	0.730	0.378	0.770	0.792	0.758
% population 18-24 who are early school leavers	0.735	0.793	0.567	0.823	0.849	0.821
% unemployed males of working age	0.645	0.828	0.843	0.867	E	E
% population living in dwellings without at least one appliance	0.594	0.358	0.530	0.683	E	E
% population living in over-crowded dwellings	0.593	0.760	0.594	0.729	0.776	0.750
% population living in dwellings without at least one basic service	0.529	:	E	E	E	E
% occupied dwellings not owner occupied	0.524	0.694	0.830	0.804	0.808	0.796
% population living in dwellings needing serious repair/dilapidated	0.523	0.366	0.580	0.643	0.653	0.619
% single parent households with dependent children	:	E	E	E	E	E
% unemployed of working age	0.595	0.824	0.864	0.862	0.817	0.756
% unemployed females of working age	:	E	E	E	E	E
% dwellings that are vacant	:	E	E	E	E	E
% population without Maltese citizenship	0.378	:	E	E	E	E
% elderly (65+) living alone	0.334	:	E	E	E	E
% population (5+) who moved 1 year from Census Day	0.392	:	E	E	E	E
% population living in dwellings without all three appliances	:	E	E	E	E	E
% population living in dwellings without at least two appliances	0.329	:	E	E	E	E
Variance (%)	37.4	44.8	58.8	60.3	61.6	57.8
No. of components	4	3	2	1	1	1
KMO sampling adequacy	0.816	0.888	0.888	0.865	0.862	0.815
Bartlett's Sig.	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

EA – Enumeration Area; : - Item did not load on primary component; E – item excluded in PCA

3.4.2.2 Excluding items from the index

Initial review of the classification of localities by deprivation quintiles showed results that were not expected for the locality of Mdina. Mdina is one of the localities with specific features that makes it different from the other localities in Malta. Firstly, it is the locality with the smallest population with 278 persons enumerated in the 2005 census and 239 in the 2011 census. Due to its size, in both censuses, it was covered by only one EA. Mdina is an area of significant historical value and was the home for noble families from the 12th century. It is a small fortified medieval town which used to be the capital of Malta prior to the construction of the current capital city, Valletta in the 16th century. Due to the historical significance of the architecture and archaeological remains, the whole of Mdina has been designated an Urban Conservation Area meaning there are restrictions on development in the area to preserve the townscape. Given this limited development and the historical background of families living in Mdina, the average age of Mdina is the highest of all localities, standing at 52.4 in 2011, compared to the average of the total population which stood at 40.5 years (National Statistics Office, 2014b).

In both the 2005 and 2011 index, Mdina was classified as deprived – most deprived (quintile 5) in 2005 and quintile 4 in 2011. The NSO produces unofficial regions for Malta basing on the Property Price Index (PPI) which is an index that measures value of properties at the time of purchase. Localities are unofficially grouped based on geographic proximity and property values. Though regions are designated a numerical value in this index, the numbers do not distinguish any ranking of price values and are arbitrary. Figure 3.4 shows the 13 unofficial regions, with Mdina marked with a white triangle. Mdina is classified with localities Mtarfa, Rabat, Mgarr and Dingli based on this unofficial classification. The quintiles assigned to these localities based on the preliminary index ranged from the 3rd quintile to least deprived in both the 2005 and 2011 index, therefore Mdina did not consistently align with the localities if it is assumed that they are similar.

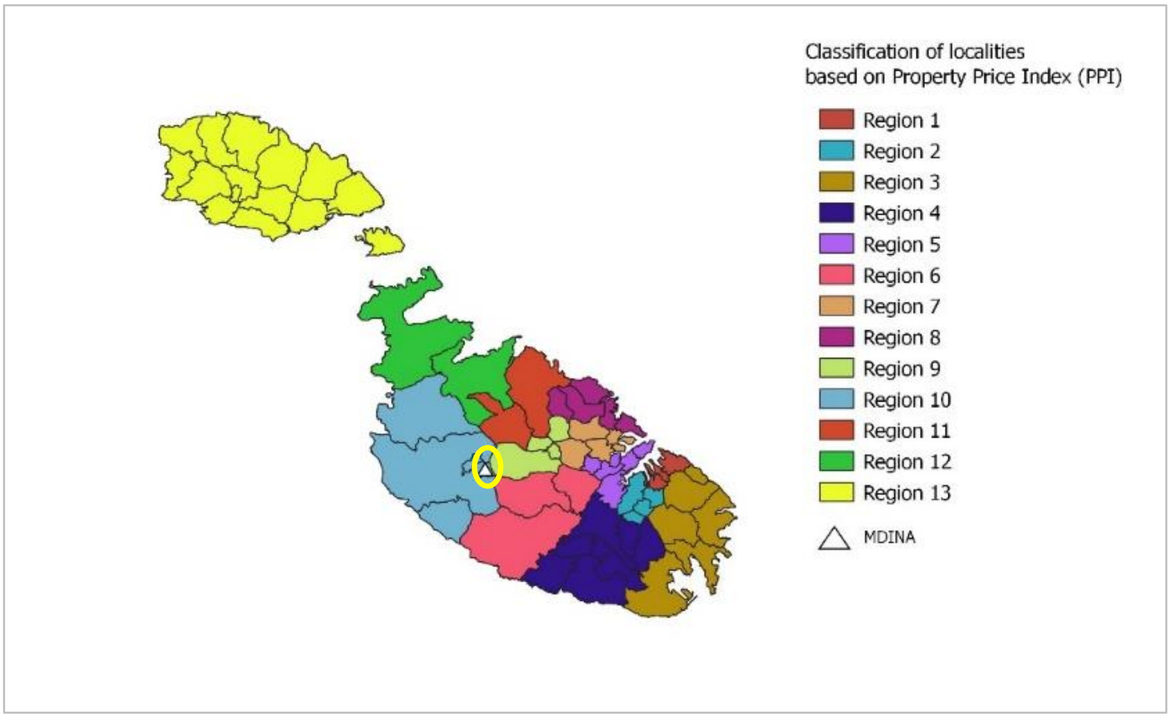


Figure 3.4: NSO unofficial classification of localities based on Property Price Index (PPI)

Given the historical background of the locality, the surrounding area, and the value of properties, it was considered inconsistent that Mdina would be classified in a high deprivation category and therefore required further testing to understand if any of the items included in the index were skewing the results. When reviewing the 2011 census report it was noted that Mdina had the highest proportion of dwellings (including vacant dwellings) which needed serious repair or were dilapidated, at 22.2% (National Statistics Office, 2014b). This relatively high proportion was most likely due to the historical nature of the buildings and the strict restrictions in place related to any demolition or construction in the area. The item % population living in dwellings needing serious repair/dilapidated (STATUS_REPAIR) was included in the computation of the deprivation score produced in the preliminary analysis.

To test whether this item was skewing the score for Mdina, the index was recalculated at the locality level for both years, excluding this item with updated factor loadings. When excluding the status of repair item, Mdina moved 2 quintiles in both years – from Most deprived (5) to quintile 3 in 2005 and from 4 to 2 in 2011. The new deprivation quintiles for each locality excluding % population living in dwellings needing serious repair/dilapidated (STATUS_REPAIR) were plotted against the proportion of persons living in dwellings needing serious repair/dilapidated for each locality. These are presented in figures 3.5 and

3.6, with Mdina highlighted in red. The graphs show that the value for Mdina (in red) is highest in both years and while in general the values for status repair are higher in the localities that are more deprived, this pattern is broken by Mdina where the value is disproportionately high when considering the deprivation quintile assigned when excluding this item.

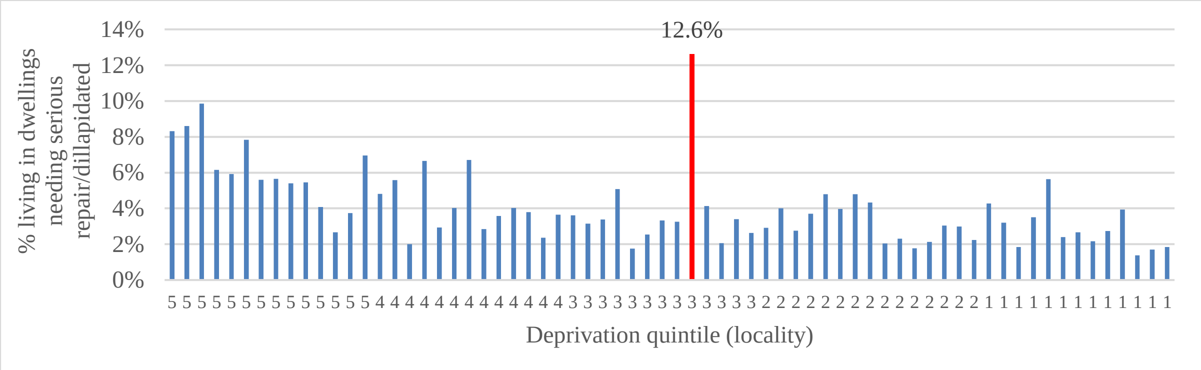


Figure 3.5: Deprivation quintile for the 68 localities against the percentage of the population living in dwellings needing serious repair/dilapidated: 2005

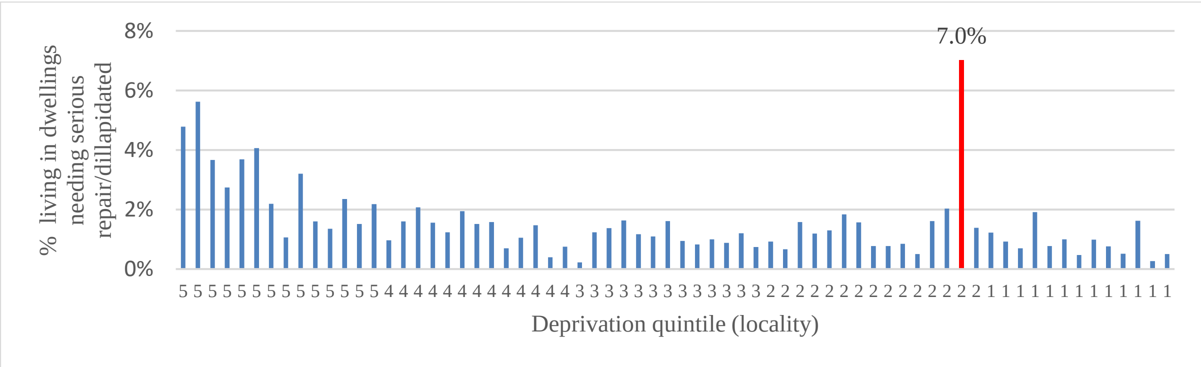


Figure 3.6: Deprivation quintile for the 68 localities against the percentage of the population living in dwellings needing serious repair/dilapidated: 2011

When comparing the updated index, with the preliminary index, apart from Mdina, 15 localities out of 68 in 2005 and 10 localities out of 68 in 2011 shifted quintile. The shifts, however, were only by one quintile, unlike what was seen for Mdina. Given the fact that the removal of the item considerably improved the index ranking for Mdina in both years and did not have significant impact on other localities, it was decided that the final index would be produced without the status of repair item. This was also deemed the most appropriate solution given that the item status of repair was the item with the lowest factor loading among all the items that loaded on the principal component for both census years. The alternative option was retaining all the items and excluding Mdina entirely, however since this meant not being able to assign a deprivation quintile to one of the localities, this was not a desired option.

3.4.3 Final index

The final index produced, following the sensitivity analysis described above, included five items – 2 within the domain of employment, 2 within the domain of living conditions and 1 within the domain of education, as presented in 3.10 below.

Table 3.10: Items included in final index

Domain	**Item**	**Computation**
Employment	Proportion of total working aged population who are unemployed	Number of unemployed aged 15 – 64/Total population aged 15 – 64
	Proportion of employed population in elementary occupations	Number of employed persons aged 15 – 64 in elementary occupations/ Total employed population aged 15 – 64
Education	Proportion of population aged 18 to 24 who are early school leavers	Number of persons 18 – 24 who obtained a secondary level of education or lower and were not students at the time of the census/ Total population aged 18 - 24
Living conditions	Proportion of population living in over-crowded occupied private dwellings	Number of persons living within over-crowded occupied private dwellings/ Total population living in occupied private dwellings
	Proportion of occupied dwellings not owner occupied	Number of occupied dwellings not owner occupied/ Total number of occupied private dwellings

A final PCA was conducted using the Z-scores for the five items listed above at the locality level to extract factor loadings to compute the final index at this geographical breakdown. Loadings and variance explained by the principal component for 2005 and 2011 using the locality level data are presented in table 3.11. Variance and factor loadings were greater at the locality level than the EA level (not shown).

Table 3.11: Results for final PCA using items at the locality level: 2005 and 2011

Z-Score of items (locality)	**2005 Factor loadings**	**2011 Factor loadings**
Proportion of population living in over-crowded occupied private dwellings	0.933	0.931
Proportion of population aged 18 to 24 who are early school leavers	0.920	0.947
Proportion of occupied dwellings not owner occupied	0.798	0.853
Proportion of total working aged population who are unemployed	0.869	0.920
Proportion of employed population in elementary occupations	0.825	0.846
Variance (%)	75.8	81.1

The final index scores for 2005 ranged from 12.8 to -5.44 at the EA level and 12.25 to -7.11 at the locality level. For 2011, the scores ranged from 14.53 to -5.4 at the EA level and 13.96 to -5.96 at the locality level. Figure 3.7 shows the scores for both indices by locality. Apart from a few localities, scores for the 2005 and 2011 index remained relatively consistent in terms of direction.

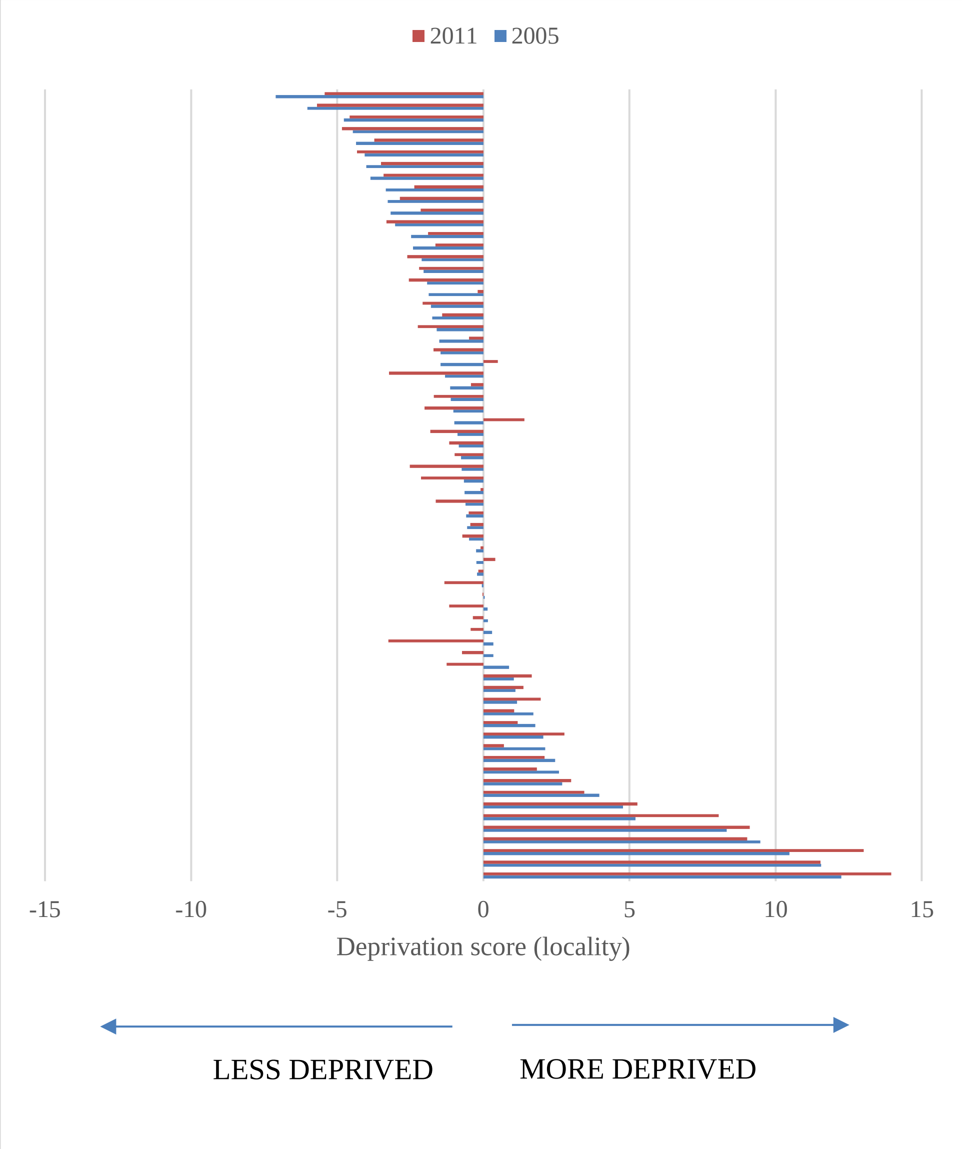


Figure 3.7: Deprivation scores from final index for each of the 68 localities: 2011 and 2005

Tables 3.12 and 3.13 show the misclassification of the EA quintile when compared to the higher-level locality quintile the EA was aggregated in. The row quintiles refer to the locality index while the column quintiles refer to the EA index. The proportions are calculated as a percentage of the row total, which is the total number of EAs within the respective locality quintiles. Cells highlighted in dark yellow indicate when the EA quintile category matches exactly with the locality quintile that it falls within.

As can be seen, the index at locality level shows more homogeneity with the EA index at the two extremes of deprivation. In 2005, 57% of EAs in localities categorised as least deprived were also categorised in this quintile using the EA index, the percentage is similar in

2011, at 58%. For the most deprived category, the homogeneity is greater. In 2005, 65% of EAs in localities categorised as most deprived were also categorised as most deprived using the EA index. The percentage is similar for 2011, at 67%. If one considers the percentage of EAs that fall within the same locality quintile or +/- 1 quintile, the results are generally good, ranging from 74% to 88% for the 2005 index and 78% to 98% for the 2011 index. Overall, the index at the locality level is more homogenous with respect to the EA in the most deprived areas and for 2011.

Table 3.12: Cross-tabulation of EA quintiles with locality quintiles: 2005

2005		EA quintile					% in same quintile or within +/- 1 quintile
		1	2	3	4	5	
Locality quintile	1	56.7%	19.6%	10.2%	4.4%	1.4%	76.3%
	2	20.7%	30.4%	22.6%	18.2%	3.8%	73.7%
	3	14.7%	25.5%	29.6%	24.9%	8.7%	79.9%
	4	6.0%	17.2%	29.0%	29.3%	21.2%	79.5%
	5	1.8%	7.4%	8.6%	23.2%	64.9%	88.1%

1: Least deprived; 5: Most deprived

Table 3.13: Cross-tabulation of EA quintiles with locality quintiles: 2011

2011		EA quintile					% in same quintile or within +/- 1 quintile
		1	2	3	4	5	
Locality quintile	1	57.8%	21.7%	10.7%	6.8%	1.6%	79.5%
	2	26.0%	32.5%	20.2%	14.3%	1.1%	78.7%
	3	9.2%	26.7%	34.5%	17.5%	6.8%	78.7%
	4	4.6%	15.8%	22.6%	31.5%	23.7%	77.8%
	5	2.3%	3.3%	11.9%	29.9%	66.8%	96.7%

1: Least deprived; 5: Most deprived

Figures 3.8 and 3.9 show the final deprivation quintile for each locality mapped to the local administrative base map for both census years. The maps show clear geographic clustering of deprivation quintiles in both census years, with most of the localities within quintile 5 (most deprived) and quintile 4 in proximity of each other and within the Southern Harbour District. The same can be said for localities classified within quintile 1 (least deprived) and quintile 2. In both years, there are no localities in Gozo classified in quintile 5 (most deprived).

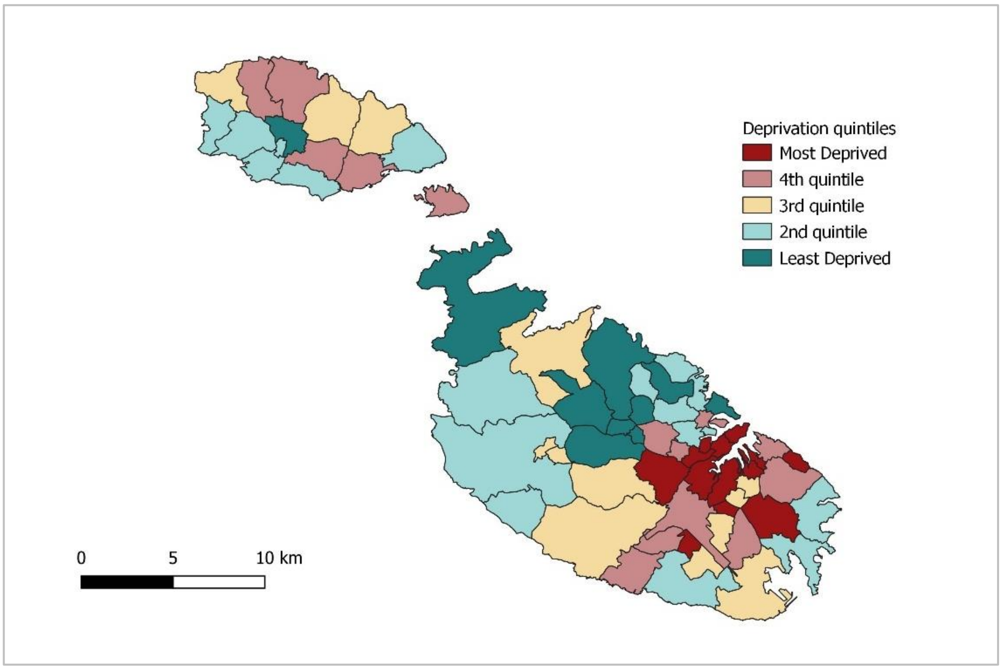


Figure 3.8: Final deprivation index quintiles for the 68 localities: 2005

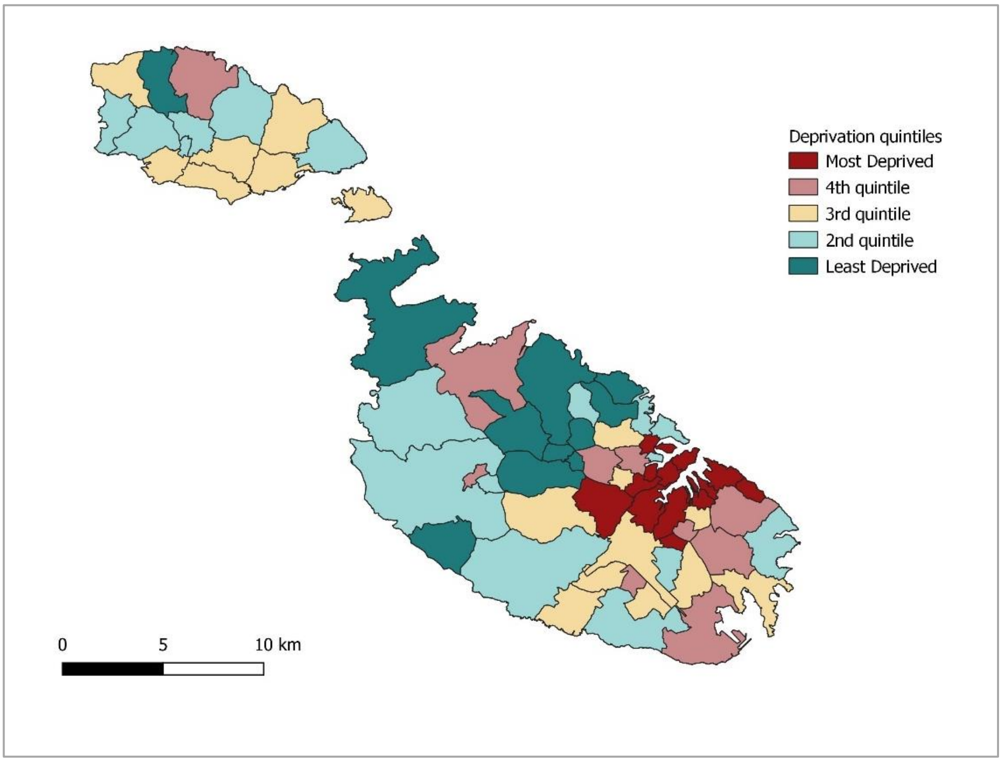


Figure 3.9: Final deprivation index quintiles for the 68 localities: 2011

When comparing localities across the two time points, 39 out of the 68 localities remained in the same quintile. The stability was not uniform across the quintiles. The least shifts were seen in quintile 5 (most deprived) and quintile 1 (least deprived) with 86% of localities classified as most deprived in 2005 remaining in this category in 2011. Similarly, 80% of localities classified as least deprived in 2005 remained in this category in 2011. There was also more stability in quintile 2 compared to the remaining two quintiles, with 61% of localities in this category in 2005 remaining there in 2011. Of the 29 localities that moved across quintiles, 13 became more deprived while 16 became less deprived.

3.4.4 Validation of the final index

The Cronbach's alpha values for the five items included in the final index at the locality level was 0.919 for 2005 and 0.941 for 2011, indicating a high level of internal consistency between them. The 19 items considered for possible inclusion in the index were compiled based on detailed review of the literature and other similar deprivation indices. The domains covered by the five items included in the final index – employment, education and living conditions – have been used widely by other authors when developing their indices – therefore they have consistently been considered able to measure some aspect of deprivation. While the notion of deprivation is relative, and there is no one clear definition of what it means to be deprived, the inclusion of domains and items consistently used by other researchers can be a proxy measure of good content validity.

Table 3.14 presents data for average disposable income per household, percentage at-risk-of-poverty and share of total district population in the unemployment register for 2011 based on data published by the NSO. Districts are ranked according to the weighted deprivation quintile estimated for them when using the 2011 index by locality. The Southern Harbour district has the highest values for the three indicators published by the NSO. This is consistent with the weighted average quintile from the 2011 district which is the highest value across all the six districts, indicating that it is the most deprived district based on the average quintile values for the localities within it.

There is less consistency between the individual indicators across districts and how they compare to the average deprivation quintile for the district. This is most likely due to the limitation of single items as a measure of deprivation, as well as the masking of pockets of deprivation when aggregating to large areas.

Table 3.14: District level estimates of three social, income and employment indicators published by the NSO and the average population weighted quintile for districts using the locality index: 2011

District	Average disposable income per household (€)[1]	% at risk of poverty[1]	% total population in unemployment register[2]	Deprivation quintile (population weighted average)
Southern Harbour	19,603	18.8	2.5%	4
Northern Harbour	21,066	17.6	1.5%	3
South Eastern	22,412	11.7	1.7%	3
Western	25,144	13.5	2.3%	3
Northern	21,655	14.7	0.9%	2
Gozo and Comino	20,169	12.6	1.1%	2

[1]Source: Statistics on Income and Living Conditions (SILC), NSO, [2]Source: Jobsplus

3.5 Discussion

This chapter presented the steps undertaken to develop and validate a small area-based deprivation index for Malta using census data. While several small area-based deprivation indices have been developed using other sources apart from the census, the census was considered the best source for this analysis given the current issues with quality and coverage of administrative registers in Malta. While data was available to the researcher at the enumeration area level, the final index was intended to be developed for the 68 localities. The choice of locality as a small area was a practical one. Enumeration areas are developed for operational use in the census and are updated each census round. Crucially, they are only used in the census and are not easily mapped to other external data sources. Since the aim of this thesis was to develop an index that could be used by researchers for inequalities analysis beyond the work presented here, locality was deemed the most appropriate option to maximise future usability of the index, as information on locality is generally available in health registers and administrative sources. While the final index was intended to be produced at the locality level, the enumeration area (EA) data was still used to select the items included in the final index and an index at the EA level was also constructed.

Based on a review of the literature and other indices developed using the census as presented in Chapter 2, nineteen (19) initial items were operationalised for possible inclusion in the index. To try and enhance, as much as possible, the stability of the index with time, especially for future updating, data from two census rounds were analysed – 2005 and 2011.

While the 2021 census was still ongoing at the time of the analysis and therefore data arising from this census could not be used to develop the index, the 2021 census questionnaire was reviewed, and ultimately certain decisions related to the final selection of items to include in the index were also based on what was potentially available from the 2021 census, once data would be available to researchers.

As discussed in Chapter 2, while some researchers have selected items for their index based on desk research or through the pooling of opinions on the relevance of items from external individuals, deemed as experts, a common method used to select and combine items for a deprivation index has been factor reduction, mainly principal component analysis (PCA) (Allik *et al.*, 2020). PCA was chosen as the method used to select and combine the items for the small area-based deprivation index described in this chapter. This choice was not only made because this method has been applied by others, but also because a data-driven method was deemed the most appropriate technique to use to attempt to develop an index that was reasonably objective. Also, the intention was to develop an index that was concise, containing a limited number of items that were deemed, when combined, to measure deprivation most effectively. The choice of the principal component emerging from the PCA and thus exclusion of items that did not load on this component, was expected to produce the most parsimonious index.

Sensitivity analysis was conducted to assess the impact of selected changes in the analytical process on the index output. While the item measuring status of repair loaded on the principal component, it led to unexpected deprivation scores for Mdina in both years. The item was excluded from the final index since it had the lowest factor loading and so as not to exclude Mdina entirely from the index score. The final index contained five items covering the domains of employment, education and living conditions. The factor loadings emerging from the final PCAs for 2005 and 2011 separately, were used to weight the items before they were summed to produce a deprivation score for each locality. While some indices presented in Chapter 2 were unweighted, weighting was deemed as appropriate in the calculation of the index created here since the loadings of individual items varied within the census PCA analysis and between the two PCA's conducted for 2005 and 2011. The overall variance explained from the first principal component for the final index at the locality level was 75.8% for the 2005 data and 81.1% for the 2011 data. This was generally high compared to the variances found in other indices which ranged from below 50% (Salmond, Crampton and Sutton, 1998; Lian *et al.*, 2011), to between 50 and 70% (Messer *et al.*, 2006; Havard *et al.*, 2008; Panczak *et al.*, 2012; Torres-Cintrón *et al.*, 2012; Z. Wang *et al.*, 2020).

As far as was possible, validation of the final index, apart from predictive validity, was conducted. The internal consistency measured through the Cronbach's alpha showed good consistency between the items and was comparable to those in other indices (Singh, 2003; Havard *et al.*, 2008; Lian *et al.*, 2011). The domains included in the index were also consistent with the common domains included in other similar indices as discussed in Chapter 2. While no external measures of deprivation were available at the locality level so that comparison could be made between the deprivation score and other measures, a weighted average quintile was estimated for the six districts for the locality index based on the 2011 census to compare to available indicators at the district level as published by the NSO.

It must be acknowledged that the analysis presented here does have some limitations. Firstly, the use of census data for the development of a deprivation index is limited by the fact that the census is only updated every ten years. While it is acknowledged that the census becomes outdated the further away one moves from the census year, it is currently the best source available for the development of such an index in Malta. Also, at the time of the analysis, the 2021 census was not available to the researcher. The 2021 census data will provide an update which is expected to align with the considerable demographic change experienced in Malta over the last ten years. Secondly, while the smallest area available from the census was the enumeration area, since this cannot be linked to other external sources, the final index had to be developed at the locality level. Notwithstanding this, when compared to the small area sizes used in other indices produced in larger countries, the locality sizes in Malta are still relatively small. Thirdly, while an attempt was made to include an iterative process to reduce the items included in the final index, the process still had some subjectivity and certain items had to be excluded because they will not be available in the 2021 census. Despite these limitations, to the author's knowledge, this is the first attempt in Malta to develop and validate a small area-based deprivation index using the census.

Chapter 4: Deprivation and mortality – testing for a health gradient.

4.1 Introduction

As outlined in Chapter 2, several small area-based census deprivation indices have been developed with the intention of being used for the analysis of health inequalities. Indeed, one of the aims for developing this small area-based deprivation index for Malta is so that it can be applied in health inequality analysis when individual socioeconomic variables are not available in health data sources.

The assessment of a gradient in risk of mortality by deprivation is a common health outcome used by researchers in the development and testing of small area-based census deprivation indices. Mortality risk can be analysed across different stratification levels, such as by sex or by age groups to assess, for example, risk of premature mortality. Age-standardised mortality rates (ASMRs) were chosen as the outcome measure to test whether the index created in this study predicts a gradient by deprivation quintile. ASMRs are a widely used outcome measure in public health because they are calculated using readily accessible mortality and population counts and are easy to understand and interpret. Furthermore, ASMRs, as opposed to crude mortality rates, consider the age structure of the population through standardisation, which allows for comparison across different categories such as geographical areas and time periods. The main analysis which will be presented and discussed in this chapter will be the testing for a gradient in all-cause mortality risk by deprivation when using the final deprivation index at the locality level. National mortality data can be stratified by locality, therefore total all-cause mortality ASMRs could be calculated for each deprivation quintile.

As discussed in Chapter 3, it must be acknowledged that a deprivation index developed at the locality level may mask pockets of deprivation found within smaller Enumeration Areas (EAs). Since EAs are created specifically for census operations and are different each census round, the deprivation index created at this level could not be easily mapped to auxiliary data sources. This is because EAs are not based on standardised boundaries like those used to outline official localities, postcode sectors or a geo-coded grid and are not used for any other spatial stratification apart from the census. Therefore, deprivation quintiles based on the EA index could not be linked directly to the data available within the mortality register, since the register only contains spatial information on locality for the records within it. However, knowing the limitation of using the locality index rather than the EA index for analysis, it was ideal to find a way to analyse a mortality output using deprivation quintiles based on both the locality and EA indices to compare the two outputs. Apart from locality, the census data files contain the EA code and therefore the EA deprivation quintile can be assigned to each record in this data. The census data files also contain a unique identifier field that allows the data to be linked directly to the mortality register. This allowed for the linking of the census data file, containing the EA and locality deprivation quintiles, to the mortality register, such that records were assigned mortality status within five years from the respective censuses. These mortality counts were used to calculate ASMRs by deprivation quintiles based on the EA and locality indices separately and were compared.

The rest of the chapter will present the outcome of the two analyses described above. Section 4.2 will describe the data sources used to estimate the respective ASMRs. Section 4.3 will describe the methods used and results will be presented in section 4.4. Finally, section 4.5 will discuss the findings presented in this chapter.

4.2 Data source

To estimate ASMRs, numerator counts of mortality and denominator counts of mid-year population were required. It is also useful to produce rates for males and females separately, since mortality risk varies across the sexes, therefore numerator and denominator counts can be further subdivided by sex. Age-standardisation of a mortality rate also requires that a standard population be used. The European Standard Population (ESP) was chosen as the population for standardisation as it is the standard population that most closely resembles the age structure for Malta. The ESP is a theoretical population adding up to a total of 100,000. The ESP was originally introduced in 1976 but was updated in 2013 to take into consideration ageing populations across Europe. The updated standard population was agreed upon following a consultation process with EU Member States. The ESP2013 was developed based on the average of Member State population projections for 2011-2030. Since Malta is also an EU Member State and therefore included in the average, this is the ESP which was considered most appropriate for this analysis. Table 4.1 shows the age distribution of the ESP. As can be seen, there are no differences in age structure between the sexes.

Table 4.1: European standard population 2013

Age group	Male	Female
0 – 4	5,000	5,000
5 – 9	5,500	5,500
10 – 14	5,500	5,500
15 – 19	5,500	5,500
20 – 24	6,000	6,000
25 – 29	6,000	6,000
30 – 34	6,500	6,500
35 – 39	7,000	7,000
40 – 44	7,000	7,000
45 – 49	7,000	7,000
50 – 54	7,000	7,000
55 – 59	6,500	6,500
60 – 64	6,000	6,000
65 – 69	5,500	5,500
70 – 74	5,000	5,000
75 – 79	4,000	4,000
80 – 84	2,500	2,500
85 – 89	1,500	1,500
90 – 94	800	800
95+	200	200
Total	100,000	100,000

The remainder of this section is divided into two – 4.2.1 explains the data sources used to estimate the ASMRs using the census linked mortality data, while section 4.2.2 presents the data sources used to estimate the ASMRs for all-cause mortality using the final index by locality. Data was not readily available and therefore provided by the National Statistics Office (NSO) upon request. Only aggregate counts were requested, and no ethical approval was required by the NSO to process the data request. Data was provided to the researcher in March 2022.

4.2.1 Census linked data for ASMRs

Mortality within 5 years of the respective censuses was taken as a proxy for total all-cause mortality at the time of the census and was used for the estimation of the ASMRs using the EA index. The individual record census files the NSO used to produce the aggregate indicators which were provided to the researcher to construct the index in Chapter 2, were linked to the mortality database for 2005- 2010, for census year 2005, and 2011-2016, for census year 2011.

Record linking between the censuses and mortality register was conducted using the unique identifier available in both sources – this is a unique identity number given to all Maltese citizens and foreign nationals who have a residence card. There were 0.5% and 0.4% of records in the 2005 and 2011 censuses which did not have an identity card number, in such cases a combination of name, surname and date of birth were used to try to link to the mortality register data. Given the low percentage of missing identifying numbers and the linking by a combination of other fields, it is assumed that the linkage between the two files was accurate. Based on the data linkage, an aggregate table with mortality counts was produced by the NSO for each census year separately, broken down by EA code, sex and 5-year age groups. The age-group provided referred to the age at death. Since the census file used for the creation of the index was used as the basis for linking to the mortality data, the deaths considered in this analysis are restricted to persons who were resident in private households at the time of the census and were not assigned to a dummy EA.

To produce a mortality rate, denominator population data is required. Therefore, the population counts for each census year by EA, sex, and 5-year age groups were also requested. The data provided related to mortality and population counts at the EA level could be summed up to the locality level based on the EA codes provided in the aggregate data files. This allowed for the estimation of ASMRs using the locality and EA deprivation index quintiles, thus direct comparison of the gradient when using deprivation quintiles calculated at different small area levels could be conducted.

One must acknowledge the limitation of using the census population as a denominator since this means the base population was kept static. Using the population data as at census means that demographic shifts occurring with time were not taken into consideration, including shifts in population structure due to aging and emigration. While the mortality counts were aged forward based on the year of death, the base population was not. Even though 5-year age groups were used, this potentially may have led to persons being counted in a different age group in the base compared to the numerator. Furthermore, persons who emigrated from the population after the census were no longer at risk of dying and therefore not excluding them could have led to an inflation of the denominator count. Between 2006-2010 an estimated 19,353 emigrants left Malta, equivalent to 5% of the enumerated population in 2005 considered for this analysis. Between 2012-2016, an estimated 29,289 emigrants left Malta, equivalent to 7% of the enumerated population in 2011 considered for this analysis. Of course, not all the emigrants within this period may have been present in the country at the time of the censuses and therefore would not be included in the enumerated populations considered here. In both time periods, slightly over 85% of the emigrants were of working age (15-64) at the time of emigration. Emigration estimates are available for the total population, therefore no information is available on locality prior to emigration. There is no way of knowing if there is a pattern of emigration by deprivation and thus the possible impact this may have had on the estimated mortality rates.

While acknowledging the possible impact of all the limitations outlined above, this analysis was conducted to assess whether there were differences in the health gradient when using the EA versus the locality index. Any possible error from this approach is assumed to have been consistent across the estimates emerging from the two indices and is not expected to have impacted upon the overall aim and the ultimate comparisons conducted.

4.2.2 All-cause mortality ASMRs

Annual mortality counts in Malta for the two census years were each below 3,500, owing to the small population size of the country. To ensure adequate numbers when breaking down the aggregates by deprivation quintile, the decision was made to analyse all-cause ASMRs using mortality counts for a five-year period rather than a single year. Mortality counts were requested broken down by sex, locality and 5-year age groups (age at death) for the 5 years around the census, that is 2003-2007 for 2005, and 2009-2013 for 2011.

Population estimates by locality produced by the NSO are not presently disseminated disaggregated by age – only total counts disaggregated by sex are currently available at the locality level. To be able to estimate the standardised rates described above, a denominator population with similar breakdowns as the mortality data, was required. Locality data disaggregated by age and sex were available from the census. It was decided that this distribution could be used to approximate the distribution by age for the mid-year population estimates available annually, by locality. Mid-year population estimates for each year from 2003-2007 and 2009-2013 by sex and locality, were provided by the NSO and proportional distributions for each locality by sex and age from the respective census rounds were used to approximate the distribution of the counts by age for the annual data. The method does assume that the age distribution by locality for the years around the censuses has remained fixed and this may not be applicable for all years or all localities. However, since the distributions used were 5-year age groups and not single years of age, the differences over time should have had less impact on the age groupings.

When comparing the average age for the localities based on the 2005 census and the 2011 census (Appendix D), all localities aged during the inter-censal period. The largest increase in average age between the two time points was 5.4 years for Mdina, while the smallest increase was and 0.5 years for Għasri. Given that these changes occurred over 6 years, it is likely that any shifts in the distribution across the 5-year age groups for the years around the census were less significant.

4.3 Methods

For both analyses discussed in this chapter, the direct standardisation method was used to estimate the ASMRs. The method involves the application of age and sex specific mortality rates for each deprivation quintile to the age and sex distribution of the reference population, in this case, the ESP, described previously. Direct standardisation ensures that the mortality rate is independent of differences in the age and sex distribution between populations. This allows for comparison of mortality risk across the different quintiles when eliminating the impact of possible age and sex structure differences in the different geographical areas and time points.

To quantify the relative socioeconomic gradient, if any, a summary measure of the linear association between deprivation and mortality was estimated. The Relative Index of Inequality (RII) was used to measure the inequality gradient in mortality. The RII is like a measure of range, in that it considers the difference between the lowest and highest group in the data being analysed, but it also takes into consideration the values for all the groups and the population size of each group. For this reason, the RII assumes the data is ranked which makes the measure appropriate for the analysis of deprivation quintiles which are ranked from least to most deprived. The RII is estimated by first estimating the Slope Index of Inequality (SII) which is calculated by fitting a linear regression to the ranked outcome measure being analysed – in this case the age-standardised rates for the five deprivation quintiles. The RII is then produced by dividing the SII by the overall value of the outcome measure for the entire population to standardise the gradient relative to the average measure of the outcome. The RII can then be expressed as a percentage with the value indicating how the most deprived group differs, in percentage terms, from the average value of the population. A 95% confidence interval around the value was also estimated using the method proposed by Sonja Lumme and colleagues (Lumme *et al.*, 2015). This method was adopted instead of the traditional 95% confidence intervals from the regression equation, since these do not consider that the observations being analysed are ranked.

All analysis was conducted in RStudio 2022.02.1 using the SocEpi package (Allik, 2018). Additional details on the estimation of the ASMRs for the two different analyses will be presented in separate sections. Section 4.3.1 explains the method used to estimate the ASMRs using the two deprivation indices developed at the locality and EA levels and the census linked mortality data, while section 4.3.2 presents the methods used to estimate the ASMRs for all-cause mortality using the final index developed at the locality. Table 4.2 provides a summary of the differences between the two analyses.

Table 4.2: Summary of the differences between the two ASMR analyses

	Mortality five years from census	**All-cause mortality**
Population denominator	Census population resident in private households who have an assigned EA.	Total mid-year resident population in Malta in reference periods 2003-2007 and 2009-2013
Mortality numerator	Deaths occurring five years after census day. Extracted from linking denominator data to mortality register.	Total resident deaths in Malta in reference periods 2003-2007 and 2009-2013
Small area deprivation index	EA and locality indices	Locality index
Outcome measure	ESP age-standardised mortality rates	ESP age-standardised mortality rates
Measure of effect	Relative Index of Inequality	Relative Index of Inequality
Aim of analysis	Comparison of performance of two deprivation indices developed at different small area levels.	Health inequality analysis of mortality risk and area-based deprivation.

EA: Enumeration Area, ESP: European Standard Population

4.3.1 Comparing index performance at two small area levels

The mortality and population data from the linked census dataset grouped at the EA and locality level which were described in section 4.2.1, were assigned the respective deprivation quintile value depending on the applicable census year. The mortality and population counts were than summed, by quintile, for each period 2003-2007 and 2009-2013 separately for the EA and locality indices.

As described in section 4.2.1, while the population bases from the censuses were static and therefore not aged five years ahead, the linked deaths five years from the census were based on age at death and therefore subject to ageing. This led to the 95+ age group having a higher count in the numerator than the population denominator. While the original ESP standard population is truncated at 95+, it was decided to truncate it to 80+ along with the mortality and population data to address this issue. While this truncation may have impacted the estimated ASMR, since the risk of dying could vary across the 5-year age bands between 80 and 95, the aim of this analysis was to assess the presence of a health gradient by deprivation and to compare this gradient when using deprivation scores estimated at different small areas, rather than specifically to analyse the estimated ASMR's.

Working with the numerator and denominator data which were grouped separately based on the quintiles assigned from the EA and locality indices, ESP ASMRs were estimated for each deprivation quintile for the total, and males and females separately for the two time periods. ESP age-standardised rates were estimated per 100,000-person year at risk. For each ASMR, 95% confidence intervals were computed as $r \pm 1.96*(r/\sqrt{h})$ where r is the standardised rate and h is the number of events, in this case, mortality counts. To compare the gradient in mortality risk by quintile when using the two different indices, the RII was estimated as described previously, and compared.

4.3.2 Application to all-cause mortality

The deprivation quintile score from the locality index was assigned to the aggregate all-cause mortality counts and mid-year population data described in section 4.2.2, using the locality code in the aggregate tables. The counts were then summed up into the five deprivation quintiles leading to the creation of a new aggregate dataset containing population and mortality counts broken down by sex, 5-year age group, deprivation quintile and time-period (2003 –

2007 and 2009 – 2013). For each time-period, ESP age-standardised mortality rates were estimated for each quintile for the total, and males and females separately. Furthermore, estimates were produced separately for the age group 0 – 69 and 70+ to assess the impact of deprivation on premature mortality. The age cut-off used to define premature mortality was in line with the cut off applied by the World Health Organisation (WHO) for the estimation of premature mortality from non-communicable diseases (World Health Organisation, 2023) and the potential years of life lost (PYLL) indicator estimated by the Organisation for Economic Co-operation and Development (OECD) (OECD, 2021). For each ASMR, 95% confidence intervals were computed as described in section 4.3.1. ESP age-standardised rates were estimated per 100,000-person year at risk. The RII was estimated as described previously, to assess whether a gradient in mortality risk exists by deprivation. Additional analysis and testing were conducted, if required, depending on the results from the first estimation of all cause ASMRs.

4.4 Results

4.4.1 Comparing index performance at two small area levels

The total population considered in this section is the same population used to build the indices, that is, those enumerated in the 2005 and 2011 census and living in private households, excluding those assigned to a dummy EA. As was described in Chapter 3, this is equal to 389,656 in 2005 and 399,901 in 2011. For both census years, 3.4% of the population died within 5 years of their enumeration. There were slightly more deaths amongst males in both years (54.5% in 2005; 54.1% in 2011), and the majority of those who died were aged 60 and over at the time of death (87% in 2005; 89% in 2011) (Table 4.3).

Table 4.3: Mortality counts within 5 years of the 2005 and 2011 censuses by age and sex

Age at death	2005-2010			2011-2016		
	Males	**Females**	**Total**	**Males**	**Females**	**Total**
0 – 9	6	8	14	8	12	20
10 – 14	10	10	20	6	5	11
15 – 19	25	7	32	13	10	23
20 – 24	47	11	58	22	17	39
25 – 29	41	17	58	27	12	39
30 – 34	51	15	66	49	18	67
35 – 39	55	25	80	55	26	81
40 – 44	91	34	125	70	45	115
45 – 49	114	99	213	105	62	167
50 – 54	247	142	389	211	130	341
55 – 59	402	249	651	378	238	616
60 – 64	630	407	1,037	619	359	978
65 – 69	700	414	1,114	917	561	1,478
70 – 74	983	686	1,669	903	605	1,508
75 – 79	1,200	984	2,184	1,172	864	2,036
80 – 84	1,213	1,225	2,438	1,235	1,276	2,511
85 – 89	909	1,017	1,926	1,029	1,195	2,224
90 – 94	345	448	793	500	651	1,151
95+	77	160	237	125	235	360
Total	**7,146**	**5,958**	**13,104**	**7,444**	**6,321**	**13,765**

Table 4.4 presents the mortality counts broken down by the assigned deprivation quintile based on the EA and locality indices. Across the two periods and when using either of the small area-based indices, the largest share of deaths within five years of the census was among those living in areas classified as most deprived. On the other hand, the smallest share of deaths was in the least deprived quintile for both periods except when using the locality index for the period 2005-2010. The smallest share of deaths based on this index was in the areas within the 3rd quintile category, at 16.3%.

Table 4.4: Mortality counts 5 years from the 2005 and 2011 censuses broken down by deprivation quintiles: EA and locality index

	2005-2010		2011-2016	
	EA index	Locality index	EA index	Locality index
Total	**13,104 (100%)**		**13,765 (100%)**	
Least deprived	1,833 (14.0%)	2,441 (18.6%)	1,901 (13.8%)	1,759 (12.8%)
2	2,168 (16.5%)	2,408 (18.4%)	2,290 (16.6%)	2,958 (21.5%)
3	2,471 (18.9%)	2,131 (16.3%)	2,643 (19.2%)	2,459 (17.9%)
4	2,788 (21.3%)	2,476 (18.9%)	3,041 (22.1%)	3,077 (22.4%)
Most deprived	3,844 (29.3%)	3,648 (27.8%)	3,890 (28.3%)	3,512 (25.5%)

EA: Enumeration Area

The estimated ESP ASMRs are presented in table 4.5. The age-standardised mortality rates for deaths five years from the census declined between the two censuses for the total population and when analysing estimates by sex, though the ASMRs for males were higher than those for females in both periods. The total estimated ESP age-standardised mortality rate for deaths occurring five years after the 2005 census was 1,075 per 100,000 (95% CI 1,060 - 1,089), reducing to 948 per 100,000 (95% CI 935 - 961) for those occurring five years after the 2011 census. Declines were also seen across all deprivation quintiles when comparing the total and sex specific rates and when using either the EA or locality index. When analysing the results by deprivation, the estimated ASMRs for the total and males, were highest in the most deprived quintiles when using either index. In all cases, the estimated ASMR was higher when using the EA index compared to the locality index. Results were less consistent for females. In the earlier period, based on the EA index, the highest ASMR remained in the most deprived quintile while when using the locality index, it was in the 3rd quintile. For the more recent period, when using the EA index, the highest ASMR was in the 4th quintile, while when using the locality index the highest ASMR was in both the 4th and most deprived quintile. While in the earlier period the ASMR for the most deprived quintile using the EA index was higher than that for the locality index, this was not the case in the more recent period.

Table 4.5: ESP ASMRs for deaths 5 years from census by sex, deprivation quintile, and period of death: EA index and locality index

Deprivation quintile		2005-2010		2011-2016	
		ESP ASMR per 100,000 (95% CI)		*ESP ASMR per 100,000 (95% CI)*	
		EA index	**Locality index**	**EA index**	**Locality index**
T	**Least deprived**	955 (918 - 991)	968 (937 - 999)	830 (799 - 861)	887 (852 - 921)
	2	1,024 (989 - 1,058)	1,044 (1,011 - 1,077)	914 (884 - 944)	861 (836 - 887)
	3	1,077 (1,043 - 1,111)	1,097 (1,060 - 1,134)	946 (917 - 976)	970 (941 - 1,000)
	4	1,086 (1,054 - 1,118)	1,093 (1,059 - 1,128)	981 (953 - 1,009)	997 (969 - 1,026)
	Most deprived	1,174 (1,145 - 1,203)	1,161 (1,132 - 1,191)	1,031 (1,005 - 1,057)	1,026 (999 - 1,053)
	Total	**1,075 (1,060 - 1,089)**		**948 (935 - 961)**	
M	**Least deprived**	1,196 (1,138 - 1,253)	1,204 (1,154 - 1,254)	1,017 (968 - 1,066)	1,032 (979 - 1,086)
	2	1,210 (1,156 - 1,264)	1,256 (1,204 - 1,309)	1,095 (1,048 - 1,143)	1,064 (1,022 - 1,106)
	3	1,290 (1,237 - 1,343)	1,287 (1,230 - 1,345)	1,153 (1,107 - 1,199)	1,184 (1,138 - 1,230)
	4	1,359 (1,308 - 1,410)	1,350 (1,295 - 1,405)	1,170 (1,126 - 1,215)	1,213 (1,168 - 1,258)
	Most deprived	1,493 (1,446 - 1,540)	1,488 (1,441 - 1,536)	1,333 (1,290 – 1,375)	1,310 (1,266 - 1,354)
	Total	**1,324 (1,301 - 1,348)**		**1,166 (1,145 - 1,186)**	
F	**Least deprived**	767 (722 - 812)	797 (759 - 835)	678 (639 - 716)	763 (719 - 807)
	2	873 (829 - 916)	871 (829 - 913)	767 (729 - 804)	711 (680 - 742)
	3	914 (872 - 956)	942 (895 - 989)	785 (749 - 822)	800 (763 - 838)
	4	879 (840 - 918)	895 (853 - 938)	841 (806 - 876)	826 (791 - 861)
	Most deprived	939 (903 - 974)	921 (885 - 957)	809 (777 - 840)	821 (788 - 855)
	Total	**883 (864 - 901)**		**781 (765 - 797)**	

ESP: European Standard Population; ASMR: Age-standardised Mortality Rate; CI: Confidence Interval; EA: Enumeration Area; T: Total; M: Male; F: Female

Figures 4.1 and 4.2 show the estimates in table 4.5 as linear plots. Reviewing the estimated ASMRs shows good comparability between the two indices, with the estimates overlapping considerably both at the point level and the 95% confidence intervals. Visual inspection of the linear trend lines between the points shows a generally upward gradient between the least and most deprived quintiles for all estimates in the two time points, with steeper gradients observed for males and the total, than for females. The gradients for the estimates based on the EA index appear steeper than those using the locality index. This is however expected since an index developed using smaller area sizes should be better able to differentiate between low and high deprivation patterns of mortality risk as it is more likely to capture pockets of deprivation in larger areas.

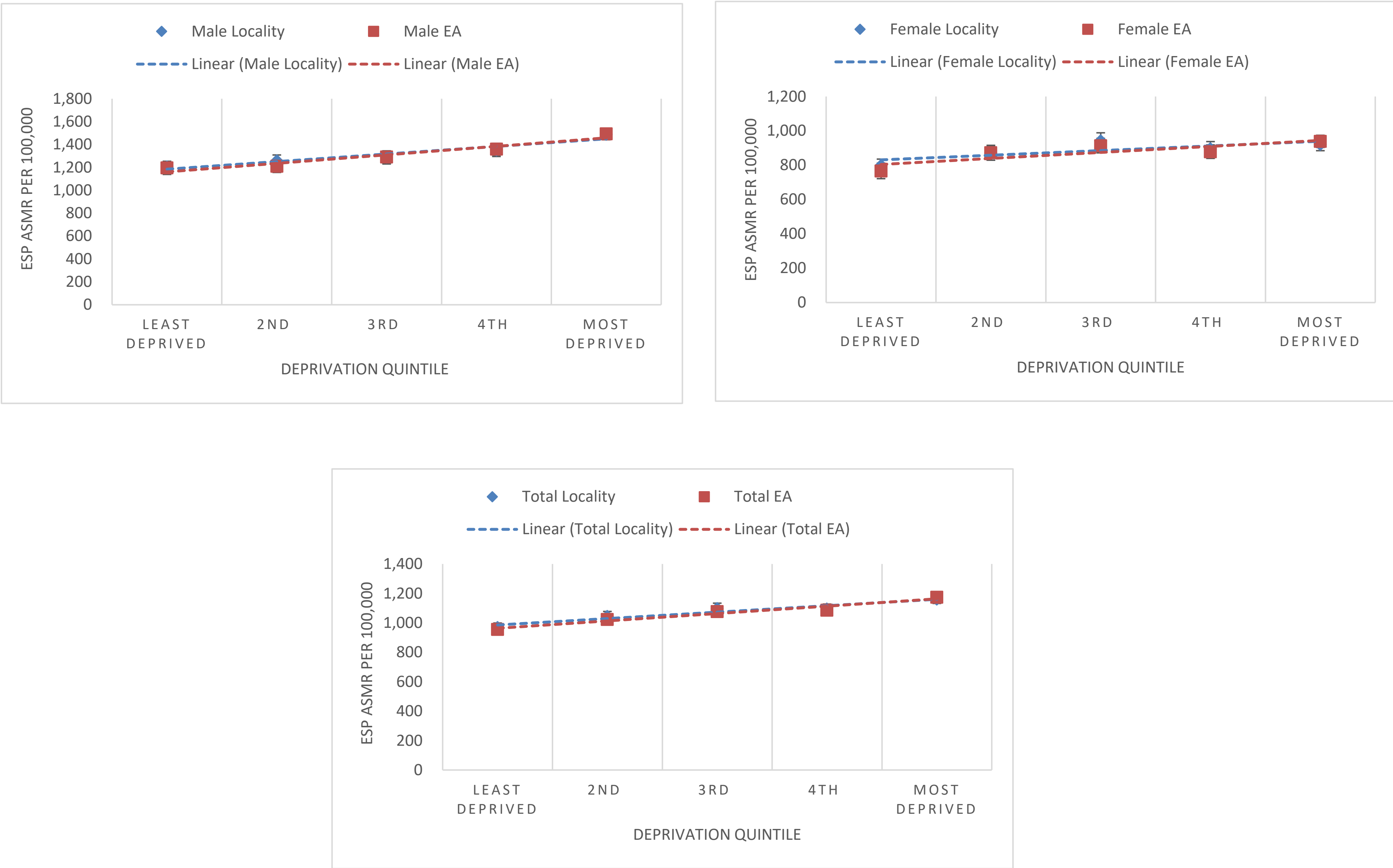


Figure 4.1: ESP ASMR's for the total and by sex for the five deprivation quintiles using EA and locality index: deaths five years from 2005 census

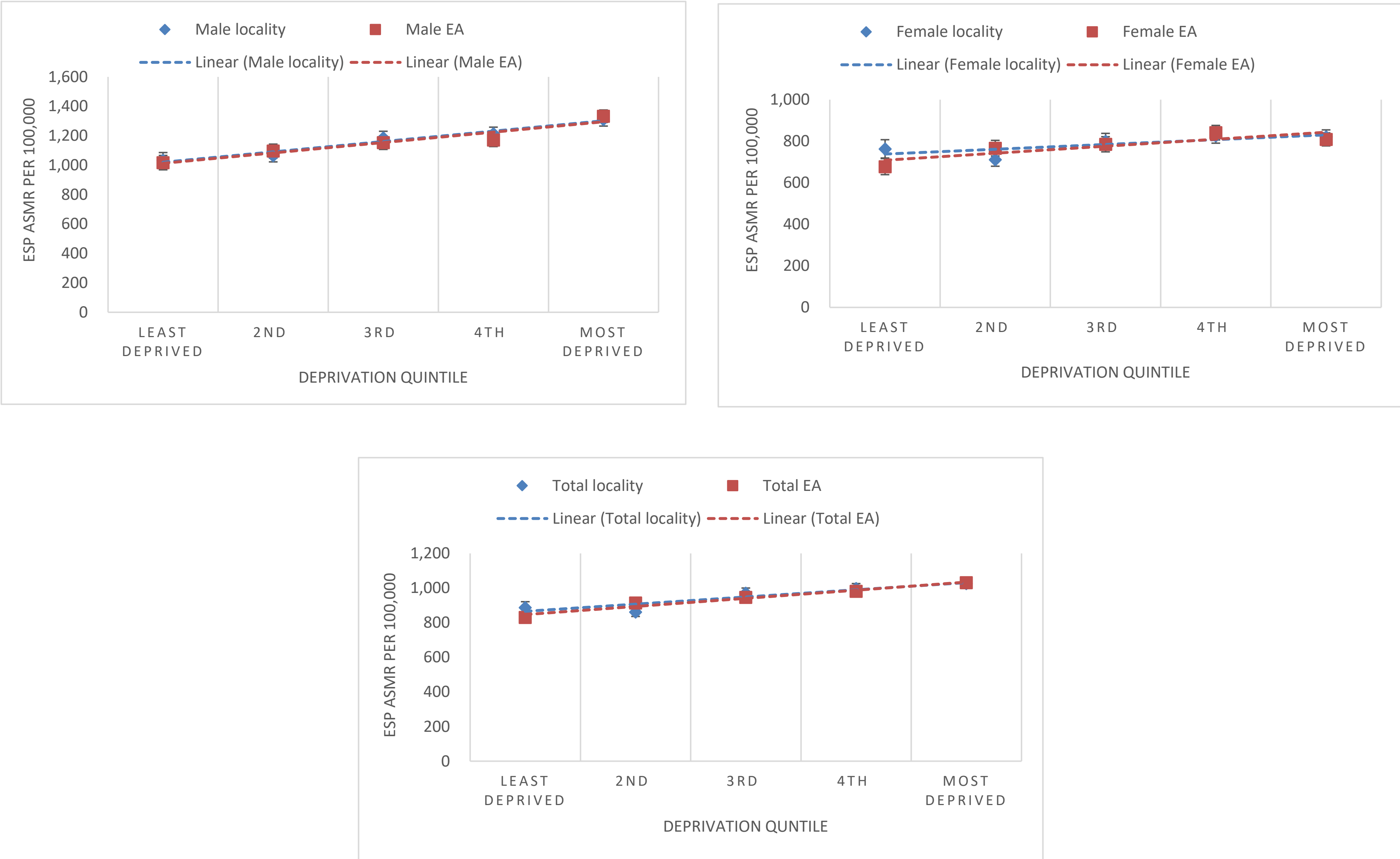


Figure 4.2: ESP ASMRs for the total and by sex for the five deprivation quintiles using EA and locality index: deaths five years from the 2011 census

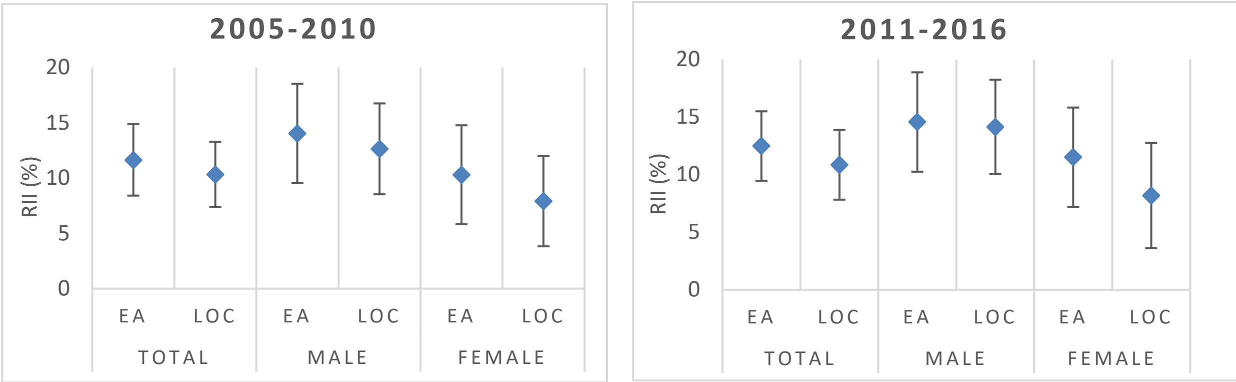


Figure 4.3: RII (%) estimates for the ASMRs for the total and by sex using the EA and locality index: deaths within five years from 2005 and 2011 census

Figure 4.3 shows the RII estimates for the ASMRs with corresponding 95% confidence intervals for the two periods for the total and by sex. Estimates have been calculated using the deprivation quintile ASMRs based on both indices and are presented as percentages. Comparing across the two time periods, the RII values have remained relatively consistent. When comparing between the two indices, the RII values using the EA index are slightly higher than those estimated using the locality index, which is consistent with the visual inspection of the linear plots. The RII values extracted from the EA index estimates do however fall within the confidence intervals of those extracted from the locality index, therefore it can be assumed that they are generally comparable.

Interpreting the RII values shows a significant linear gradient in mortality risk by deprivation when using either index. When assessing the values estimated using the EA ASMRs, for the total population, those in the most deprived areas had a 12% greater risk of dying within 5-years of the census when compared to the average risk of the population (2005: 11.6% 95% CI 8.4% - 14.8%; 2011: 12.5% 95% CI 9.5% - 15.5%). Slightly higher RII values were estimated for males when compared to females. Males living in the most deprived areas had an approximately 14% greater risk of dying within 5-years of the census when compared to the average risk of the population (2005: 14.0% 95% CI 9.5% - 18.7%; 2011: 14.6% 95% CI 10.3% - 18.6%). For females the increased risk was 10.3% in 2005 (95% CI 5.8% - 14.8%) and 11.5% in 2011 (95% CI 7.2% - 16.4%).

4.4.2 Application to all-cause mortality

There were a total of 15,602 all-cause deaths between 2003-2007 and 16,152 between 2009-2013. In both periods, there were slightly more deaths amongst males (50.8%) and the majority (73%) of those who died were aged 70 and over at the time of death. When assessing the breakdown of the deaths by deprivation quintile, the largest share of deaths in both periods was in the most deprived quintile, at 26.8% in 2003-2007 and 24.4% in 2009-2013. The same pattern is observed when comparing by sex (Table 4.6).

Table 4.6: All-cause mortality counts by age groups, sex, and deprivation quintile: 2003-2007 & 2009-2013

Age at death	**2003-2007**			**2009-2013**		
	Male	**Female**	**Total**	**Male**	**Female**	**Total**
0-9	74	63	137	70	70	140
10-14	12	10	22	6	8	14
15-19	36	9	45	24	9	33
20-24	59	17	76	44	14	58
25-29	50	13	63	45	28	73
30-34	51	26	77	63	22	85
35-39	53	35	88	74	23	97
40-44	106	52	158	86	45	131
45-49	136	114	250	136	79	215
50-54	282	161	443	243	163	406
55-59	429	309	738	402	213	615
60-64	532	339	871	703	425	1,128
65-69	792	500	1,292	835	489	1,324
70-74	1,135	809	1,944	955	641	1,596
75-79	1,285	1,201	2,486	1,296	1,164	2,460
80-84	1,425	1,595	3,020	1,353	1,576	2,929
85-89	904	1,295	2,199	1,198	1,652	2,850
90-94	449	794	1,243	535	955	1,490
95+	113	337	450	139	369	508
Total	**7,923**	**7,679**	**15,602**	**8,207**	**7,945**	**16,152**
Deprivation quintile	**Male**	**Female**	**Total**	**Male**	**Female**	**Total**
Least deprived	1,493	1,611	3,104	1,085	1,146	2,231
2	1,358	1,376	2,734	1,743	1,801	3,544
3	1,337	1,126	2,463	1,464	1,370	2,834
4	1,624	1,502	3,126	1,851	1,744	3,595
Most deprived	2,111	2,064	4,175	2,064	1,884	3,948
Total	**7,923**	**7,679**	**15,602**	**8,207**	**7,945**	**16,152**

Table 4.7 presents the estimated ASMRs for the two time periods. Inspection of the total ASMRs shows that, while a general linear gradient is present with increasing deprivation, estimates specifically for the 3rd quintile in both periods are outside the linear trend, with a peak for the period 2003-2007 and a dip in 2009-2013. When assessing the estimates separately for premature mortality (0-69) and old age mortality (70+), the pattern outlined above may be specifically related to old-age mortality, therefore estimates for the two age groups will be analysed separately.

Table 4.7: ESP ASMRs for all-cause mortality by sex, deprivation quintile, and period of death: 2003-2007 & 2009-2013

	Deprivation quintile	2003-2007 ESP ASMR per 100,000 (95% CI)			2009-2013 ESP ASMR per 100,000 (95% CI)		
		Total	0-69	70+	Total	0-69	70+
T	Least deprived	915 (884-947)	234 (217-250)	5,101 (4,901-5,302)	889 (852-925)	183 (168-197)	5,226 (4,982-5,469)
	2	956 (921-991)	226 (209-242)	5,440 (5,211-5,669)	951 (921-981)	204 (190-218)	5,540 (5,342-5,738)
	3	1,250 (1,199-1,301)	262 (243-281)	7,320 (6,977-7,663)	942 (908-977)	204 (190-219)	5,477 (5,248-5,706)
	4	999 (964-1,033)	269 (251-287)	5,486 (5,266-5,706)	1,112 (1,076-1,149)	246 (231-261)	6,434 (6,190-6,678)
	Most deprived	1,079 (1,045-1,112)	286 (270-303)	5,946 (5,729-6,163)	1,158 (1,123-1,194)	278 (261-295)	6,569 (6,338-6,799)
	Total	**1,012 (996-1,028)**	**256 (248-264)**	**5,657 (5,556-5,759)**	**1,009 (994-1,025)**	**224 (217-230)**	**5,835 (5,734-5,936)**
M	Least deprived	882 (838-927)	294 (267-320)	4,498 (4,225-4,770)	1,112 (1,041-1,182)	227 (204-250)	6,545 (6,063-7,028)
	2	924 (875-973)	275 (249-301)	4,912 (4,601-5,222)	1,137 (1,084-1,189)	258 (235-280)	6,536 (6,185-6,888)
	3	1,366 (1,289-1,443)	327 (297-357)	7,745 (7,224-8,266)	1,158 (1,097-1,219)	257 (234-280)	6,689 (6,278-7,101)
	4	1,019 (970-1,068)	338 (310-367)	5,198 (4,894-5,502)	1,316 (1,252-1,379)	309 (286-332)	7,499 (7,070-7,928)
	Most deprived	1,057 (1,010-1,103)	349 (322-375)	5,406 (5,114-5,698)	1,438 (1,375-1,502)	372 (344-400)	7,989 (7,569-8,409)
	Total	**1,006 (984-1,028)**	**317 (305-329)**	**5,239 (5,100-5,379)**	**1,231 (1,204-1,258)**	**286 (275-297)**	**7,036 (6,852-7,219)**
F	Least deprived	945 (900-989)	175 (155-195)	5,671 (5,380-5,961)	760 (718-803)	139 (121-157)	4,577 (4,293-4,862)
	2	986 (936-1,036)	177 (156-198)	5,955 (5,620-6,290)	813 (777-849)	153 (135-170)	4,868 (4,633-5,103)
	3	1,138 (1,072-1,205)	198 (175-221)	6,916 (6,462-7,369)	791 (750-832)	153 (136-171)	4,708 (4,437-4,979)
	4	978 (930-1,026)	199 (178-221)	5,764 (5,448-6,080)	962 (918-1,006)	186 (168-204)	5,734 (5,441-6,028)
	Most deprived	1,099 (1,051-1,147)	226 (205-246)	6,464 (6,148-6,781)	955 (914-997)	186 (166-206)	5,681 (5,412-5,951)
	Total	**1,015 (993-1,037)**	**196 (186-205)**	**6,048 (5,902-6,194)**	**851 (833-869)**	**164 (156-172)**	**5,070 (4,951-5,189)**

ESP: European Standard Population; ASMR: Age-standardised Mortality Rate; CI: Confidence Interval; T: Total; M: Male; F: Female

4.4.2.1 Premature mortality

Figure 4.4 shows the age-standardised premature mortality rates (0-69) by deprivation quintile for the two periods. Visual inspection of the linear trends show that the rates have declined between the two time periods however the decline is not equal across the deprivation quintiles. While the premature ASMR for the least deprived quintile reduced from 234 per 100,000 (95% CI 217 - 250) in the earlier period, to 183 per 100,000 (95% 168 - 197) in the more recent period, shifts in the most deprived quintile values are much smaller, from 286 per 100,000 (95% 270 - 303) to 278 (95% CI 261 - 295) and the confidence intervals overlap between the estimates for the two time points.

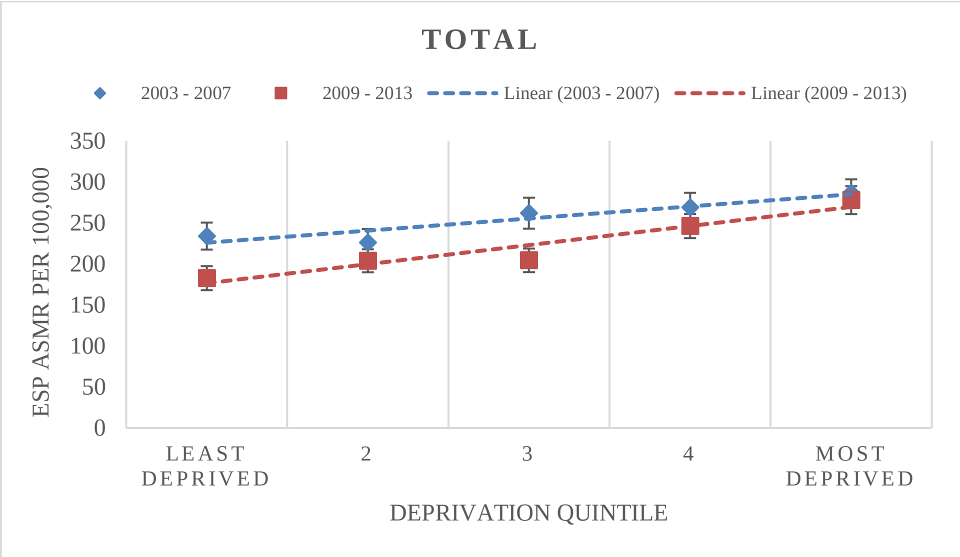


Figure 4.4: ESP ASMRs age 0-69 for the total by deprivation quintile: 2003-2007 & 2009-2013

When analysing the rates by sex, premature mortality was higher amongst males when compared to females across the deprivation quintiles and in both periods. While there were declines for both sexes when comparing the two periods, the declines were not consistent across sex. Figure 4.5 shows the estimates for males with a similar pattern as seen for the total. The estimated rate for the least deprived quintile is 294 per 100,000 (95% CI 267 - 320) in the earlier period reducing to 227 per 100,000 (95% CI 204 - 250) in the most recent period, while for the most deprived quintile the rate has increased in the most recent period from 349 per 100,000 (95% CI 322 - 375) to 372 per 100,000 (95% CI 344 - 400). Figure 4.6 presents estimates for females. While estimates between the two periods have also gone down, the declines in absolute mortality across the deprivation quintiles is more consistent, with the linear trends remaining parallel. The estimate for the most deprived quintile decreased from 226 per 100,000 (95% CI 205 - 246) to 186 per 100,00 (95% CI 166 - 206) while for the least deprived it decreased from 175 per 100,000 (95% CI 155 - 195) to 139 per 100,000 (95% CI 121 - 157).

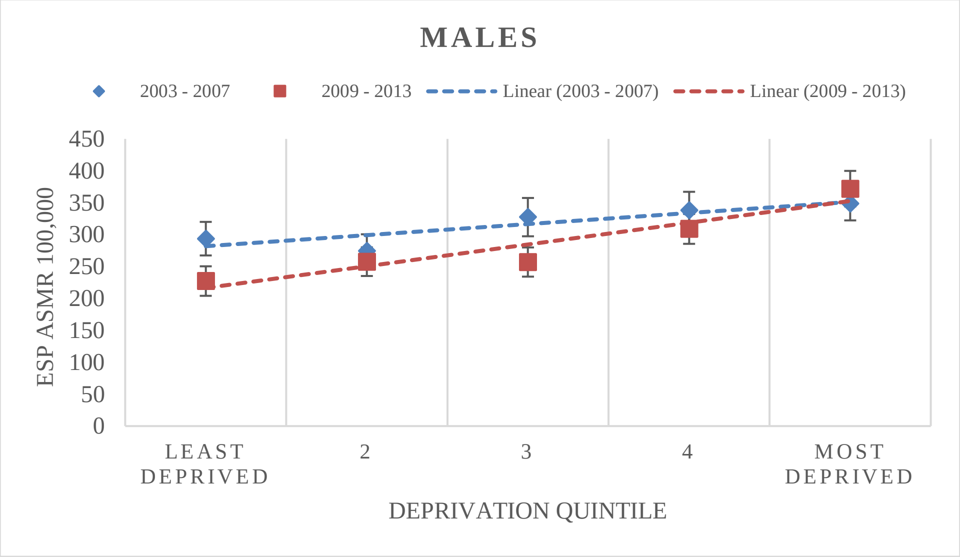


Figure 4.5: ESP ASMRs age 0-69 for males by deprivation quintile: 2003-2007 & 2009-2013

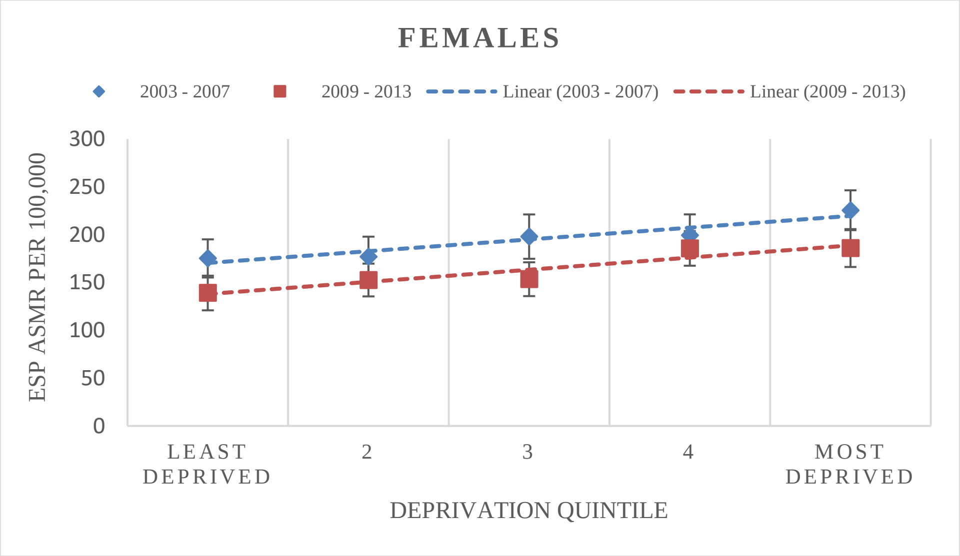


Figure 4.6: ESP ASMRs age 0-69 for females by deprivation quintile: 2003-2007 & 2009-2013

To assess the linear gradient by deprivation quintile, the RII was estimated. Figure 4.7 shows the estimates for the total and by sex for the two periods. For both periods there is a linear gradient for premature mortality by deprivation quintile, though the trend by sex varies. While a slight gap between the RII estimates for males and females exists for the earlier period, with females having a slightly higher RII compared to males (F:15.9% 95% CI 7.6% - 25.0%; M:13.7% 95% CI 7.5% - 20.3%), the gap grew considerably in the more recent period, mainly driven by the increased RII for males. The RII for males increased by 15% while the RII for females increased by only 3%.

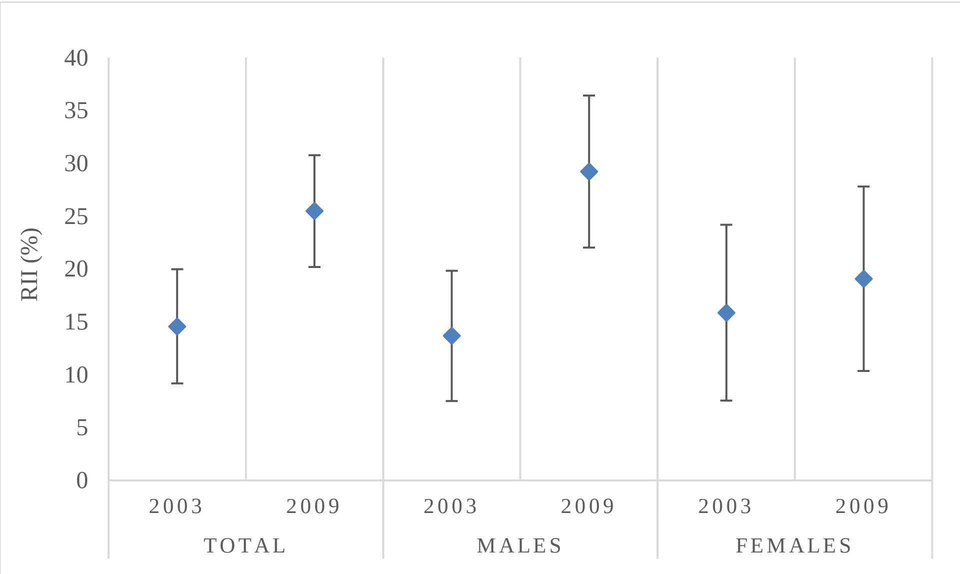


Figure 4.7: RII (%) estimates for premature ASMRs or the total and by sex: 2003-2007 & 2009-2013

In the most recent period, the risk of premature mortality for the total population was 25.5% (95% CI 20.2% - 30.8%) higher for those living in the most deprived areas when compared to the average risk for the total population. For males, the increased risk stood at 29.2% (95% CI 22.0% - 36.1%) in the most deprived areas while for females it was 19.1% (95% CI 10.4% - 28.0%).

4.4.2.2 Old-age mortality

Figure 4.8 shows the old-age (70+) age-standardised mortality rates by deprivation quintile for the two periods. As described previously, the estimates for the 3rd quintile were out of the linear trend, especially when using the 2005 index on the 2003-2007 data. Additional analysis was conducted to try to explain the reason for this pattern.

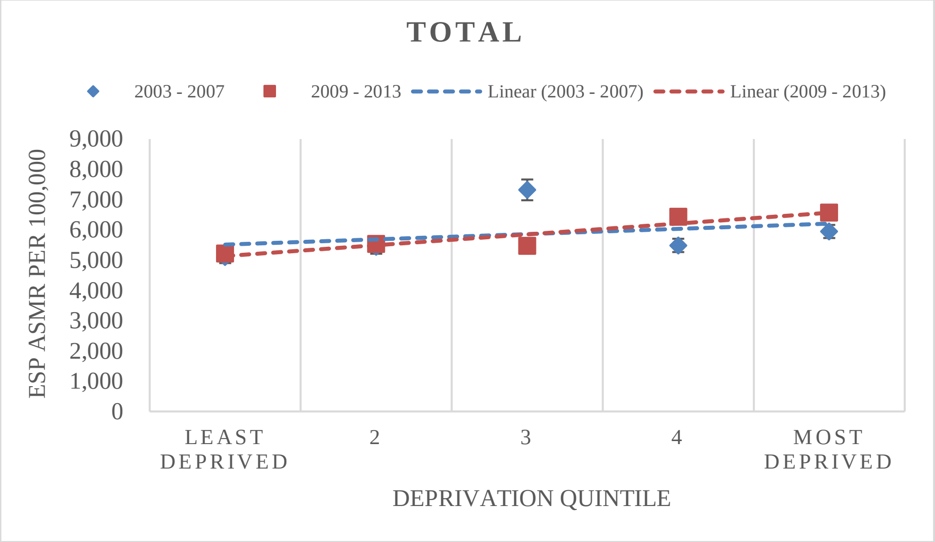


Figure 4.8: ESP ASMRs age 70+ for the total by deprivation quintile: 2003-2007 & 2009-2013

A possible reason for the trends observed for old-age mortality and deprivation may have been the impact of very-old age mortality on the estimates. It is more likely that the very old population live in an institution – mainly homes for the elderly – rather than their private residences within the community. Due to the small size of the country, elderly homes are not present in all localities in Malta and movement into an elderly home may not necessarily be based on the persons previous private residential location but the availability of beds. Based on the 2005 census, there were 22 localities out of 68 with homes for the elderly, while in 2011 this went up to 29. This means that less than half of the localities had homes for the elderly making it possible that mortality counts in the very old were higher in some localities and not others, because of the presence of an elderly home. The location of elderly homes in selected localities, would also be expected to influence the share of the very old in the denominator population counts. These may however have not been adequately accounted for since the breakdown of the population by age for the localities used for the ASMRs had to be estimated using the census data as this data was not available from the NSO. It is therefore possible that there was some internal migration of this age group after the censuses because of movement into selected localities which had institutions. This internal migration pattern may be biased by locality for the old age group but not for the rest of the ages. Considering these possible limitations, estimates of old age mortality were recalculated, truncating the oldest age bracket at 84. This age was chosen as the cut-off point based on the average age of the population enumerated in homes for the elderly in the 2005 and 2011 censuses, which stood at 81.5 and 81.8, respectively.

Figure 4.9 shows the updated old-age age-standardised mortality rates by deprivation quintile for the two periods for deaths at age 70 to 84. Excluding the very old age mortality improved the linear trend, especially for the 2009-2013 estimates. While the estimate for the 3rd quintile for 2003-2007 reduced and was closer to the linear trend, it was still elevated. Overall, total old-age mortality declined between the two periods, from 3,850 per 100,000 (95% CI 3,765 - 3,936) for deaths in 2003-2007 to 3,633 per 100,000 (95% CI 3,550 - 3,717) for deaths in 2009-2013. The mortality rate declined marginally for the most deprived (3,979; 95% CI 3,815 - 4,143 to 3,908, 95% CI 3,732 - 4,084) and 4th quintile (3,890; 95% CI 3,697 - 4,083 to 3,859; 95% CI 3,670 - 4,047) while the greatest declines were seen for the 2nd quintile (3,702; 95% CI 3,504 - 3,901 to 3,359; 95% CI 3,192 - 3,526) when excluding the difference between the 3rd quintile.

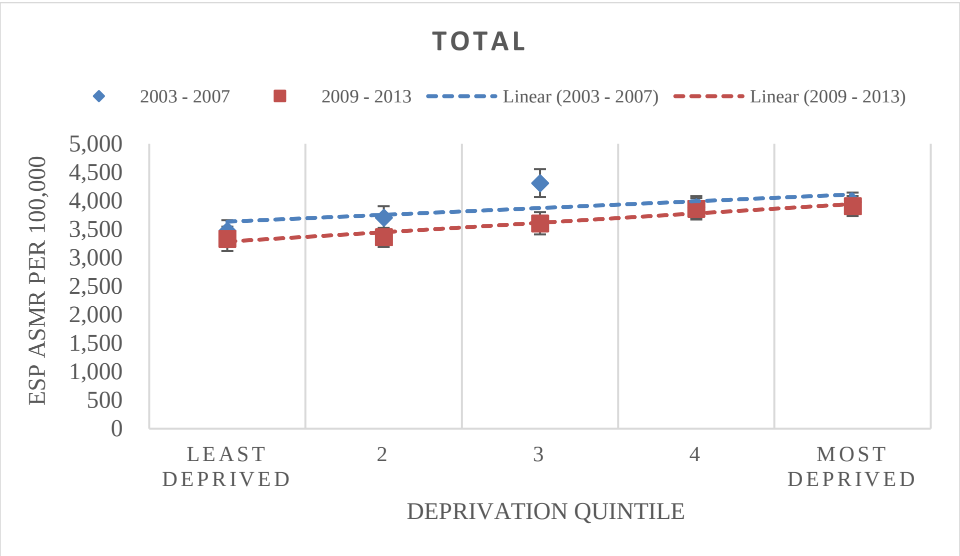


Figure 4.9: ESP ASMRs age 70-84 for the total by deprivation quintile: 2003-2007 & 2009-2013

Figures 4.10 and 4.11 present the ESP ASMRs for old-age mortality by sex for both periods. Visual inspection of the graphs shows that the pattern observed with the 3rd quintile in the total estimates were attributed to the estimates for males and not for females. The linear trend for males showed that age-standardised mortality rates increased in the most recent period in the 4th (4,149; 95% CI 3,866 - 4,432 to 4,819; 95% CI 4,494 - 5,144) and most deprived

(4,103; 95% CI 3,866 - 4,341 to 5,128; 95% CI 4,809 - 5,447) quintiles. This indicates an increasing gap in mortality risk for deaths at age 70 to 84, between the least deprived and most deprived quintiles. This increasing gap among males was similar to the pattern which was observed for premature mortality.

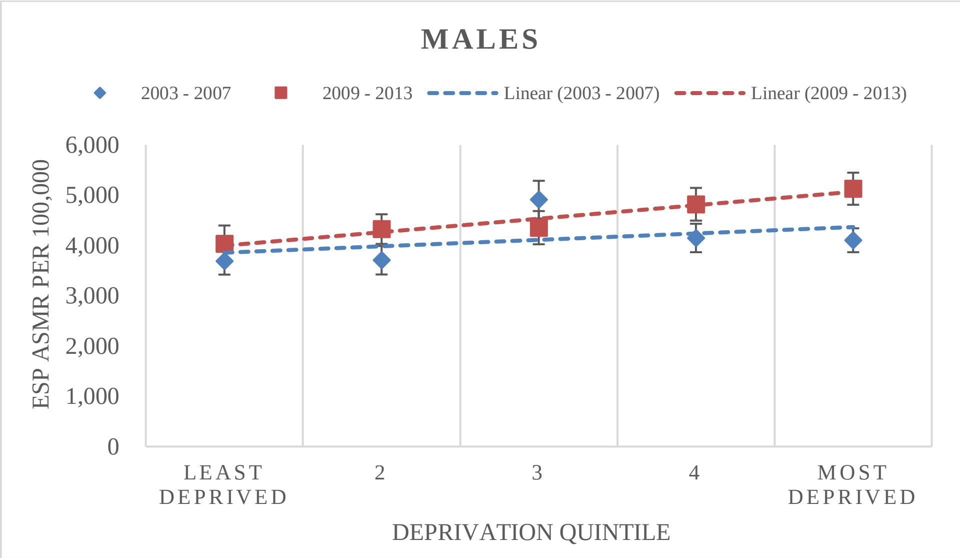


Figure 4.10: ESP ASMRs age 70-84 for males by deprivation quintile: 2003-2007 & 2009-2013

The pattern for females was also like that observed for premature mortality, with declines in the ASMRs for deaths between age 70-84 across all the quintiles. This led to a shifting down of the linear trend line indicating a decrease in absolute mortality across quintiles, though the gradient of the linear trend remained consistent.

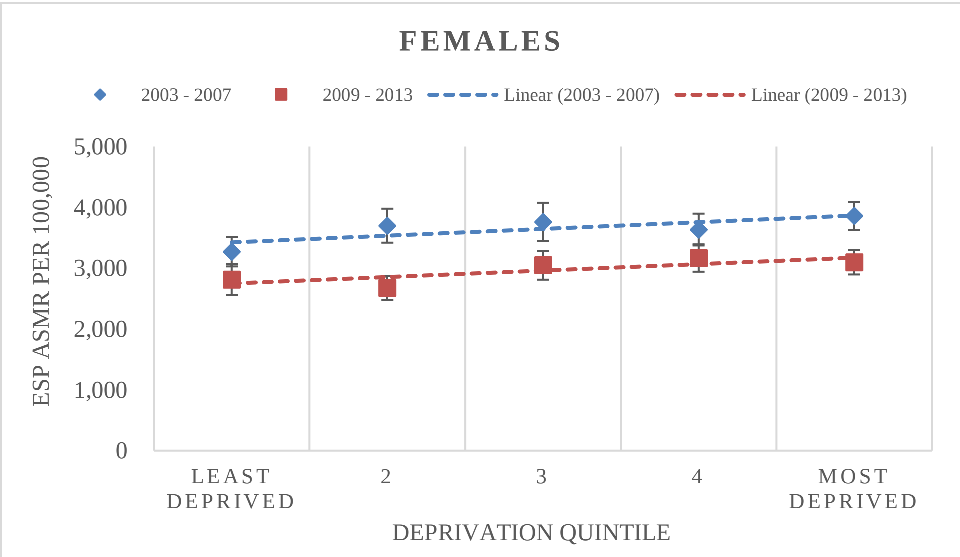


Figure 4.11: ESP ASMRs age 70-84 for females by deprivation quintile: 2003-2007 & 2009-2013

To assess the linear gradient of mortality risk for old-age mortality (70-84) by deprivation quintile, the RII was estimated. Figure 4.12 shows the estimates for the total and by sex for the two periods. The RII for females remained relatively similar between the two periods at 8.3% for deaths occurring between 2003-2007 (95% CI 2.3% - 13.8%) and 8.9% for deaths occurring between 2009-2013 (95% CI 3.0% - 14.7%). The estimate for the total increased slightly in the more recent period, however this difference must be interpreted with caution considering the possible impact of the spike in mortality in the 3rd quintile on the RII estimate for deaths in 2003-2007. For old age mortality in 2009-2013 the risk of mortality for the total population aged 70 to 84 was 11.2% (95% CI % 6.8% - 15.3%) higher for those living in the most deprived areas when compared to the average risk for the population.

For males, the RII value nearly doubled between the two periods, which is in line with the increased gradient observed in the linear trends discussed previously. The RII for 2003-2007 was 7.7% (95% CI 2.3% - 13.8%) increasing to 14.5% (95% CI 8.5% - 20.2%) for 2009-2013. However, this difference must also be interpreted with caution considering the possible impact of the spike in mortality in the 3rd quintile on the RII estimate for deaths in 2003-2007.

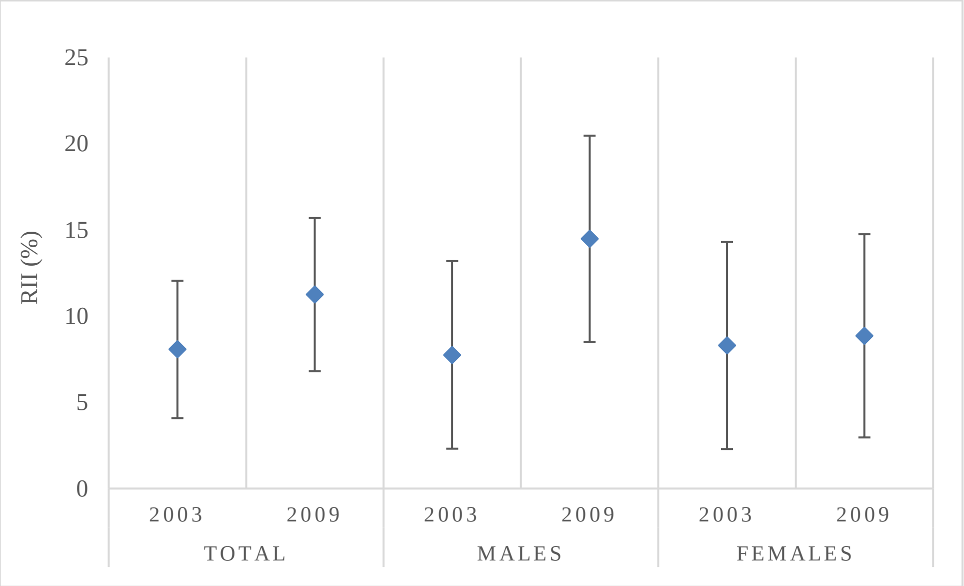


Figure 4.12: RII (%) estimates for old-age mortality (70-84) ASMRs for the total and by sex: 2003-2007 & 2009-2013

Further analysis was conducted to try to try to explain the pattern observed in the estimates for old age mortality for 2003-2007, among males, after the exclusion of the very old age group. The overall performance of the 2005 index for this specific age group may be a result of varying performance of the individual items. To assess this, each individual item from the 2005 census data used to create the composite index, was used to create single-item indices. This was done by using the individual unweighted Z-scores for each item and compiling deprivation quintiles in the same way as was done for the composite index – first ranking localities by score, then by population size and finally segmenting the ranked scores into quintiles based on a proportionate distribution of the population. Once the quintiles were assigned to each locality, the mortality and population data was assigned the respective quintile value for each item index separately. ESP ASMRs were estimated, by deprivation quintile, for males aged 70-84 using all 5 indices. To allow for comparison, the same estimates were produced for the premature mortality category (0-69). The RII was estimated for the resulting ESP ASMRs and are presented in figure 4.13.

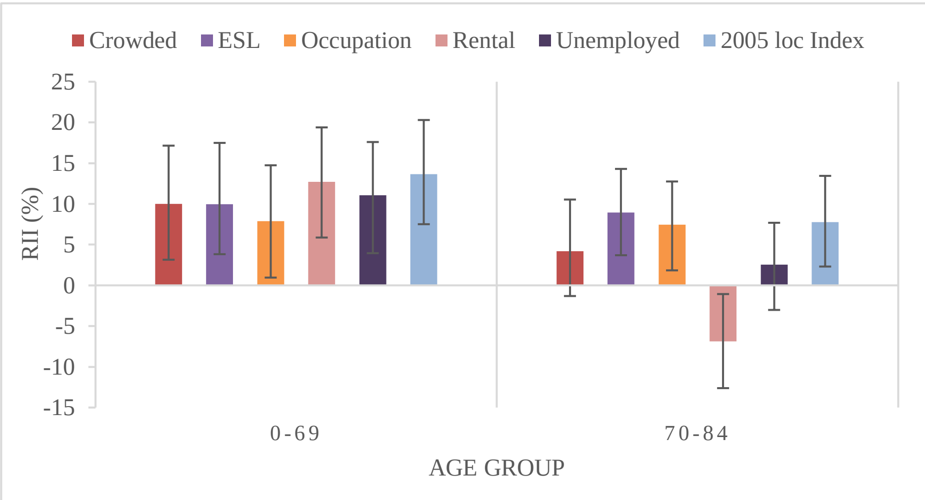


Figure 4.13: RII (%) estimates for premature mortality (0-69) and old-age mortality (70-84) ASMRs for males: 2003 – 2007

The RII values for the gradient of mortality risk by deprivation for premature mortality (0-69) among males were relatively consistent when using the single items as separate measures of deprivation compared to the composite index. While the RII for ASMRs by deprivation quintile estimated using the composite index was 13.7% (95% CI 7.5% – 20.3%) the values for the individual items ranged from 7.9% (95% CI 1.0% - 14.7%) for the item related to occupation (share of employed population in elementary occupations) to 12.7% (95%

5.9% - 19.4%) for the item related to rental of dwelling (share of occupied dwellings not owner occupied). On the other hand, there was less consistency in the RII estimates for old-age mortality across the items. While the RII for the composite index was 7.7% (95% CI 2.3% - 13.4%), the RII based on the single item indices varied considerably from -6.9% (95% CI -12.6 - -1.1%) for the item on rental of dwelling (share of occupied dwellings not owner occupied) to 9.0% (95% CI 3.7% - 14.3%) for the item on early school leavers (share of population aged 18 to 24 who are early school leavers). Based on these RII estimates, it was possible that some of the items included in the composite index did perform as well in predicting a gradient in old-age mortality by deprivation as for premature mortality among the male population for deaths occurring in 2003-2007.

4.5 Discussion

This chapter sought to test the use of a small area-based deprivation index in health inequalities analysis as applied to mortality. Age-standardised mortality rates (ASMRs) were chosen as the outcome measure to test whether the index created in this study predicted a gradient by deprivation quintile. The use of mortality estimates to assess the predictive validity of small area-based deprivation indices is common across indices as was discussed in Chapter 2. For the all-cause mortality analysis, estimates were produced for deaths occurring in Malta between 2003-2007 and 2009-2013 for the total, as well as for premature mortality (0-69) and old-age mortality (70+). Apart from the analysis of all-cause mortality to assess the predictive validity of the index at the locality level, additional analysis was conducted using mortality five years from the census to compare the performance of the index at the enumeration area and locality levels. This was conducted to assess the comparative performance of the locality index, since based on cross-tabulations presented in Chapter 3, the deprivation index developed at the locality level masked pockets of deprivation found within the smaller enumeration areas. For both analyses, the relative index of inequality (RII) was also estimated to quantify the relative socioeconomic gradient, if any, between deprivation and mortality.

The results from this chapter have illustrated several key points. Firstly, while comparative analysis between the enumeration area and locality level indices showed that the enumeration area index predicted slightly stronger gradients using the relative index of inequality (RII), the results using the locality index were still comparable. This shows that the small area-based deprivation index developed in this research is a useful tool for analysing health inequalities when data on locality are available in administrative health sources, even though these areas are not as small as the EAs. In the case of the data provided for this analysis, all records contained a locality therefore linkage to the deprivation quintile was complete.

While premature mortality rates declined between the two periods, the linear gradient for premature mortality by deprivation quintile, increased. This suggests that the declines in mortality risk were not equal across deprivation quintiles. This was most notable for males in the most deprived areas, where the premature mortality rate in the most recent period increased. Similar patterns in premature mortality have been noted by other researchers, and this has been attributed to increases in mortality related to lifestyle factors such as alcohol and drug use, and mental health related causes in more deprived areas (Leyland *et al.*, 2007; Norman *et al.*, 2011) Results for old age mortality were less consistent. Old-age mortality rates estimated for the third quintile were outside the linear trend. While exclusion of mortality in the very old age group (85+) attenuated the results when using the 2011 index, estimates for males in quintile 3 remained out of the trend line when using the 2005 index. Additional analysis on individual items showed that it is possible that some of the items included in the composite index do not predict old age mortality as well as premature mortality when using the 2005 index.

It must be acknowledged that the analysis presented here does have some limitations. Firstly, the enumeration area (EA) index could not be mapped to the mortality register since EAs are only developed for census operations and are not easily linked to other external data sources, therefore the locality index could only be used. As was shown in the comparative analysis of the indices at EA and locality level, the use of the EA index may have led to stronger gradients in deprivation as it captures pockets of deprivation within localities which may be masked when considering the deprivation quintile for the locality. Secondly, only all- cause mortality data was available which limited the ability to analyse specific causes of mortality which may have helped explain the trends seen in premature mortality amongst males. To avoid the analysis of small counts, mortality data was grouped into five-year periods, however some of the analysis was still based on relatively small numbers per quintile which led to large margins of error around the results. This is an issue unique to small countries that must be acknowledged when interpreting findings. Thirdly, population estimates at locality level by age and sex were not available therefore they had to be approximated using the age and sex breakdowns available from the censuses. This method assumed that the age distribution by locality for the years around the censuses remained fixed and this may not be applicable for all years or all localities. However, since the distributions used for analysis were 5-year age groups, the differences over time were expected to have had less impact than would have been the case for single years of age. Despite these limitations, results from the analysis are comparable to those found in other similar analysis. To the author's knowledge, this is the first attempt at analysing the social determinants of mortality in Malta, beyond patterning by age, sex, and district of residence.

Chapter 5: Deprivation and cancer incidence – application to cancer registry data

5.1 Introduction

Cancer is the leading cause of death worldwide, accounting for nearly 10 million deaths in 2020. Several cancers have been linked to specific lifestyle risk factors such as tobacco consumption, diet, and physical inactivity (World Health Organisation, 2022). Health inequalities exist across the entire cancer pathway from varying exposure to risks, likelihood of incidence, late diagnosis, access to screening programmes, delay and overall access to treatments, prognosis, survivorship, and mortality (Lortet-Tieulent *et al.*, 2019). Different outcome measures can be used to assess these health inequalities, including but not limited to, incidence rates, screening uptake, treatment delay, recurrence rates, survivorship, and mortality. Each outcome measure would be providing a snapshot of the inequalities along the cancer care pathway and several competing risk factors may all contribute to the relationship between area level deprivation and the outcome measure being analysed, such as stage at diagnosis, access to screening programmes, lifestyle, age, sex, and environmental exposures, to name a few.

For the purposes of this chapter, incidence was selected as the outcome measure for analysis. Neighbourhoods and area deprivation may be associated with cancer incidence in several ways. Environmental factors related to the area itself may increase cancer incidence due to carcinogen exposure, such as exposure to air pollution and other hazardous substances, more common in urban areas or neighbourhoods near polluting industries. For example, several studies have estimated an increased risk in lung cancer incidence due to exposure to air pollutants (Krewski *et al.*, 2009; Raaschou-Nielsen *et al.*, 2013)Access to healthy foods may also vary in neighbourhoods. The concentration of fast-food options has been found to be higher in more deprived areas, especially urbanised areas (Kwate *et al.*, 2009; Macdonald *et al.*, 2018) and this may impact upon the spatial patterning of cancer incidence as research has shown a link between the consumption of ultra-processed foods and certain cancers such as colon-rectum and ovarian cancer (Chang *et al.*, 2023; Isaksen and Dankel, 2023). Furthermore aspects of the built up environment, such as walkability and greenspaces, are known to promote healthy behaviours and these may vary by area (Larsen, Rydz and Peters, 2023). Apart from the factors related to the environment of neighbourhoods and areas themselves, evidence has shown that persons in more deprived areas are more likely to adopt unhealthy lifestyles that are known risk factors for cancer, such as excess alcohol consumption (Bellis *et al.*, 2016), smoking and physical inactivity (Algren *et al.*, 2015). The spatial patterning of cancer incidence may also yield paradoxical associations with area deprivation because of the impact of screening uptake. For example, a number of studies have found that incidence rates of breast cancer are lower in more socioeconomically deprived areas while mortality rates are higher. Such findings have been attributed to the fact that women in more deprived areas may be less likely to participate in screening (Smith *et al.*, 2019).

While analysing cancer incidence will only address one part of the possible inequalities along the cancer pathway and therefore provides only a preliminary study of the relationship between deprivation and cancer, the aim of this analysis was to present another example of how the small area-based deprivation index created in Chapter 3 can be used to measure health inequalities in Malta using administrative data. The rest of the chapter will be divided as follows: Section 5.2 will describe the data sources used to estimate the age-standardised incidence rates; Section 5.3 will describe the methods used to estimate the outputs analysed in this chapter, while Section 5.4 will present the results. Finally, Section 5.5 will discuss the findings presented in this chapter.

5.2 Data source

Incident cases of cancer were sourced from the National Cancer Register which is managed by the Directorate for Health Information and Research (DHIR) within the Ministry for Health, Malta. The register started operating in 1985, first collecting data from the main public health hospital at the time. As of 1991, the register obtained full coverage by collecting all incident cases diagnosed in all hospitals, clinics and laboratories operating across Malta. Data held within the register is coded using the international classification of diseases for oncology (ICD-O) (Directorate for Health Information and Research Malta, 2023). Underpinning the register is the Notification of Cancer Act, legislation which was originally enacted in 1957, and which makes it a legal obligation for all medical practitioners to notify the public health authorities of newly diagnosed cancer cases. The legislation also stipulates the details to be provided through the notification of cancer form. The locality of residence of the patient is collected as part of the notification form, therefore a small area-based deprivation index developed at the locality level is an appropriate measure to use to analyse health inequalities in cancer, based on data from this register. This is especially since currently only basic patient characteristics are collected from the cancer register notification form and no socioeconomic details of the patient are available (Government of Malta, 1957).

Prior to requesting the data, an initial meeting was held with the manager of the register. Since analysis was to be conducted by cancer site, it was agreed that analysing single-year counts would not be possible due to the small counts. It was therefore agreed, that as with the analysis conducted in Chapter 4, grouped count data would be requested for the period 2003-2007 and 2009-2013, that is, the five years around the two censuses. In 2011, 2,006 incident cases of cancers (excluding cutaneous basal cell and squamous cell carcinomas) were diagnosed and registered in Malta. Of these, 51% were females and 71% of all incident cases were diagnosed at age 60 and over. Paediatric (childhood) cancers, or cancers diagnosed before the age of 15, accounted for approximately 1% of all incident cases (National Cancer Register, 2023a). Female breast cancer was the most common cancer (343) (National Cancer Register, 2023c) followed by colorectum (218) (National Cancer Register, 2023b), male prostate (210) (National Cancer Register, 2023e) and lung (152) (National Cancer Register, 2023d). These four cancer sites accounted for a little under half of the incident cases. In agreement with the manager, analysis was restricted to the top six cancer sites to avoid analysing small counts that may have led to possible large margins of error. These were female breast (C50), colorectum (C18, C20), male prostate (C61), lung/bronchus (C34), bladder (C67) and lymphoid (histology code 959-972). Since paediatric cancers are still considered rare and their occurrence, cause, prognosis, and treatment may vary considerably from adult cancers, the analysis was restricted to cancers diagnosed at 15 years of age and over.

While initially data was to be requested by cancer site, sex, age group and locality of residence, such detailed breakdown of small counts by locality of residence would have potentially been deemed identifiable. It was instead agreed that the researcher would map the quintile category from the relevant small area-based index to the locality list used by the register. The researcher provided the list to the register manager, and it was used to produce the aggregate data broken down by deprivation quintile, period of diagnosis (5-years), 5-year age group at diagnosis, sex, and cancer site. Since aggregate data was requested, grouped across a 5-year period, and excluding any data on location of residence, DHIR deemed that the data was not identifiable and therefore ethical approval was not required to process the request. Data was provided to the researcher in October 2022.

Population data used for the analysis was sourced from the National Statistics Office (NSO). The mid-year population estimates by deprivation quintile, age and sex which were described in Chapter 4, were used as the denominator data for the analysis which will be presented in this chapter.

5.3 Methods

The numerator and denominator data described above were used to estimate age-standardised incidence rates (ASIRs). As described in Chapter 4, the direct standardisation method was used to estimate the ASIRs and corresponding 95% confidence intervals. The method involves the application of age and sex specific incidence rates for each deprivation quintile and cancer site to the age and sex distribution of the reference population, in this case, the European Standard Population (ESP), described previously. This method ensures that the incidence rate is independent of differences in the age and sex distribution between populations within different deprivation quintiles. The ASIRs were estimated per 100,000-person time at risk. The relative index of inequality (RII) was also estimated to measure the inequality gradient by deprivation quintile in cancer incidence. The RII can be expressed as a percentage with the value indicating how the most deprived group differs, in percentage terms, from the average value of the population. A 95% confidence interval around the value was also estimated. The method used to estimate the RII and confidence intervals are described in more detail in Chapter 4.

5.4 Results

There was a total of 9,515 incident cancer cases diagnosed in the period under analysis, 4,088 cases in the period 2003-2007 and 5,247 in the period 2009-2013. Over the period under analysis, female breast cancer accounted for 29.9% of cases with an additional 20.6% of cases being colon-rectum cancer. Many of the incident cancers were diagnosed at 55 years of age and over (84%), and slightly more cases were diagnosed among males (51.2%). When considering deprivation quintile, 22.7% of the incident cases were diagnosed among individuals living in areas categorised within the most deprived quintile (Table 5.1).

Table 5.1: Incident malignant cancer cases by cancer site, sex, age at diagnosis and deprivation quintile: 2003-2007 & 2009-2013

		2003-2007	2009-2013	Total
		N (%)	**N (%)**	**N (%)**
Total		4,088 (100)	5,427 (100)	9,515 (100)
Cancer site	Breast (female)	1,233 (30.2)	1,616 (29.8)	2,849 (29.9)
	Prostate (male)	708 (17.3)	1,015 (18.7)	1,723 (18.1)
	Colon-rectum	857 (21.0)	1,102 (20.3)	1,959 (20.6)
	Bronchus-lung	686 (16.8)	833 (15.3)	1,519 (16.0)
	Lymphoid	311 (7.6)	502 (9.3)	813 (8.5)
	Bladder	293 (7.2)	359 (6.6)	652 (6.9)
Sex	Female	1,938 (47.4)	2,707 (49.9)	4,645 (48.8)
	Male	2,150 (52.6)	2,720 (50.1)	4,870 (51.2)
Age at diagnosis	15-24	17 (0.4)	24 (0.4)	41 (0.4)
	25-34	39 (1.0)	70 (1.3)	109 (1.1)
	35-44	165 (4.0)	174 (3.2)	339 (3.6)
	45-54	455 (11.1)	561 (10.3)	1,016 (10.7)
	55-64	1,052 (25.7)	1,313 (24.2)	2,365 (24.9)
	65-74	1,154 (28.2)	1,619 (29.8)	2,773 (29.1)
	75-84	966 (23.6)	1,220 (22.5)	2,186 (23.0)
	85+	240 (5.9)	446 (8.2)	686 (7.2)
Deprivation quintile	Least deprived	897 (21.9)	891 (16.4)	1,788 (18.8)
	2	767 (18.8)	1,079 (19.9)	1,846 (19.4)
	3	635 (15.5)	1,079 (19.9)	1,714 (18.0)
	4	759 (18.6)	1,248 (23.0)	2,007 (21.1)
	Most deprived	1,030 (25.2)	1,130 (20.8)	2,160 (22.7)

5.4.1 Female breast cancer

The total age-standardised incidence rate for female breast cancer increased between the two periods, from 158 per 100,000 (95% CI 149 - 166) to 185 per 100,000 (95% CI 176 - 194). For incident cases diagnosed between 2003-2007, the highest rate was within the least deprived quintile with 179 incident cases per 100,000 (95% CI 158 - 199), while the lowest rate was in the 3rd quintile at 136 per 100,000 (95% CI 116 - 156). For the period 2009-2013 the highest incidence rate was in the 4th quintile at 210 per 100,000 (95% CI 190 - 231), while the lowest rate was in the most deprived quintile at 160 per 100,000 (95% CI 141 - 178). Incidence rates increased across all quintiles except for the most deprived quintile where the incidence rate remained the same. The RII estimates in both periods were negative, indicating a negative gradient by deprivation, though both periods had confidence interval ranges that included 0. The RII estimate for 2009-2013 became less negative at -1.6% (95% CI -10% - 6.6%) from -

5.3% (95% CI -15% - 4.5%) in 2003-2007 (Table 5.2). These RII values indicated that the risk of being diagnosed with incident female breast cancer as slightly lower for persons in the most deprived areas compared to the average risk for the total population.

Table 5.2: ESP ASIRs per 100,000 for female breast cancer by deprivation quintile and RII (%) estimates: 2003-2007 & 2009-2013

	2003-2007	**2009-2013**
	ESP ASIR per 100,000 (95% CI)	***ESP ASIR per 100,000 (95% CI)***
Total	**158 (149 - 166)**	**185 (176 - 194)**
Least deprived	179 (158 - 199)	184 (153 - 207)
2	154 (135 - 174)	174 (155 - 193)
3	136 (116 - 156)	196 (175 - 217)
4	159 (138 - 179)	210 (190 - 231)
Most deprived	160 (142 - 179)	160 (141 - 178)
RII (%)	-5.3% (-15.0% - 4.5%)	-1.6% (-10.0% - 6.6%)

ESP: European standardised rate; ASIR: Age-standardised incidence rate; CI: Confidence interval; RII: Relative index of inequality

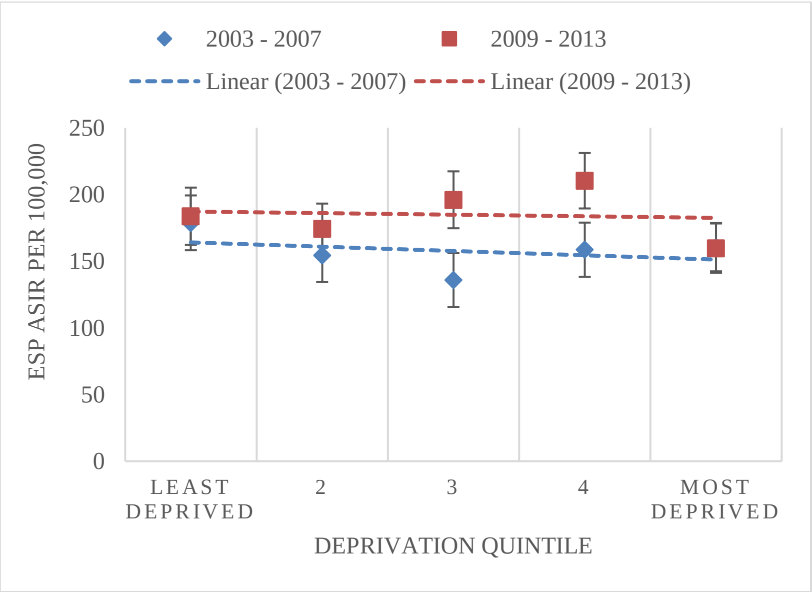


Figure 5.1: ESP standardised incidence rates per 100,000 for female breast cancer by deprivation quintile: 2003-2007 & 2009-2013

5.4.2 Male prostate

The total age-standardised incidence rate for male prostate cancer increased between the two periods, from 101 per 100,000 (95% CI 94 - 108) to 149 per 100,000 (95% CI 140 - 158). For incident cases diagnosed between 2003-2007, the highest rate was within the 3rd quintile with 127 incident cases per 100,000 (95% CI 105 - 150), while the lowest rate was in the 4th quintile at 84 per 100,000 (95% CI 69 - 99). For the period 2009-2013 the pattern was reversed, with the highest incidence rate in the 4th quintile at 157 per 100,000 (95% CI 137 - 178), while the lowest rate was in the 3rd quintile at 144 per 100,000 (95% CI 124 - 165). Incidence rates increased across all quintiles in the more recent period with the gap between the highest and lowest rates also decreasing. The RII estimate for the earlier period was negative, indicating that the risk of being diagnosed with male prostate cancer was 16.6% lower in persons in the most deprived area compared to the average risk for the total population (95% CI -29.0% - -3.8%). In the more recent period, the value became marginally positive, though the confidence interval included 0 (Table 5.3).

Table 5.3: ESP ASIRs per 100,000 for male prostate cancer by deprivation quintile and RII (%) estimates: 2003-2007 & 2009-2013

	2003-2007	2009-2013
	ESP ASIR per 100,000 (95% CI)	***ESP ASIR per 100,000 (95% CI)***
Total	**101 (94 - 108)**	**149 (140 - 158)**
Least deprived	111 (94 - 129)	146 (123 - 169)
2	108 (90 - 126)	150 (131 - 170)
3	127 (105 - 150)	144 (124 - 165)
4	84 (69 - 99)	157 (137 - 178)
Most deprived	90 (77 - 104)	146 (126 - 165)
RII (%)	-16.6% (-29.4% - -3.8%)	1.0% (-11.0% - 12.0%)

ESP: European standardised rate; ASIR: Age-standardised incidence rate; CI: Confidence interval; RII: Relative index of inequality

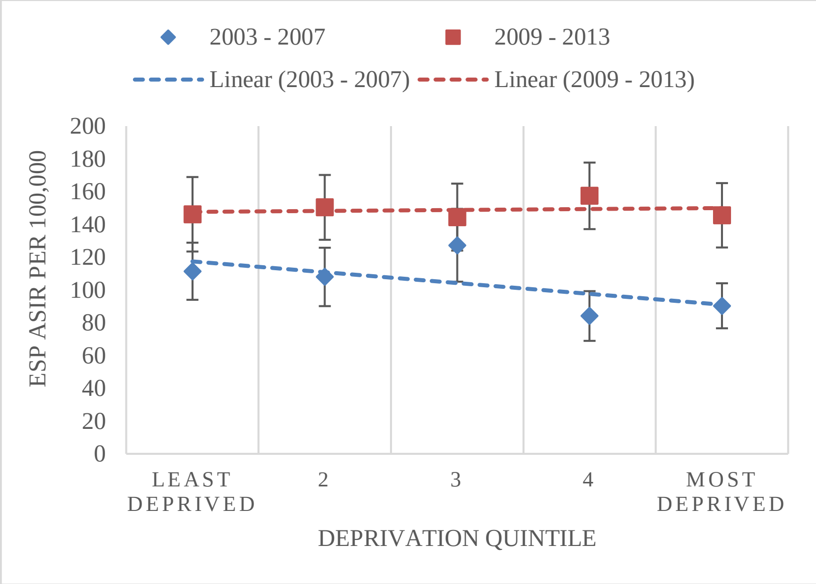


Figure 5.2: ESP standardised incidence rates per 100,000 for male prostate cancer by deprivation quintile: 20032007 & 2009-2013

5.4.3 Colon-rectum

The total age-standardised incidence rate for colon-rectum cancer increased between the two periods, from 58 per 100,000 (95% CI 54 - 62) to 71 per 100,000 (95% CI 67 - 75). For incident cases diagnosed between 2003-2007, the highest rate was within the 2nd quintile with 66 incident cases per 100,000 (95% CI 57 - 76), while the lowest rate was in the 3rd and most deprived quintiles at 55 per 100,000 (95% CI 46 - 65; 47 - 62). In the most recent period, the highest rate was within the least deprived quintile at 80 per 100,000 (95% CI 69 - 91), while the lowest rate was within the 2nd quintile at 63 incident cases per 100,000 (95% CI 55 - 72). Incidence rates increased across all quintiles in the more recent period except in the 2nd quintile where a slight decrease was noted. The RII estimate for both periods were negative and remained relatively the same, with the risk of being diagnosed with colon-rectum cancer being 7% lower in persons in the most deprived compared to the average risk for the total population (95% CI -19.2% - 4.6%; -17.5% - 4.4%) (Table 5.4). Further analysis sub-divided by sex was not conducted due to the small counts and the CI ranges of the RII values presented for the totals.

Table 5.4: ESP ASIRs per 100,000 for colon-rectum cancer by deprivation quintile and RII (%) estimates: 2003-2007 & 2009-2013

	2003-2007	2009-2013
	ESP ASIR per 100,000 (95% CI)	***ESP ASIR per 100,000 (95% CI)***
Total	**58 (54 - 62)**	**71 (67 - 75)**
Least deprived	57 (49 - 66)	80 (69 - 91)
2	66 (57 - 76)	63 (55 - 72)
3	55 (46 - 65)	77 (67 - 87)
4	56 (47 - 64)	72 (63 - 81)
Most deprived	55 (47 - 62)	65 (56 - 73)
RII (%)	-7.1% (-19.2% - 4.6%)	-7.3% (-17.5% - 4.4%)

ESP: European standardised rate; ASIR: Age-standardised incidence rate; CI: Confidence interval; RII: Relative index of inequality

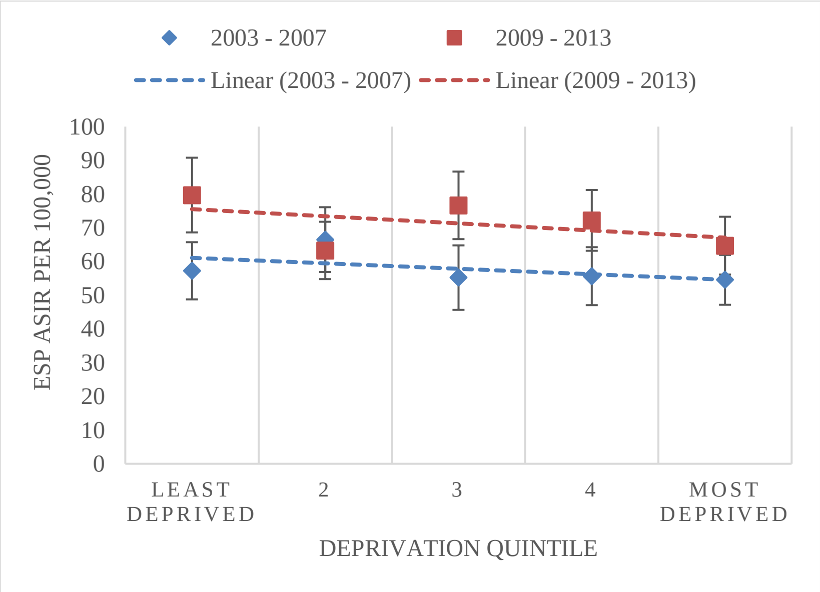


Figure 5.3: ESP standardised incidence rates per 100,000 for colon-rectum cancer by deprivation quintile: 2003-2007 & 2009-2013

5.4.4 Bronchus-lung

The total age-standardised incidence rate for bronchus-lung cancer increased between the two periods, from 46 per 100,000 (95% CI 43 - 50) to 53 per 100,000 (95% CI 50 - 57). For incident cases diagnosed between 2003-2007, the highest rate was within the 3rd quintile with 52 incident cases per 100,000 (95% CI 42 - 61), while the lowest rate was in the 2nd quintile at 37 per 100,000 (95% CI 30 - 45). In the most recent period, the highest rate was within the most deprived quintile at 67 per 100,000 (95% CI 58 - 76), while the lowest rate was within the 2nd quintile at 40 incident cases per 100,000 (95% CI 34 - 47). Incidence rates increased across all quintiles in the more recent period except in the 2nd quintile. The RII estimates for both periods were positive, with the RII more than doubling in the most recent period. The confidence interval for the earlier period does include 0. In the most recent period, the risk of being diagnosed with bronchus-lung cancer was 27.5% higher for persons in the most deprived areas compared to the average risk for the total population (95% CI 15.9% - 38.9%) (Table 5.5).

Table 5.5: ESP ASIRs per 100,000 for bronchus-lung cancer by deprivation quintile and RII (%) estimates: 2003-2007 & 2009-2013

	2003-2007	**2009-2013**
	ESP ASIR per 100,000 (95% CI)	***ESP ASIR per 100,000 (95% CI)***
Total	**46 (43 - 50)**	**53 (50 - 57)**
Least deprived	45 (38 - 53)	48 (39 - 56)
2	37 (30 - 45)	40 (34 - 47)
3	52 (42 - 61)	52 (43 - 60)
4	47 (39 - 55)	61 (53 - 70)
Most deprived	50 (43 - 57)	67 (58 - 76)
RII (%)	11.1% (-1.4% - 23.3%)	27.5% (15.9% - 38.9%)

ESP: European standardised rate; ASIR: Age-standardised incidence rate; CII: Confidence interval; RII: Relative index of inequality

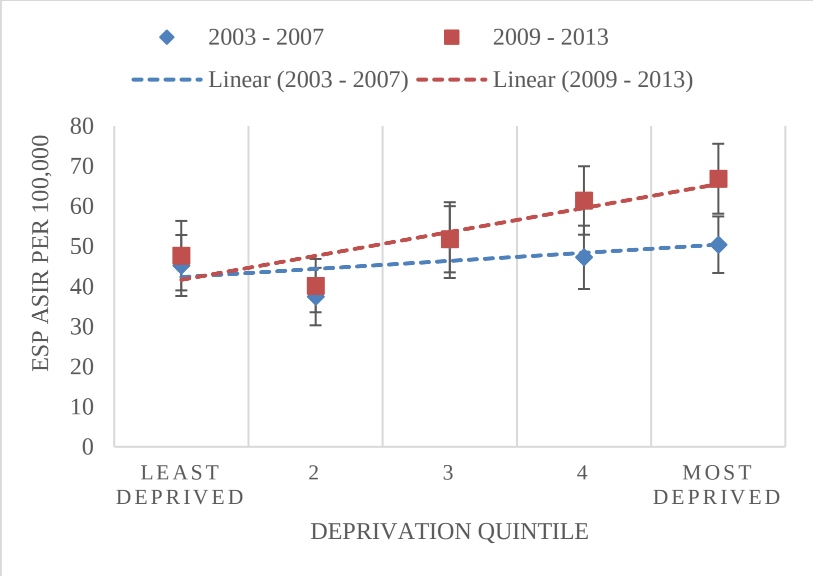


Figure 5.4: ESP standardised incidence rates per 100,000 for bronchus-lung cancer by deprivation quintile: 2003-2007 & 2009-2013

For the more recent period, further analysis was conducted by sex. Table 5.6 presents ASIRs and RII estimates by sex for the incident cases of bronchus-lung cancer diagnosed between 2009 – 2013. At the total level and across all deprivation quintiles, the incidence rate was higher for males than females with overall 89 incident cases per 100,00 (95% CI 82 - 96) compared to 28 per 100,000 (95% CI 24 - 31), respectively. For both sexes, the highest incidence rate was in the most deprived quintile while the lowest incidence rate was in the 2nd quintile. RII estimates were both positive with a higher value for females, though this estimate has a much wider confidence interval. The risk of being diagnosed with bronchus-lung cancer for males was 26.4% higher for those in the most deprived compared to the average risk for the total male population (95% CI 12.1% - 20.2%), for females the RII was 33.8% (95% CI 9.2% - 56.3%).

Table 5.6: ESP ASIRs per 100,000 for bronchus-lung cancer by deprivation quintile and sex and RII (%) estimates: 2009-2013

	Males	Females
	ESP ASIR per 100,000 (95% CI)	**ESP ASIR per 100,000 (95% CI)**
Total	**89 (82 - 96)**	**28 (24 - 31)**
Least deprived	75 (58 - 91)	27 (18 - 36)
2	69 (56 - 83)	19 (13 - 25)
3	90 (74 - 107)	23 (16 - 31)
4	104 (87 - 121)	30 (22 - 38)
Most deprived	105 (88 - 122)	40 (31 - 49)
RII (%)	26.4% (12.1% - 40.2%)	33.8% (9.2% - 56.3%)

ESP: European standardised rate; ASIR: Age-standardised incidence rate; CI: Confidence interval; RII: Relative index of inequality

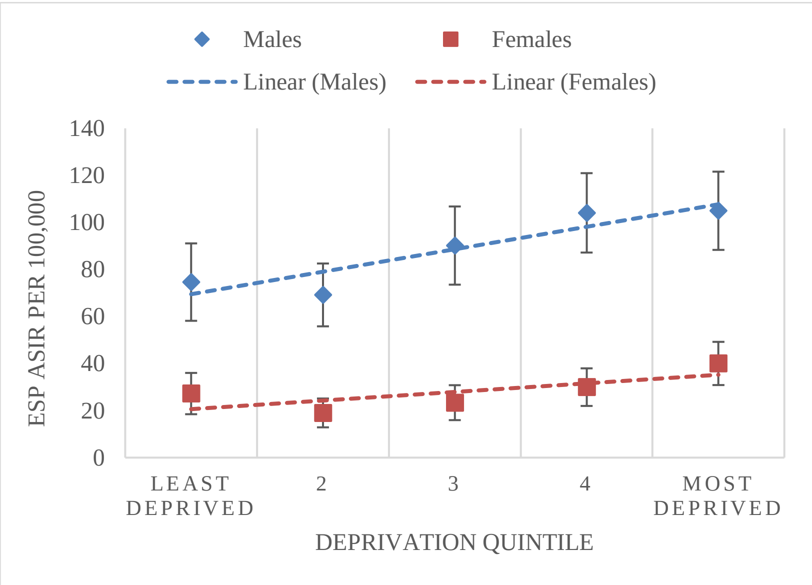


Figure 5.5: ESP standardised incidence rates per 100,000 for bronchus-lung cancer by deprivation quintile and sex: 2009-2013

5.4.5 Lymphoid

The total age-standardised incidence rate for lymphoid cancer increased between the two periods, from 20 per 100,000 (95% CI 18 - 22) to 30 per 100,000 (95% CI 28 - 33). For incident cases diagnosed between 2003-2007, the highest rate was within the most deprived quintile with 24 incident cases per 100,000 (95% CI 19 - 29), while the lowest rate was in the 2nd quintile at 17 per 100,000 (95% CI 12 - 22). In the most recent period, the highest rate was within the 3rd quintile at 36 per 100,000 (95% CI 29 - 43), while the lowest rate was within the 4th quintile at 27 incident cases per 100,000 (95% CI 22 - 33). Incidence rates increased across all quintiles in the more recent period. The RII estimates for the two periods were approximately 6% but in opposite directions, with a positive value for the earlier period and a negative value in the recent period. The confidence intervals were very wide for both estimates and included 0 (Table 5.7). Further analysis sub-divided by sex was not conducted due to the small counts and the CI ranges of the RII values presented for the totals.

Table 5.7: ESP ASIRs per 100,000 for lymphoid cancer by deprivation quintile and RII (%) estimates: 2003-2007 & 2009-2013

	2003-2007	**2009-2013**
	ESP ASIR per 100,000 (95% CI)	***ESP ASIR per 100,000 (95% CI)***
Total	**20 (18 - 22)**	**30 (28 - 33)**
Least deprived	22 (17 - 27)	30 (24 - 37)
2	17 (12 - 22)	30 (24 - 35)
3	19 (13 - 25)	36 (29 - 43)
4	18 (13 - 23)	27 (22 - 33)
Most deprived	24 (19 - 29)	28 (22 - 33)
RII (%)	5.8% (-14.4% - 25.3%)	-6.4% (-20.9% - 8.4%)

ESP: European standardised rate; ASIR: Age-standardised incidence rate; CII: Confidence interval; RII: Relative index of inequality

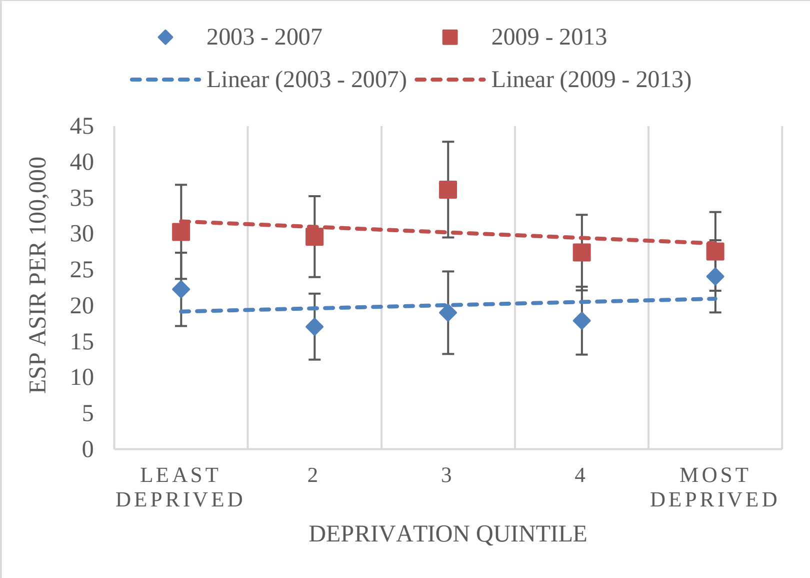


Figure 5.6: ESP standardised incidence rates per 100,000 for lymphoid cancer by deprivation quintile: 2003-2007 & 2009-2013

5.4.6 Bladder

The total age-standardised incidence rate for bladder cancer increased between the two periods, from 20 per 100,000 (95% CI 18 - 23) to 25 per 100,000 (95% CI 22 - 27). For incident cases diagnosed between 2003-2007, the highest rate was within the 3rd quintile with 24 incident cases per 100,000 (95% CI 17 - 31), while the lowest rate was in the most deprived quintile at 18 per 100,000 (95% CI 14 - 23). In the most recent period, the highest rate was within the 3rd quintile and most deprived quintile at 28 per 100,000 (95% CI 21 - 34; 22 - 33), while the lowest rate was within the 2nd quintile at 21 incident cases per 100,000 (95% CI 16 - 26). Incidence rates increased across all quintiles in the more recent period except for the least deprived quintile. The RII estimates for the two periods were in opposite directions, with a negative value for the earlier period and a positive value in the recent period. The confidence intervals were very wide for both estimates and included 0 (Table 5.8). Further analysis sub-divided by sex was not conducted due to the small counts and the CI ranges of the RII values presented for the totals.

Table 5.8: ESP ASIRs per 100,000 for bladder cancer by deprivation quintile and RII (%) estimates: 2003-2007 & 2009-2013

	2003-2007	2009-2013
	ESP ASIR per 100,000 (95% CI)	***ESP ASIR per 100,000 (95% CI)***
Total	**20 (18 - 23)**	**25 (22 - 27)**
Least deprived	22 (17 - 27)	22 (16 - 28)
2	20 (15 - 25)	21 (16 - 26)
3	24 (17 - 31)	28 (21 - 34)
4	20 (14 - 25)	24 (18 - 29)
Most deprived	18 (14 - 23)	28 (22 - 33)
RII (%)	-9% (-27.3% - 10.2%)	12.8% (-4.6% - 30.3%)

ESP: European standardised rate; ASIR: Age-standardised incidence rate; CII: Confidence interval; RII: Relative index of inequality

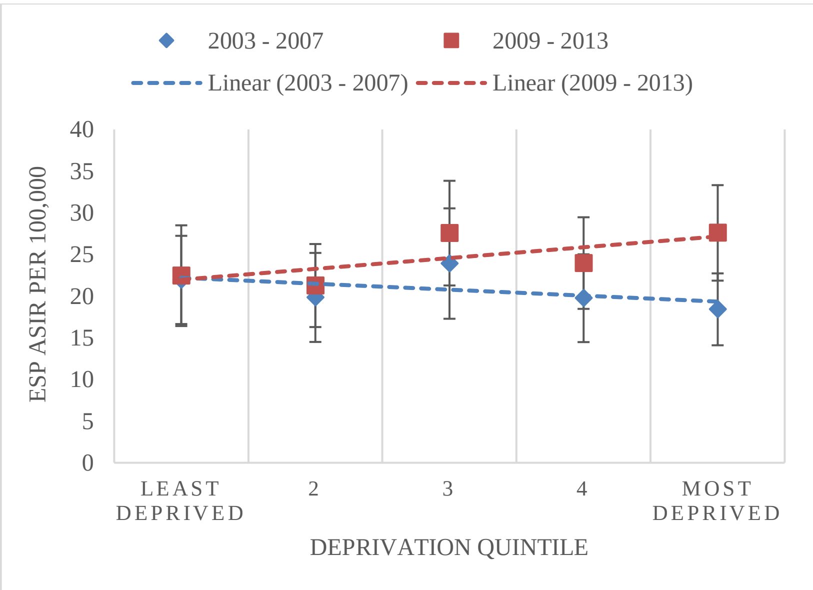


Figure 5.7: ESP standardised incidence rates per 100,000 for cancer by deprivation quintile: 2003-2007 & 2009-2013

5.5 Discussion

This chapter sought to apply a small area-based deprivation index for the analysis of health inequalities in cancer incidence for the six most common cancer sites diagnosed in Malta between 2003-2007 and 2009-2013. The outcome measures selected for this analysis was age-standardised incidence rates using the European standard population (ESP). While different outcome measures can be used to assess health inequalities along the cancer care pathway, incidence rates were selected to provide a preliminary analysis of the possible relationship between deprivation and cancer in Malta. Standardising the rates using a standard population allowed for comparison across deprivation quintiles independent of differences in the age and sex distribution between populations within the quintiles. The RII was selected to quantify the relative socioeconomic gradient, if any, and is a summary measure of the linear association between deprivation and incidence. The RII assumes the data is ranked which makes the measure appropriate for the analysis of deprivation quintiles which are ranked from least to most deprived. This preliminary analysis also served to test the application of the small area-based index to health inequalities analysis using another administrative health data source, where individual level variables on socioeconomic status are not collected but data on locality of residence is available and therefore can be linked to the index.

The results from this chapter have illustrated several key points. Firstly, the small area-based deprivation index developed in this research is a useful tool for analysing health inequalities when data on locality are available in administrative health sources. While the area-based measure may also serve to compliment individual data on socioeconomic indicators, the index is a useful proxy when socioeconomic data is absent from health registers as is the case for the cancer register. In the case of the data provided for this analysis, all records contained a locality, therefore linkage to the deprivation quintile was complete. Overall results show that age adjusted incidence rates for all the cancers analysed in this chapter have increased in the more recent period. This could partly be explained by increased awareness and better diagnostic tools that are detecting cancers earlier, as well as the introduction of screening programmes. However, the increases may also be related to the impact of changes in lifestyle choices such as cigarette smoking, diet, alcohol consumption, sun exposure, obesity, and physical inactivity and exposure to environmental pollutants.

Unlike the analysis presented for mortality in Chapter 4, the relationship between deprivation and cancer incidence was not consistent across cancer sites and periods. For female breast cancer and colon-rectum cancer, the RII estimates were negative in the two periods, notwithstanding the wide confidence intervals, this suggests that the risk of incidence was higher in the lowest deprivation quintile when compared to the most deprived quintile. This paradoxical pattern by deprivation has been well documented for female breast cancer, with one reason for the pattern being the possibility that people from more deprived areas are less likely to undertake regular breast screening. There are, however, other factors that increase breast cancer incidence in women in lower deprived areas, such as having children at an older age, and higher use of the oral contraceptive pill (Smith *et al.*, 2019). Organised population-based cancer screening programmes were introduced in Malta in 2009, with the breast screening programme being the first to be rolled out. When initially launched, women aged 50-60 were invited every 3 years to participate (Ministry for Health Malta, 2017). The introduction of the programme corresponds with the start of the second period of analysis and may explain the decline in the RII, which became less negative. Incidence rates in the 3rd and 4th quintiles increased in the second period which may be the result of uptake of free screening after the launch of the programme. Notwithstanding this, the aged-standardised incidence rate did not change in the most deprived quintile, possibly indicating that the uptake of screening was not equal across the quintiles. A 2017 study on the determinants of uptake of breast cancer screening in Malta found that fear, negative expectation of the screening experience and embarrassment, were the main barriers for screening uptake amongst participants in the study. The study however also found other barriers such as lack of time, transportation issues and lower family income (Marmarà, Marmarà and Hubbard, 2017). The colorectal screening programme was introduced in 2012 for men and women aged 60-64 who were invited to participate every 2 years (Ministry for Health Malta, 2017), this falls towards the end of the most recent period of analysis and most likely had less impact on the overall incidence rates in this 5 year period when compared to the impact of the introduction of breast screening. This may explain why the RII for the two periods remained stable.

For male prostate and lymphoid cancers, the pattern by deprivation was in opposing directions for the two periods. For male prostate incidence, the RII in the earlier period was considerably negative, mirroring the pattern observed for female breast and colon-rectal cancer. In the most recent period, this became marginally positive. Unlike for female breast cancer, no national screening programme for prostate cancer was introduced in the period under analysis. Researchers have posited that increases in incidence of prostate cancer are partly a result of greater availability of opportunistic screening through prostate-specific antigen blood testing (PSA) (Znaor *et al.*, 2013; Mihor *et al.*, 2020). The pattern observed for lymphoid cancer is more difficult to explain, especially given the unclear aetiology of these cancers. These mixed results are in line with those found by other researchers where no clear link between socioeconomic factors and haematological cancer incidence could be drawn (Mihor *et al.*, 2020).

The results for bronchus-lung cancer incidence present the strongest deprivation gradient in this analysis, especially for the most recent period. For the period 2009-2013, the risk of being diagnosed with bronchus-lunch cancer was 27.5% higher for persons living in more deprived areas were 27.5% compared to the average risk for the total population. This deprivation gradient in bronchus-lung cancer incidence has been documented by other researchers. Inequalities in smoking patterns are said to be the major contributory factor for the higher risk associated with people in more deprived areas, with smoking inequalities accounting for between 40-70% of the increased risk (Kuznetsov *et al.*, 2011; Riaz *et al.*, 2011; Li *et al.*, 2015; Mihor *et al.*, 2020). Occupational exposures may also contribute to the increased risk. Analysis from a large prospective cohort study found that after adjusting for smoking habits and dietary factors, occupational exposure to specific carcinogens explained approximately 14% of the inequalities that remained in lung cancer incidence (Menvielle *et al.*, 2010). Smoking prevalence data is not available prior to 2008, which may not directly relate to the long-term risk exposure within the cohort under study. The smoking prevalence data which is available does however show consistently lower prevalence estimates of daily tobacco smoking for those in the highest education category (Table 5.9).

Table 5.9: Prevalence of daily smoking of tobacco products by sex and educational attainment level[1] (%): 2008, 2014 & 2019

Educational attainment	**Total**			**Males**			**Females**		
	2008	**2014**	**2019**	**2008**	**2014**	**2019**	**2008**	**2014**	**2019**
Total	19.2	20.1	20.6	23.8	23.3	23.6	15.1	17.0	17.2
Pre-primary, primary and lower secondary education (levels 0-2)	18.1	22.9	23.6	24.3	27.3	27.4	13.6	18.9	19.6
Upper secondary and post-secondary non-tertiary education (levels 3 and 4)	21.3	18.7	22.6	25.3	21.3	26.9	17.6	15.6	18.2
First and second stage of tertiary education (levels 5 and 6)	13.8	12.5	13.2	17.7	13.7	14.4	9.3	11.2	11.9

National data regarding exposure to carcinogens at the workplace is not available, however several civil cases have been brought against the Government of Malta by former employees of the Malta Drydocks, seeking damages for known exposure to asbestos. In 2021,

[1] [1]Educational attainment classified using the international standard of education (ISCED). Data for 2008 is based on the 1997 version of the classification, while data for 2014 and 2019 is based on the 2011 update.

the Government of Malta acknowledged this occupational asbestos exposure and launched a process for one-time compensation for the workers or their legal heirs (Government of Malta, 2021).

Like lung cancer, tobacco smoking and certain occupational exposures are the most substantial risk factors for bladder cancer (Jubber *et al.*, 2023), however the analysis showed contradictory findings for the two periods. This may be because of the smaller number of cases of bladder cancer when compared to lung cancer, leading to larger margins of error around the estimates. Notwithstanding this, the most substantial increases in incidence of bladder cancer between the two periods were within the most deprived quintile, a pattern similar to what was seen for lung cancer incidence.

It must be acknowledged that the analysis presented here does have some limitations. Firstly, despite grouping data into five-year periods to be able to analyse larger counts, some of the analysis was still based on relatively small numbers per quintile which led to large margins of error around the results. This made it difficult to draw conclusions on the association between deprivation and cancer incidence. Secondly, the enumeration area (EA) index could not be mapped to the cancer register since EAs are only developed for census operations and are not easily linked to other external data sources, therefore the locality index could only be used. As was shown in Chapter 4, the use of the EA index may have led to stronger gradients in deprivation as it captures pockets of deprivation within localities which may be masked when considering the deprivation quintile for the locality. Finally, additional data, such as stage of diagnosis, was not available in the cancer register, therefore more in-depth analysis could not be conducted to consider other factors which may have influenced the relationship between deprivation and incidence. Despite these limitations, results from the analysis are comparable to those found in other similar analysis. To the author's knowledge, this is the first attempt at analysing the social determinants of cancer incidence in Malta in general, beyond patterning by age, sex, and district of residence.

Chapter 6: Person and place – multi-level analysis of area deprivation and individual level characteristics on selected health outcomes

6.1 Introduction

The analyses presented in Chapter 4 and Chapter 5 aimed to test and validate the use of the small area-based deprivation index created in Chapter 3 for application in the analysis of health inequalities. This was done by assigning a deprivation quintile to individual record level data from two national health registers using locality of residence data routinely collected in the registers. In both analyses, individual level socioeconomic characteristics were not collected as part of the register, therefore the locality level deprivation quintile assigned to the individual was used as a measure of socioeconomic status. When individual level socioeconomic data are available, multi-level modelling can be used to assess whether area level measures of deprivation still contribute to the analysis of health inequalities when adjusting for individual level characteristics. As presented in Chapter 2, similar techniques have been used by other researchers to assess the relationship between area level measures of deprivation and health outcomes, while adjusting for individual level factors.

Since the census file used to build the enumeration area (EA) index in Chapter 3 was linked to mortality register data for analysis in Chapter 4, multi-level analysis of mortality 5 years from the census could be conducted using the EA area deprivation index and individual level socioeconomic characteristics collected in the census. The census also contains a small number of self-reported health measures, such as self-reported prevalence of any chronic condition/long-term illness, allowing analysis to be extended to self-reported health outcomes.

This chapter will therefore present the multi-level analysis of selected health outcomes and mortality by combining two measures of socioeconomic status at two hierarchical levels – area deprivation based on the EA index developed in this research and individual socioeconomic characteristics collected in the census. While the results will also provide an analysis of the relationship between individual level socioeconomic predictors and health, the main aim of this chapter is to assess whether the small area-based deprivation index still contributes to the analysis of health inequalities when considering individual characteristics.

The rest of the chapter will be divided as follows: Section 6.2 will describe the data source used for the multi-level modelling; Section 6.3 will describe the methods used to estimate the outputs, while Section 6.4 will present the results. Finally, Section 6.5 will discuss the findings presented in this chapter.

6.2 Data source

While in Chapter 3, census data from two time points – 2005 and 2011 – were used, the analysis in this chapter was restricted to the 2011 census since this time point represents the most current data source available. The analysis was also restricted to the census population aged 25 to 59, this was done to reduce heterogeneity in the category of inactive persons in relation to employment status. While generally the working age population is defined as the population aged 15 to 64 (National Statistics Office, 2023i), stricter exclusion criteria were applied for the age bracket considered for analysis based on the following:

- The established definition for a dependent child as applied within the census for classification of household type is any person aged less than 18; and any person aged 18 to 24 who is economically inactive and living with at least one parent (National Statistics Office, 2023h). The second category generally covers students still in full-time tertiary education who are not economically active. Based on this definition, a lower age cut-off of 25 was selected.
- Up until 2012, the pensionable age in Malta was 61 for men and 60 for women (Grech, 2016b). Since the 2011 census was being used for the analysis, an upper cut-off of 59 years was taken to exclude persons who retired after having reached pensionable age.

The census and linked mortality data used in Chapter 3 and Chapter 4, were provided as aggregate outputs since micro-data was not necessary for that analysis, therefore this data file could not be used for the multi-level modelling presented in this chapter. A separate data request was therefore made to the National Statistics Office (NSO) to obtain a micro-level data file from the census containing the anonymised individual level data linked to mortality. Anonymised micro-data is provided by the NSO within strict confidentiality rules to minimise the risk of identification of statistical units. Microdata is processed by means of tailored statistical tools developed for the purpose of statistical disclosure control to ensure that even with the removal of identifying fields such as name and surname, the combination of variables requested in the data file does not indirectly lead to the identification of statistical units when there are small counts. When such a risk is still possible because of the combination of variables, for example, if a statistical unit's date of birth, sex and location of residence is so unique it makes that record identifiable, the individual level record is suppressed in the data file provided to the researcher. This can be avoided by restricting the number of fields requested, especially fields that provide potentially identifiable information about the statistical unit, and categorising fields into broad groups to minimise the possibility of small counts.

Therefore prior to making the formal request for the data, the census questionnaire was reviewed in detail to ascertain the essential individual level fields required for the analysis, and categories were created for each field to produce groups that could allow for meaningful comparison but also prevent, as much as possible, small counts. Also, to request micro-level data, a clear justification needs to be provided for each field requested, therefore it was important to ensure that fields included in the request were deemed relevant for the analysis being conducted.

In a multi-level model, data is analysed within a hierarchy, meaning that more than one level is attributed to the individual records, and these levels are said to be nested within each other. In this analysis, individuals (level 1) are nested within EAs (level 2). A description of the fields included in the data request will be presented below, broken down by the levels considered in this analysis.

6.2.1 Enumeration area fields (level 2)

In a multi-level model, a field is required to cluster persons within their level 2 category – in this case the EA. Since the EA code identifies the locality of residence in which a person lives, it is considered a highly identifiable field and including it in the data file along with other variables made it more likely that suppression was required. Since the precise EA was not required for the analysis, it was agreed that the EA would be anonymised into a random number which could be used as a grouping characteristic to nest persons within level 2.

Apart from the anonymised EA code, the additional level 2 field required for the analysis was the deprivation quintile assigned based on the EA index developed in Chapter 3. Since this needed to be linked to the individual census records through their EA code, and the EA code was to be anonymised as described above, it was agreed that the researcher would provide the NSO a list of the EA codes and corresponding deprivation quintiles, and the deprivation quintile would be added to the micro-data by the NSO prior to anonymisation of the file.

6.2.2 Individual fields (level 1)

The individual level fields considered for inclusion were those related to demographic and socioeconomic characteristics. Dwelling characteristics were not considered for the analysis since these characteristics would have constituted another level in which individuals were nested – a mid-way level between the individual and the EA.

When planning the data request, contingencies were outlined a priori, such that if the NSO needed to suppress records of the fields requested, certain fields could be excluded, and categories regrouped to reduce partially, or fully, the suppression of records.

The demographic fields included in the data request were the following:

- **Age**
  Age at time of enumeration. This field was requested categorised in seven 5-year age bands ranging from 25-29 to 55-59. Since this field was considered as an essential predictor to be included in the analysis, it was not considered for exclusion should suppression be required. Wider age bands, such as 10-year age bands were considered as a contingency.

- **Sex**
  Sex was categorised as male and female, and no other categorisation was foreseen for this field. Sex was considered an essential predictor to include in the analysis, therefore it was not considered for exclusion should suppression be required.

- **Citizenship**
  Since only 4.9% of the enumerated population in the 2011 census was non-Maltese (National Statistics Office, 2014b), the citizenship field was not broken down by single country of citizenship, but a binary field was requested. The citizenship field categorised persons as having or not having Maltese citizenship. From the three demographic predictors, this field was considered for possible exclusion should

suppression be required. While citizenship may be an important predictor, the fact that the share of persons who were foreign was just 5% may have made it more likely that this field would lead to small counts when combined with other fields.

The socioeconomic fields included in the data request were the following:

- **Highest level of education**

  In the census, the highest level of education variable was classified using the International Standard Classification of Education (ISCED). This was developed by the United Nations Education, Culture and Science Organisation (UNESCO) and classifies education into nine major categories ranging from 0, no formal education or less than primary to 8, Doctoral or equivalent. Since there may be considerable variation by age group for the number of persons in the lowest and highest education categories, the nine ISCED levels were collapsed into three categories. This was done based on the groupings used by Eurostat, the statistical office of the European Union (EU), and consequently used by the NSO for its comparison across education levels. The groupings were: Low education = ISCED levels 0-2; Medium education= ISCED levels 3-4; High education ISCED levels 5-8 (Eurostat, 2023d). Highest level of education was considered an essential predictor for the analysis, therefore was not considered for exclusion should suppression be required.

- **Employment status**

  Employment status was measured in the census through several questions. First a general question was asked on employment status during the week of the census using seven categories: employed, unemployed, student, retired, unable to work due to an illness or disability, fulfilling domestic tasks, and other. For those who did not work during the week of the census, additional questions were asked regarding the reason for not working and whether they were actively seeking employment and prepared to take up employment, if available. Based on these questions, the International Labour Organisation (ILO) classification was used to group employment status into three categories, employed, unemployed, and inactive. Persons were classified as unemployed if they were unemployed during census week, were available to take up employment within two weeks and were actively seeking a job in the previous four weeks. Those not classified as unemployed or employed were classified as “inactive” or the population outside the labour force (Eurostat, 2023b). Employment status was considered an essential predictor for the analysis, therefore was not considered for exclusion should suppression be required.

- **Literacy**

  Persons were classified as literate if they declared that they were able to read and write a simple sentence, therefore the original field in the Census was categorised as “yes” or “no”. No further recategorization could be conducted on this field. While literacy may be an important predictor, the fact that the share of persons who were classified as illiterate in the census was only 6.4% (National Statistics Office, 2014b), may have made it more likely that this field would lead to small counts when combined with other fields, therefore this field was considered for possible exclusion should suppression be required.

- **Occupation skill level**

  Occupations were classified in the census based on the ILO International Classification of Occupations (ISCO) which groups them into ten major categories based on skill level with 1 being the category assigned to managers and 9 being the category for elementary occupations. Those with any occupation within the armed forces are classified in a separate category numbered as 0. The ILO further groups the major occupation categories into three skill levels – ISCO 1-3 are high skill; ISCO 4-8 are medium skill, and ISCO 9 is low skill. Since the category for armed forces may include a variety of skill levels it is not classified within this grouping (International Labour Office, 2023). This skill categorisation was used to reduce the full ISCO list to three categories. Since this categorisation was selected, armed forces occupations had to be excluded. While it is not possible to assess the impact of excluding this category on the analysis, occupations within the armed forces category accounted for 0.9% of all the occupations within the enumerated population. The majority of those in armed forces occupations were male (95.6%) and aged between 20 and 29 (87.3%) (National Statistics Office, 2014b). While employment status may be an important predictor, those who were classified as unemployed or inactive and never worked, did not have an occupation skill level. This means that a category of “not applicable” had to be included for this field. Since employment status categorised all individuals, and occupation skill level was a subset of this field, occupation skill level was considered for exclusion if suppression was required.

6.2.3 Health outcome fields (level 1)

Three health outcomes were considered for analysis as follows:

- **Mortality five years after the census**
  The linking conducted between the census file and the mortality register for reference years 2011 to 2016 to extract aggregate counts for Chapter 4, was also used to assign mortality status for records in the micro-data file. As outlined in Chapter 4, record linking between the census and mortality register was conducted using the unique identifier available in the census – this is a unique identity number given to all Maltese citizens and foreign nationals who have a residence card. There were 0.4% of records in the census file which did not have an identity card number, in such cases a combination of name, surname and date of birth were used to try to link to the mortality register. It was not possible to ascertain the number of records which would be considered lost to follow-up due to emigration over the five-year period. It is possible that a portion of persons in the census file were not linked to the mortality register because they left the country in this period and some of these may have subsequently died abroad. Between 2012-2016, an estimated 29,289 emigrants left Malta, equivalent to 7% of the total enumerated population in 2011. A portion of the emigrants may also be immigrants who entered the country after the census, therefore the share of emigrants from the enumerated population in 2011 would be expected to be less than 7%. Since the outcome of interest was overall mortality within five years of follow-up, this field was requested as a binary variable categorising records as "yes" if a person died at any point within the five-year follow-up period.

- **Self-reported long-term illness/disease and/or chronic condition**
  Persons enumerated in the 2011 census were asked whether they had any long-term illness, disease and/or chronic condition. In the census a long-term/chronic condition was defined as an illness/disease or condition that may be expected to require a long period of supervision, observation, or care. Examples given in the questionnaire were

asthma, diabetes, and heart disease. Respondents were not asked to specify the condition; therefore, the outcome was binary, categorised as “yes” or “no”.

- **Self-reported mental health condition**
  Persons enumerated in the 2011 census were asked whether they had a mental health condition. Unlike for the question on self-reported chronic conditions, no examples were provided. Respondents were not asked to specify the condition; therefore, the outcome was binary, categorised as “yes” or “no”.

The request for data based on the full list of fields outlined above was made on 8 September 2022 through the formal micro-data request form available on the NSO website. A list of all the fields requested, their categories and justification for their request with respect to the intended analysis, was included with the form. The NSO does not require ethical approval to process anonymised micro-data requests by researchers but makes its decision to provide data to researchers based on the information provided in the application. The researcher was notified on 5 October 2022 that the combination and categories for the fields requested would not require any record suppression and the request could proceed, as is. This means that all the fields included in the request were retained in the file and no exclusions or re-grouping of categories was required from the researcher.

As part of the micro-data request process, researchers are required to sign a contract with the NSO which stipulates, amongst others, the conditions upon which data must be stored, who is able to access the data, how results from the data should be disseminated and the agreed data retention period. Once the contract was duly signed by both parties (researcher and representative of the NSO), a password protected data file was shared, including meta-data, on 6 October 2022. Table 6.1 provides a summary of the data fields included in the data file provided. The 2011 census was weighted to calibrate for the 3.8% under-coverage due to non-response (National Statistics Office, 2014a). The weight for each record was also included in the file to ensure that the population considered in the analysis was the same as that covered by the census. While the meta-data file provided by the NSO included a value label of -1, denoting missing data for the fields “Sex”, “Citizenship”, “Literacy”, “Self-reported long-term illness/disease and/or chronic condition”, and “Self-reported mental health condition”, upon inspection, none of the records had any missing data.

Table 6.1: Description of data fields in 2011 data file used for multi-level modelling analysis

	Type	Categories
Level 2 – Enumeration Area		
Enumeration area code	Numeric	Random number ranging from 1 to 1022
Deprivation quintile	Ordinal categorical	1 = Least deprived 2 = 2nd quintile 3 = 3rd quintile 4 = 4th quintile 5 = Most deprived
Level 1 – Individual		
Age	Ordinal categorical	1 = 25-29 2 = 30-34 3 = 35-39 4 = 40-44 5 = 45-49 6 = 50-54 7 = 55-59
Sex	Binary	1 = Male 2 = Female
Citizenship	Binary	1 = Yes, Maltese 2 = No, not Maltese
Highest level of education	Ordinal categorical	1 = Low education 2 = Medium education 3 = High education
Employment status	Nominal categorical	1 = Employed 2 = Unemployed 3 = Inactive
Literacy	Nominal categorical	1 = Literate 2 = Illiterate
Occupation skill level	Ordinal categorical	1 = High skill 2 = Medium skill 3 = Low skill 4 = Not applicable
Mortality 5 years after census	Binary	1 = Yes 2 = No
Self-reported long-term illness/disease and/or chronic condition	Binary	1 = Yes 2 = No
Self-reported mental health condition	Binary	1 = Yes 2 = No
Weight	Numeric	Weighting factor assigned in the census

As described in Chapter 3, persons assigned to dummy EAs in the census – mainly those living in institutions or records which were imputed – were excluded from the creation of the index. This means that a quintile could not be assigned to any records in the census file with a dummy EA. The population therefore considered in the analysis were persons enumerated in the 2011 census who were aged between 25 and 59 and living within private households that were not assigned to dummy EAs. Apart from this, as described previously, persons categorised within the occupation category 0=Armed Forces, based on the ISCO classification, were also excluded from the file. The total number of records, after applying the calibration weight was 191,896, which is 95.3% of the total population enumerated within the target age group in the 2011 census. The unweighted number of records in the file was 189,859 therefore calibrating the data for non-response added an additional 2,037 persons, an increase of 1.1%.

6.3 Methods

The three health outcomes analysed in this chapter were all binary, therefore logistic regression was chosen as the method for analysis. Since the data was hierarchical, a multi-level framework was used. This was considered appropriate since the health outcomes experienced by the individuals clustered within EAs may have been correlated because of shared context, therefore once including the area-based deprivation quintile as a predictor in the analysis, the independence of observations expected in ordinary logistic regression did not hold.

While it was expected that there was variability between the individual outcomes, some of the variability may be attributed to differences between the EAs. For such analysis it was not necessary that the number of level 1 units clustered within the level 2 units is equal. There were 1,022 EAs at level 2 dividing the 191,896 individuals at level 1. EA size ranged from 10 to 412, with an average size of 187.

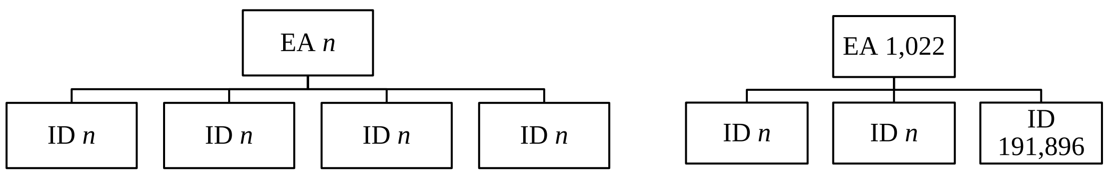


Figure 6.1: Schema for the multi-level framework applied in the analysis

Prior to building the multi-level models, descriptive statistics were extracted to describe the study population according to the individual and area level characteristics described previously. Univariate odds ratios and corresponding confidence intervals were estimated for each level 1 and level 2 predictor against the individual health outcomes. The category deemed as most advantageous was set as the reference to estimate the ratios. This served as a preliminary analysis to understand the relationship between each individual predictor and the health outcomes prior to building the models. The next section describes the modelling strategy applied for the multi-level models.

6.3.1 Modelling strategy

The health outcomes were considered separately, and none were included as predictors within the analyses, therefore three separate models were built – one for each outcome. The strategy applied to build the models was the same for all three outcomes. Multi-level analysis was conducted in MLwiN 3.05.

Two-level models (level 1= individual; level 2= enumeration area) were fitted with a random intercept for each enumeration area to examine the associations between area deprivation and the health outcome before and after adjusting for individual-level demographic and socioeconomic characteristics. In a random intercept multilevel logistic regression model, the probability for success or failure with respect to the binomial outcome (π_{ij}) is modelled as a linear combination of a series of predictors x_{pij} together with a random effect for the higher-level unit u_{0j} as illustrated below.

$$\text{logit}(\pi ij) = \log\left(\frac{\pi ij}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} + \ldots + u_{0j}$$

Taking the example of mortality as the outcome, this is operationalised as follows:

$$\text{logit}(\text{mortality } \pi ij) = \log\left(\frac{\text{mortality } \pi ij}{1-\text{mortality } \pi ij}\right) = \beta_0 + \beta sex_{ij} + \beta age\_group_{1ij} + \beta education_{1ij} + \beta_3 x_{1ij} \ldots + EAu_{0j}$$

The logit-link function was selected since parameter estimates for the predictors were estimated as log odds ratios and therefore, when exponentiated, could be interpreted as odds ratios. Since the outcome data are binary, a binomial distribution was assumed to apply for the variation at level 1 (individual) in the data. MLwiN uses an iterative estimation procedure that involves transforming the data and fitting a linear model. The estimation type used to calculate the predicted values can do so from the fixed part only, marginal quasi-likelihood (MQL), or from both the fixed and random parts of the model, penalised quasi-likelihood (PQL). MQL is the simplest procedure but tends to underestimate the variances and produces biased results, though it is assumed an adequate tool for model building. The rule of thumb approach was followed, whereby the MQL procedure was used initially to build the models, and the PQL procedure was used to extract the results once a final model was established (Leyland and Groenewegen, 2020). In practice, when the PQL procedure was applied to the final models there were only marginal changes to the model outputs compared to those which were extracted using MQL.

A forward stepwise approach was used to build the models. First, the mean model was estimated, this model estimated the mean odds of experiencing the health outcome in the total study population while letting it vary across EAs. Thus, this model could be used to estimate the percentage of variance in the outcome that was explained by differences in EAs, before adjusting for any other covariates. This model was used as one of the reference models (Model 1). Two measures were used to estimate the magnitude of the variance between the EAs. Firstly, the intra-class correlation coefficient (ICC) was approximated as follows, where $\sigma^2 u0$ is the variance for the constant estimated from the model:

$$\text{ICC} = \frac{\sigma^2 u0}{\sigma^2 u0 + 3.29}$$

The second measure was the median odds ratio (MOR). The MOR quantified the variation between the enumeration areas by taking the median of all random pairwise comparisons between two individuals selected randomly from two EAs with identical covariates. The measure showed the extent to which the individual probability of having the health outcome was determined by the person’s EA and was quantified as follows:

$$\text{MOR} = \exp\left(0.954 \sqrt{\sigma^2 u0}\right)$$

The ICC and MOR were estimated for each model produced in the analysis.

The second reference model (Model 2) included age group and sex as control variables and small area deprivation quintile as predictor variable. This model was included as a reference model to provide an estimate of the odds of experiencing the health outcome by small area deprivation quintile when adjusting only for the main demographic characteristics of the study population.

The final model was built by adding the remaining individual level explanatory variables to the second reference model, in a stepwise fashion, including a new individual level explanatory variable at each step. To avoid redundancy in the model, items were only retained based on an assessment of the Wald test for the item as well as the p-values for the pairwise odds ratios for each category within the explanatory variables. Once the final model was estimated (Model 3a), residual plots were extracted and analysed. The final model was also refitted with the small area-based deprivation measure set as a continuous variable type in the modelling software. The variable itself was not changed, with the five quintiles of deprivation retained, however allowing the model to treat it as a continuous variable produced an estimate of the change in the odds of the respective health outcome with every unit change in deprivation quintile (Model 3b).

Since the aim of this analysis was to assess the contribution of the small area-based deprivation measure when considering individual characteristics, a simple model structure was maintained. Random slopes were not included, and interaction terms were only used when deemed necessary to try to explain model outputs.

6.4 Results

6.4.1 Descriptive statistics

Table 6.2 presents descriptive statistics of the study population. The sex distribution was equal. The distribution by age was generally even, with approximately 15% of the population within each 5-year age band. The only exceptions were the age groups 40 to 44 and 45 to 49 with approximately 13% in each category. Only 5% of the population were foreign. With respect to the socioeconomic predictors, 5% were illiterate and 70% were employed. Of those who reported an occupation, 36% were engaged in a medium skill occupation while only 7% were engaged in a low skill occupation. The most common education category was the lowest category with 59% of the population having attained an educational level of ISCED 2 or lower. When considering the area-level deprivation quintiles, distribution across the quintiles was generally similar, ranging from 21% in the least deprived and 2nd quintile, to 19% in the most deprived quintile.

Table 6.2: Descriptive statistics of the study population

	n	%
Total	191,896	100
Sex		
Male	96,072	50.1
Female	95,824	49.9
Age group		
25-29	28,490	14.8
30-34	28,442	14.8
35-39	27,202	14.2
40-44	23,987	12.5
45-49	25,749	13.4
50-54	29,530	15.4
55-59	28,496	14.8
Maltese citizenship		
Yes	182,165	94.9
No	9,731	5.1
Small area deprivation quintile		
Least deprived	39,595	20.6
2	39,474	20.6
3	38,661	20.1
4	37,943	19.8
Most deprived	36,223	18.9
Literacy		
Literate	182,284	95.0
Illiterate	9,612	5.0
Highest education level achieved		
ISCED 5-8 (High)	35,181	18.3
ISCED 3-4 (Medium)	43,018	22.4
ISCED 0-2 (Low)	113,697	59.2
Employment status		
Employed	134,166	69.9
Unemployed	7,865	4.1
Inactive	49,865	26.0
Occupation skill level		
High	56,502	29.4
Medium	69,885	36.4
Low	13,778	7.2
Not applicable	51,731	27.0

ISCED: International standard classification of education

6.4.2 Univariate odds ratios

Table 6.3 presents univariate odds ratios for the individual level 1 and level 2 predictors and the three health outcomes. For all three health outcomes, males had higher odds of experiencing the health outcome when compared to females and this was highest for the mortality outcome, with the odds of males dying within 5 years of the census being 1.67 times greater when compared to females (95% CI 1.52 - 1.84). Age shows a continuous positive gradient across the categories when using the youngest age group (25-29) as reference. Like sex, the largest odds ratios were found for the mortality outcome with the odds of those aged 55-59 dying within 5 years of the census being 17.73 times higher when compared to those age 25-29 (95% CI 12.99 - 24.21). The gradients by age for the two self-reported outcomes were nearly identical. Being foreign was a protective factor when considering this predictor individually and not adjusting for other co-variates. This was especially for the outcome related to self-reported mental health conditions, with the odds of foreigners reporting a mental health condition being 75% lower when compared to the Maltese citizens in the study population (OR 0.25, 95% CI 0.19 - 0.33).

All the socioeconomic predictors showed a positive gradient with the three health outcomes when using the most advantageous category within each predictor as reference. Unlike with the demographic predictors where the largest odds ratios were for the mortality outcome, when considering socioeconomic predictors, the largest univariate odds ratios were seen for the mental health outcome. The odds of those living in the most deprived areas reporting a mental health condition was 2.22 times higher when compared to those in the least deprived areas (95% CI 1.99 - 2.48). For those who were illiterate and in the lowest education category, respectively, the odds of reporting a mental health condition was 6.2 times higher compared to those who were literate (95% CI 5.72 - 6.72) and those in the highest education category (95% CI 5.26 - 7.31). With respect to employment, the odds of those who were inactive reporting a mental health condition was 11.39 times higher when compared to those in employment (95% CI 10.46 - 12.4), while the odds of those in the lowest occupation category reporting a mental health condition was 3.57 times higher when compared to the highest occupation category (95% CI 2.92 - 4.36).

Table 6.3: Univariate odds ratios for the level 1 and level 2 predictors and the three health outcomes

	Mortality		**Chronic health condition**		**Mental health condition**	
	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI
Sex						
Female		Ref. category		Ref. category		Ref. category
Male	1.67	(1.52-1.84)	1.1	(1.07-1.13)	1.11	(1.04-1.19)
Age group						
25-29		Ref. category		Ref. category		Ref. category
30-34	1.81	(1.24-2.65)	1.23	(1.15-1.32)	1.31	(1.08-1.59)
35-39	2.15	(1.48-3.11)	1.67	(1.57-1.79)	1.97	(1.64-2.37)
40-44	4.03	(2.86-5.69)	2.32	(2.17-2.47)	2.78	(2.33-3.32)
45-49	6.32	(4.55-8.78)	3.32	(3.13-3.53)	3.31	(2.79-3.93)
50-54	10.58	(7.71-14.51)	4.81	(4.55-5.1)	4.66	(3.96-5.49)
55-59	17.73	(12.99-24.21)	7.18	(6.78-7.59)	5.84	(4.97-6.85)
Maltese citizenship						
Yes		Ref. category		Ref. category		Ref. category
No	0.76	(0.6-0.97)	0.45	(0.41-0.48)	0.25	(0.19-0.33)
Small area deprivation quintile						
Least deprived		Ref. category		Ref. category		Ref. category
2	0.99	(0.84-1.17)	1.12	(1.07-1.17)	1.3	(1.16-1.47)
3	1.29	(1.1-1.51)	1.12	(1.07-1.16)	1.37	(1.22-1.54)
4	1.38	(1.18-1.61)	1.26	(1.21-1.31)	1.69	(1.51-1.89)
Most deprived	1.92	(1.66-2.23)	1.47	(1.41-1.53)	2.22	(1.99-2.48)
Literacy						
Literate		Ref. category		Ref. category		Ref. category
Illiterate	3.62	(3.18-4.12)	2.63	(2.51-2.75)	6.2	(5.72-6.72)
Highest education level achieved						
ISCED 5-8		Ref. category		Ref. category		Ref. category
ISCED 3-4	1.45	(1.19-1.78)	1.25	(1.19-1.31)	2.42	(2.01-2.91)
ISCED 0-2	2.83	(2.39-3.35)	1.89	(1.82-1.96)	6.2	(5.26-7.31)
Employment status						
Employed		Ref. category		Ref. category		Ref. category
Unemployed	1.95	(1.58-2.42)	1.13	(1.06-1.21)	3.56	(2.96-4.28)
Inactive	2.78	(2.53-3.06)	2.31	(2.25-2.38)	11.39	(10.46-12.4)
Occupation skill level						
High		Ref. category		Ref. category		Ref. category
Medium	1.48	(1.28-1.72)	1.07	(1.04-1.11)	1.53	(1.29-1.81)
Low	2.12	(1.73-2.61)	1.32	(1.25-1.4)	3.57	(2.92-4.36)
Not applicable	3.59	(3.12-4.11)	2.36	(2.28-2.44)	15.01	(13.04-17.28)

CI: Confidence interval; ISCED: International standard classification of education

6.4.3 Multi-level logistic regression

6.4.3.1 Mortality five years from census

Table 6.4 presents the outputs of the multi-level modelling of mortality five years from the census. Overall, 0.9% (95% CI 0.8% - 1.1%) of the study population died within 5 years of enumeration in the 2011 census. Model 1 presents estimates of the ICC and MOR for the empty model. This represents the difference in mortality risk between enumeration areas before adjustment for any co-variates. The ICC shows that before adjustment, 2.5% of the difference in mortality risk could be attributed to differences between the 1,022 enumeration areas. The median of all pairwise comparisons between the EAs was 1.32.

Model 2 shows the model output after including sex, age group and area level deprivation quintile as covariates. Both the ICC and MOR reduced when compared to the empty model. The ICC reduced from 2.5% to 0.8%, while the MOR reduced from 1.32 to 1.16, suggesting that part of the variation between EAs with respect to mortality risk, could be explained by differences in the age and sex structure of the population within the EAs, as well as the area deprivation. After adjusting for age and sex, mortality risk exhibited a gradient by deprivation. While there were no differences between the least deprived and second quintile, a gradient was present from the middle quintile when compared to the least deprived quintile. The odds of those living in the most deprived areas having died within 5 years of the census was 1.89 times higher when compared to those living in the least deprived areas (95% CI 1.63 - 2.21; $p<0.001$). When keeping deprivation constant, the odds of males having died within 5 years of the census was higher compared to females (OR 1.7; 95% CI 1.54 - 1.87; $p<0.001$) and odds of dying followed a clear gradient by age, with increasing odds with increasing age group.

Model 3a shows the final model after inclusion of all individual socioeconomic covariates and citizenship. Maltese citizenship and highest level of education were not significant predictors in the model and were excluded. While the item occupation skill level, overall, did not reach significance based on the Wald test, some of the pairwise comparisons between the categories within the item and the reference category were significant, therefore the item was retained. The inclusion of literacy, employment status and occupation skill level in the model reduced the ICC and MOR only marginally when compared to the reference model which included only age, sex, and deprivation quintile, suggesting that the differences between

EAs adjusted for when including age, sex, and deprivation quintile in model 2 were partially related to differences in individual socioeconomic characteristics. In fact, while the risk pattern by sex and age remained the same, the odds ratio for males compared to females increased when compared to model 2, and all odds ratios for the age groups compared to the reference category, decreased.

Assessment of the individual socioeconomic characteristics showed gradients when comparing the most advantageous to least advantageous categories. When maintaining all other covariates equal, the odds of those who were illiterate having died within 5 years of the census was 1.65 higher when compared to those who were literate (95% 1.43 - 1.89, p<0.001). With respect to employment status, the odds of those who were unemployed having died within 5 years of the census was 1.51 times higher when compared to those who were employed (95% CI 1.18 - 1.94, p<0.001), while the odds of those who were inactive was 2.35 times higher (95% CI 1.45 - 3.82). Finally, occupation skill level also showed a gradient, with the odds of those in a medium skill occupation having died within 5 years of the census being 1.18 times higher when compared to those in a high skill occupation (95% CI 1.01 - 1.38), while the odds of those in the low skill category having died within 5 years was 1.30 times higher (95% CI 1.05 - 1.61, p<0.05).

Even after adjusting for individual socioeconomic characteristics, area level deprivation quintile continued to contribute to the model. Adjusting for the individual characteristics removed the difference between quintile 3 and the least deprived quintile seen in model 2, however the differences between quintile 4 and the most deprived category compared to the least deprived category remained, albeit with smaller odds ratios. When taking into account age, sex and individual socioeconomic predictors, the odds of those living in enumeration areas classified as being in the 4^{th} quintile of deprivation having died within 5 years of the census was 1.18 times higher when compared to those in the least deprived areas (95% CI 1.00 - 1.38, p<0.05), while the odds of those living in enumeration areas which were classified as most deprived having died within 5 years of the census was 1.48 times when compared to those living in the least deprived enumeration areas (95% CI 1.27 - 1.73, p<0.001). Model 3b shows the final model with area deprivation as a continuous rather than categorical predictor. The odds of having died within five years of enumeration in the census increased by 11% (OR 1.11 95% CI 1.08 - 1.15; p<0.001) for each increase in deprivation quintile, even when adjusting for individual level demographic and socioeconomic predictors.

Table 6.4: Multi-level binary logistic regression models for mortality 5 years from census

		Model 1 (empty model)			
ICC		2.5%			
MOR		1.32			
	Model 2	**Model 3a (final model)**			**Model 3b**
	OR (95% CI)	**OR (95% CI)**	**Wald Test Chi**	**df**	**OR (95% CI)**
Sex (ref. category females)					
Males	1.70** (1.54-1.87)	2.77** (2.46-3.11)	295.238**	1	2.76** (2.46-3.10)
Age group (ref. category 25-29)					
30-34	1.82* (1.25-2.66)	1.76* (1.21-2.57)			1.76* (1.21-2.57)
35-39	2.19** (1.51-3.16)	2.04** (1.41-2.96)			2.04** (1.41-2.96)
40-44	4.14** (2.93-5.84)	3.68** (2.61-5.20)	677.443**	6	3.69** (2.61-5.21)
45-49	6.51** (4.69-9.05)	5.40** (3.88-7.51)			5.42** (3.90-7.53)
50-54	10.82** (7.89-14.85)	8.20** (5.97-11.27)			8.22** (5.98-11.29)
55-59	17.81** (13.04-24.31)	12.0** (8.80-16.49)			12.07** (8.82-16.53)
Small area deprivation quintile (ref. category least deprived)					
2	1.02 (0.86-1.21)	0.92 (0.78-1.10)			
3	1.30* (1.11-1.53)	1.14 (0.97-1.34)			
4	1.40** (1.19-1.64)	1.18* (1.00-1.38)	44.544**	4	1.11** (1.08-1.15)
Most deprived	1.89** (1.63-2.21)	1.48** (1.27-1.73)			
Literacy (ref. category literate)					
Illiterate		1.65** (1.43-1.89)	50.132**	1	1.65** (1.43-1.89)
Employment status (ref. category employed)					
Unemployed		1.51** (1.18-1.94)	15.037**	1	1.52* (1.19-1.95)
Inactive		2.35** (1.45-3.82)			2.35* (1.45-3.81)
Occupation skill level (ref. high skill)					
Not applicable		1.46 (0.90-2.38)			1.46 (0.89-2.37)
Low		1.30* (1.05-1.61)	7.763	3	1.29* (1.04-1.60)
Medium		1.18* (1.01-1.38)			1.17* (1.01-1.36)
ICC	0.8%	0.5%			0.7%
MOR	1.16	1.13			1.15

ICC: Intra-class correlation coefficient; MOR: Median odds ratio; OR: Odds ratio; CI: Confidence interval; df: Degrees of freedom
Maltese citizenship and highest level of education achieved were not significant and were excluded from the final model.
*** P<0.001; *P<0.05*

Figure 6.2 shows the residual plot for the enumeration areas for the final model (model 3a). Since the residuals from the model are on a log scale they were converted to odds ratios. The odds ratios reflect the difference between the individual EAs and the average EA. All the 95% confidence intervals around the residual odds ratios for the 1,022 had a range that surpassed 1, on either side, suggesting that the variance found in mortality risk between EAs was accounted for by the predictors included in the final model.

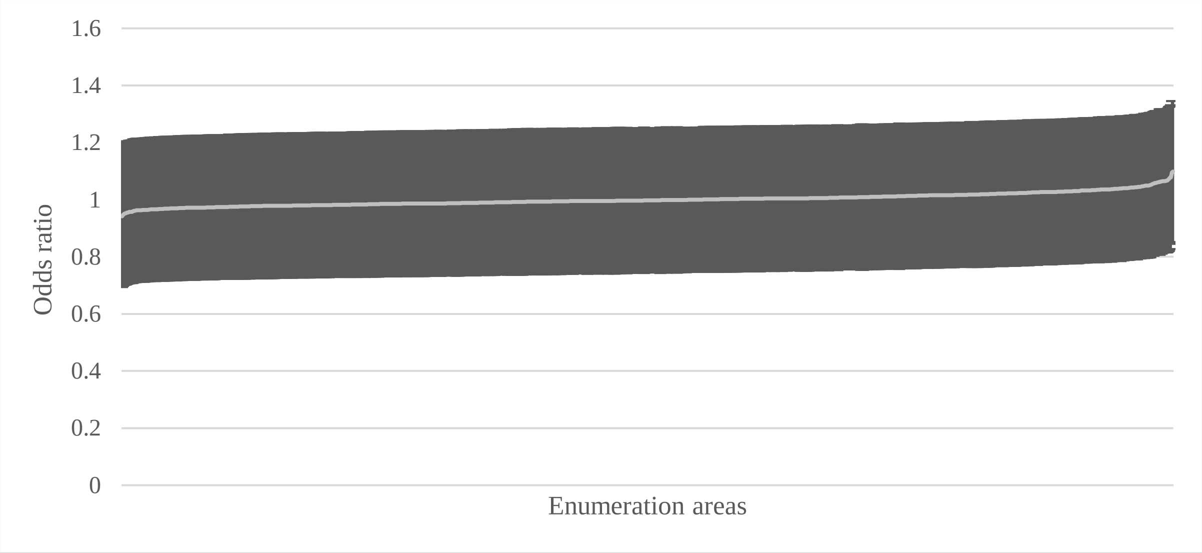


Figure 6.2: Residual plot for final model (3a) for mortality five years from censu

6.4.3.2 Self-reported chronic health condition

Table 6.5 presents the outputs of the multi-level modelling of the odds of reporting a chronic health condition. Overall, 14.6% (95% CI 13.4% - 15.9%) of the study population reported having a chronic health condition at the time of enumeration in the census. Model 1 presents estimates of the ICC and MOR for the empty model. This represents the difference between enumeration areas before adjustment for any co-variates. The ICC showed that before adjustment, 1.8% of the difference in self-reported prevalence of a chronic health condition could be attributed to differences between the 1,022 enumeration areas. The median of all pairwise comparisons between the EAs was 1.27.

Model 2 shows the model output after including sex, age group and area level deprivation quintile as covariates. Both the ICC and MOR reduced only marginally when compared to the empty model. The ICC reduced from 1.8% to 1.3%, while the MOR reduced from 1.27 to 1.22, notwithstanding the marginal decline, all predictors included in the model were significant. After adjusting for age and sex, the odds of reporting a chronic condition increased along a deprivation gradient. Unlike for mortality, the gradient started from the second quintile, though the odds compared to the least deprived quintile were the same for the second and third quintile. The odds of those living in the most deprived area having reported a chronic health condition was 1.50 higher when compared to those in the least deprived areas (95% CI 1.41 - 1.59; p<0.001). The odds for males having reported a chronic health condition was also higher when compared to females (OR 1.12; 95% CI 1.09 - 1.15; $p<0.001$) and the odds of having reported a chronic health condition followed a clear gradient by age, with increasing odds with increasing age.

Model 3a shows the final model after inclusion of all individual socioeconomic covariates and citizenship. Occupation skill level was not a significant predictor in the model and was excluded. The inclusion of the individual level socioeconomic predictors and citizenship to the model had little impact on the ICC and MOR when compared to model 2 which adjusted for age, sex, and deprivation quintile only, suggesting that part of the differences seen between EAs when adjusting for these covariates in model 2, were related to differences in individual socioeconomic characteristics. In fact, while the risk pattern by sex and age remained the same, the odds ratio for males compared to females increased when compared to model 2, and all odds ratios for the age groups compared to the reference category, decreased.

Not having Maltese citizenship was a protective factor, with the odds of this population group having reported a chronic health condition was 53% lower when compared to those with Maltese citizenship (OR 0.47 95% CI 0.43 - 0.51, $p<0.001$). For the individual socioeconomic characteristics, odds of reporting a chronic health condition increased when comparing the most advantageous to least advantageous categories, except for highest level of education achieved. While for the second highest education category (ISCED 3-4) there was no difference compared to the highest education category (ISCED 5-8), being in the lowest category appeared to be protective with respect to reporting chronic health conditions. The odds of those in the lowest education category (ISCED 0-2) having reported a chronic health condition was 13% lower when compared to those in the highest education category (ISCED 5-8) (OR 0.87 95% CI 0.83 - 0.91, $p<0.001$). Since this did not follow the typical gradient seen for the other socioeconomic characteristics, further analysis was conducted.

Firstly, the field of highest level of education was removed from the final model. Removal of the field had little impact on the model outputs, only attenuating the odds ratios marginally. Secondly, the final model was run again including the interaction of age group and highest level of education achieved. While the inclusion of the interaction term eliminated the significance of the comparison between the lowest education category (ISCED 0-2) and the highest education category (ISCED 5-8) (OR 0.92, 95% CI 0.82 - 1.05, $p=0.22$), the interaction term itself was not significant in the model (Wald Chi-squared 17.804, df 12; $p=0.122$) and none of the pairwise interaction terms compared with the reference category were significant. The final model with the highest education field, excluding the interaction term, was retained.

The odds of those who were illiterate having reported a chronic health condition was 1.66 times higher when compared to those who were literate (95% 1.58 - 1.75, $p<0.001$). With respect to employment status, there was no difference between the unemployed and employed categories, while the odds of those who were inactive having reported a chronic health condition was 2.22 times higher (95% CI 2.14 - 2.30).

Even after adjusting for individual socioeconomic characteristics, area level deprivation continued to contribute to the model. Adjusting for the individual characteristics did not remove any of the differences between all the deprivation quintiles and the reference category, though the odds ratios decreased for all pair-wise comparisons. When considering age, sex, citizenship, and individual socioeconomic predictors, the odds of those living in enumeration areas classified as most deprived having reported a chronic health condition was 1.37 times higher when compared to those living in the least deprived enumeration areas (95% CI 1.16 - 1.45, $p<0.001$). Model 3b shows the final model output including area deprivation as a continuous rather than categorical predictor. After adjusting for all individual covariates, the odds of having reported a chronic health condition increased by 8% (95% CI 1.06 - 1.09; $p<0.001$) for every increase in deprivation quintile.

Table 6.5: Multi-level binary logistic regression models for self-reported chronic health condition

	Model 1 (empty model)				
ICC		1.8%			
MOR		1.27			
	Model 2	**Model 3a**			**Model 3b**
	OR (95% CI)	**OR (95% CI)**	**Wald Test Chi**	**df**	**OR (95% CI)**
Sex (ref. category females)					
Males	1.12** (1.09-1.15)	1.54** (1.50-1.59)	727.208**	1	1.54** (1.50-1.59)
Age group (ref. category 25-29)					
30-34	1.23** (1.15-1.32)	1.20** (1.12-1.29)	6542.803**	6	1.20** (1.12-1.29)
35-39	1.69** (1.58-1.80)	1.61** (1.51-1.72)			1.61** (1.51-1.72)
40-44	2.36** (2.21-2.51)	2.19** (2.05-2.33)			2.19** (2.05-2.33)
45-49	3.42** (3.22-3.63)	3.03** (2.85-3.22)			3.03** (2.85-3.22)
50-54	4.91** (4.64-5.21)	4.12** (3.88-4.37)			4.12** (3.88-4.37)
55-59	7.25** (6.85-7.67)	5.70** (5.38-6.04)			5.70** (5.37-6.04)
Small area-based deprivation quintile (ref. category least deprived)					
2	1.15** (1.09-1.23)	1.12** (1.05-1.18)	130.288**	4	1.08** (1.06-1.09)
3	1.15** (1.08-1.22)	1.10* (1.03-1.16)			
4	1.31** (1.24-1.39)	1.23* (1.16-1.31)			
Most deprived	1.50** (1.41-1.59)	1.37** (1.16-1.45)			
Maltese citizenship (ref. category yes)					
No		0.47** (0.43-0.51)	334.334**		0.47** (0.43-0.51)
Literacy (ref. category literate)					
Illiterate		1.66** (1.58-1.75)	396.792**		1.66** (1.58-1.75)
Occupation status (ref. category employed)					
Unemployed		1.05 (0.98-1.13)	2065.261**	2	1.05 (0.98-1.13)
Inactive		2.22** (2.14-2.30)			2.22** (2.14-2.30)
Highest level of education (ref. category ISCED 5-8 High)					
ISCED 3-4		0.98 (0.93-1.03)	60.942**	2	0.98 (0.93-1.03)
ISCED 0-2		0.87** (0.83-0.91)			0.87** (0.83-0.91)
ICC	1.3%	1.2%			1.2%
MOR	1.22	1.21			1.21

ICC: Intra-class correlation coefficient; MOR: Median odds ratio; OR: Odds ratio; CI: Confidence interval; df: Degrees of freedom ISCED: International standard classification of education
Occupation skill level was not significant and was excluded from the final model.
** P<0.001; *P<0.05

Figure 6.3 shows the residual plot for the enumeration areas from the final model (model 3a). Since the residuals from the model are on a log scale they were converted to odds ratios. The odds ratios reflect the difference between the individual EAs and the average EA as extracted from the final model. The majority of the 95% confidence intervals around the residual odds ratios for the 1,022 enumeration areas had a range that surpassed 1 on either side

(96.6%), suggesting that the variance found in the odds of reporting a chronic health condition between EAs was accounted for by the predictors included in the model.

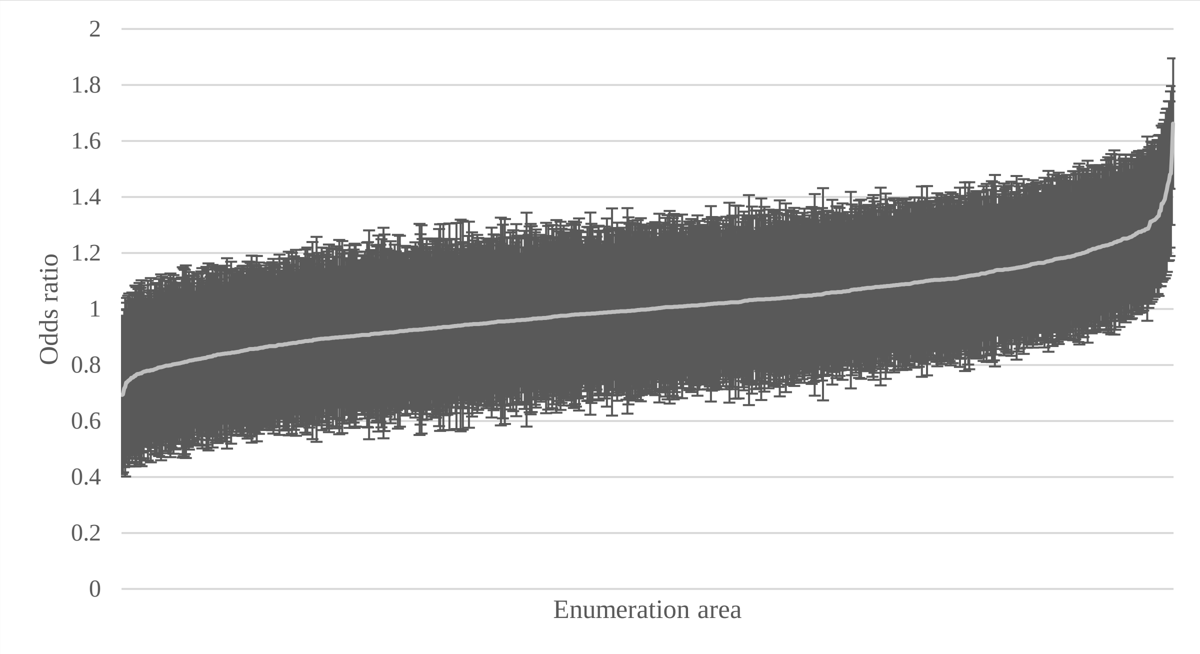


Figure 6.3: Residual plot for final model (3a) for self-reported chronic health condition

6.4.3.3 Self-reported mental health condition

Table 6.6 presents the outputs of the multi-level modelling of the odds of reporting a mental health condition. Overall, 1.8% (95% CI 1.6% - 2.1%) of the study population reported having a mental health condition at the time of enumeration in the census. Model 1 presents estimates of the ICC and MOR for the empty model. This represents the difference between enumeration areas before adjustment for any co-variates. The ICC showed that before adjustment, 5.8% of the difference in self-reported prevalence of a mental health condition could be attributed to differences between the 1,022 enumeration areas. The median of all pairwise comparisons between the EAs was 1.54.

Model 2 shows the model output after including sex, age group and area level deprivation quintile as covariates. Both the ICC and MOR reduced when compared to the empty model. The ICC reduced from 5.8% to 3.8%, while the MOR reduced from 1.54 to 1.41, suggesting that some of the variation between EAs with respect to self-reported prevalence of mental health conditions, could be explained by differences in the age and sex structure of the population within the EAs, as well as the area deprivation. After adjusting for age and sex, the odds of having reported a mental health condition followed a gradient by deprivation. Unlike for mortality, the gradient started from the second quintile. Compared to those living in the least deprived area, the odds of those living in the most deprived area having reported a mental health condition was 2.23 times higher (95% CI 1.95 - 2.54; p<0.001). The odds for males was also higher compared to females (OR 1.12; 95% CI 1.05 - 1.20; p<0.05) and the odds of reporting a mental health condition followed a clear gradient by age, with increasing odds with increasing age.

Model 3a shows the final model after inclusion of all other individual socioeconomic covariates and citizenship. All individual level socioeconomic predictors and citizenship were significant and therefore retained in the final model. Their inclusion in the model reduced the ICC and MOR only marginally when compared to the reference model which included only age, sex, and deprivation quintile, suggesting that part of the differences seen between age, sex, and deprivation quintile, in model 2 were related to differences in individual socioeconomic characteristics. In fact, while the risk pattern by sex and age remained the same, the odds ratio for males compared to females increased when compared to model 2, and all odds ratios for the age groups compared to the reference category, decreased.

Not having Maltese citizenship was a protective factor, the odds of this population group having reported a mental health condition was 76% lower compared to those with Maltese citizenship (95% CI 0.18 - 0.32, p<0.001). For the individual socioeconomic characteristics, odds of reporting a mental health condition increased when comparing the most advantageous to least advantageous categories, except for highest level of education achieved. While for education, both categories exhibited higher odds of having reported a mental health condition compared to the highest education group (ISCED 5-8), the higher odds were within the middle category (ISCED 3-4) at 1.32 (95% CI, 1.09-1.62; p<0.05) rather than the lowest education category (ISCED 0-2) which had an odds ratio of 1.23 (95% CI 1.01 - 1.49; p<0.05). The confidence intervals for both estimates did however overlap, suggesting the increased odds for both education categories compared to the highest education group, were similar.

The odds of those who were illiterate having reported a mental health condition was 2.84 times higher compared to those who were literate (95% 2.58 - 3.11, p<0.001). With respect to employment status, the odds of those who were unemployed having reported a mental health condition was 2.76 times higher compared to those who were employed (95% CI 2.25 - 3.40, p<0.001), while the odds of those who were inactive was 22.31 times higher (95% CI 14.37 -

34.61). Finally, for occupation skill level, low skill occupation was the only category, which was significantly different from the high skill category, the odds of those in a low skill occupation having reported a mental health condition was 1.89 times higher (95% CI 1.52 - 2.36, $p<0.001$).

Even after adjusting for individual socioeconomic characteristics, area level deprivation quintile continued to contribute to the model. Adjusting for the individual characteristics removed the differences between quintiles 2 and 3 when compared to the least deprived quintile seen in model 2, however the difference between quintile 4 and the most deprived category to the least deprived category remained, albeit with smaller odds ratios. When taking into account age, sex, citizenship and individual socioeconomic predictors, the odds of those living in enumeration areas classified as being in the 4^{th} quintile of deprivation having reported a mental health condition was 1.23 times higher compared to those in the least deprived areas (95% CI 1.07 - 1.41, $p<0.05$), while the odds of those in enumeration areas which were classified as most deprived having reported a mental health condition was 1.30 times higher when compared to those living in the least deprived enumeration areas (95% CI 1.14 - 1.49, $p<0.001$). Model 3b shows the final model output including area deprivation as a continuous rather than categorical predictor. After adjusting for all individual covariates, the odds of reporting a mental health condition increased by 7% (95% CI 1.03 - 1.10; $p<0.001$) for every increase in deprivation quintile.

Table 6.6: Multi-level binary logistic regression models for self-reported mental health condition

	Model 1 (empty model)				
ICC		5.8%			
MOR		1.54			
	Model 2	**Model 3a (final model)**			**Model 3b**
	OR (95% CI)	**OR (95% CI)**	**Wald Test Chi**	**df**	**OR (95% CI)**
Sex (ref. category females)					
Males	1.12* (1.05-1.20)	3.93** (3.62-4.26)	1079.206**	1	3.93** (3.62-4.26)
Age group (ref. category 25-29)					
30-34	1.32* (1.08-1.60)	1.20 (0.98-1.47)	145.68**	6	1.20 (0.98-1.47)
35-39	2.01** (1.68-2.42)	1.65** (1.37-2.00)			1.65** (1.37-2.00)
40-44	2.88** (2.41-3.43)	2.11** (1.76-2.53)			2.11** (1.76-2.53)
45-49	3.44** (2.90-4.08)	2.05** (1.72-2.45)			2.05** (1.72-2.45)
50-54	4.76** (4.04-5.60)	2.23** (1.89-2.65)			2.23** (1.88-2.65)
55-59	5.84** (4.97-6.85)	2.04** (1.73-2.41)			2.04** (1.72-2.41)
Small area deprivation quintile (ref. category least deprived)					
2	1.35** (1.17-1.55)	1.11 (0.96-1.28)	19.7*	4	1.07** (1.03-1.10)
3	1.40** (1.22-1.61)	1.07 (0.93-1.23)			
4	1.75** (1.53-2.01)	1.23* (1.07-1.41)			
Most deprived	2.23** (1.95-2.54)	1.30** (1.14-1.49)			
Maltese citizenship (ref. category yes)					
No		0.24** (0.18-0.32)	89.63**	1	0.24** (0.18-0.32)
Literacy (ref. category literate)					
Illiterate		2.84** (2.58-3.11)	476.981**	1	2.84** (2.59-3.12)
Occupation status (ref. category employed)					
Unemployed		2.76** (2.25-3.40)	204.043**	2	2.77** (2.25-3.40)
Inactive		22.31** (14.37-34.61)			22.29** (14.35-34.61)
Highest level of education (ref. category ISCED 5-8 High)					
ISCED 3-4		1.32* (1.09-1.62)	7.794*	2	1.32* (1.08-1.61)
ISCED 0-2		1.23* (1.01-1.49)			1.23* (1.01-1.49)
Occupation skill level (ref high skill)					
Not applicable		1.01 (0.64-1.59)	45.87**	3	1.01 (0.64-1.59)
Low		1.89** (1.52-2.36)			1.89** (1.52-2.35)
Medium		1.07 (0.89-1.29)			1.07 (0.89-1.29)
ICC	3.8%	3.5%			3.5%
MOR	1.41	1.39			1.39

ICC: Intra-class correlation coefficient; MOR: Median odds ratio; OR: Odds ratio; CI: Confidence Interval; df: Degrees of freedom; ISCED: International standard classification of education
*** P<0.001; *P<0.05*

Figure 6.4 shows the residual plot for the enumeration areas from the final model (model 3a). Since the residuals from the model are on a log scale they were converted to odds ratios. The odds ratios reflect the difference between the individual EAs and the average EA as extracted from the final model. The majority of the 95% confidence intervals around the residual odds ratios for the 1,022 enumeration areas had a range that surpassed 1 on either side (98.6%), suggesting that the variance found in the odds of reporting a mental health condition between EAs was accounted for by the predictors included in the model.

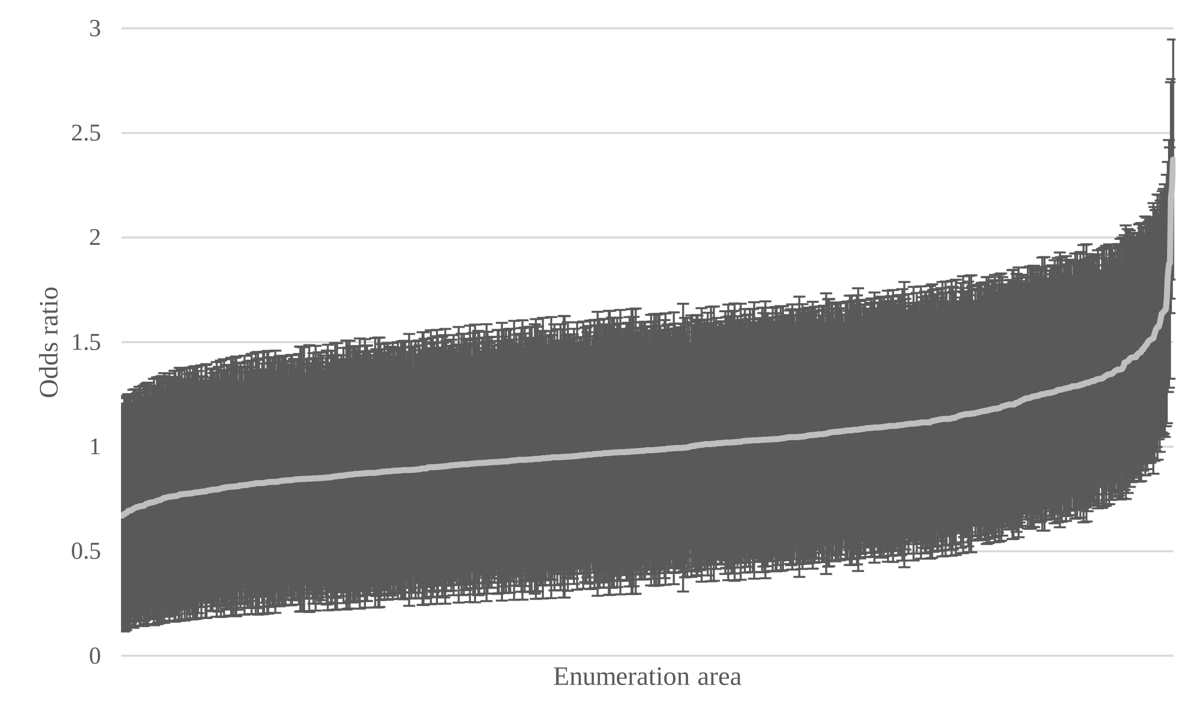


Figure 6.4: Residual plot for final model (3a) for self-reported mental health condition

6.5 Discussion

This chapter sought to use multi-level modelling to analyse whether the area-based measure of deprivation developed in this thesis still contributed to the analysis of health inequalities when individual level characteristics were included in the analysis. Data from the 2011 census was used, with analysis restricted to the private household resident population aged 25 to 59. Individual level demographic and socioeconomic predictors were extracted from the census data, while the area-based deprivation quintile assigned to each EA as described in Chapter 3, was linked to the EA code available in the census. Three binary health outcomes were selected for the analysis – mortality within five years of enumeration in the census, which was extracted by linking the individual census data to the national mortality register, and two self-reported health outcomes taken from the census data itself, prevalence of a chronic health condition and prevalence of a mental health condition. Since all outcomes were binary, multi-level binary logistic regression was used to analyse the data. While this analysis mainly served to assess the impact of including individual characteristics along with area-based measures of deprivation in health inequalities analysis, the analysis also provided a study of the relationship between individual socioeconomic predictors and the three health outcomes.

The results from this chapter have illustrated several key points. Firstly, the small area-based deprivation index developed in this research still contributed to the analysis of health inequalities even when adjusting for individual level socioeconomic predictors. For all three health outcomes, a gradient was still seen with increasing deprivation, though the odds were attenuated when including individual predictors. The strongest relationships were maintained for the most deprived quintile, for all three outcomes. The increased odds for those living in the most deprived areas compared to those living in the least deprived areas ranged from 30% for the prevalence of self-reported mental health condition and 37% for the prevalence of self-reported chronic health conditions, to 48% for mortality within five years of enumeration in the census. This was after adjusting for all significant individual level socioeconomic and demographic characteristics. The low variances remaining in the models after adjusting for all individual and area predictors, suggested that the variation seen between the EAs for the health outcomes analysed was explained by the predictors included in the models.

Other studies using multi-level techniques to analyse mental health outcomes in relation to area and individual level characteristics, presented similar results, with increased risk seen in the most deprived areas even when adjusting for individual factors (Skapinakis *et al.*, 2005; Fone *et al.*, 2007, 2014). Researchers attempting to understand the association between area or neighbourhood deprivation and mental health, posit that part of the association may be related to differences in the social environment; such as lower social cohesion, trust and social capital in more deprived areas (Visser *et al.*, 2021). A meta-analysis of 18 multi-level studies which used area-level measures to analyse mortality risk when controlled for individual socioeconomic status, found that those living in areas defined as having a low area level socioeconomic status experienced higher mortality than those living in higher level socioeconomic areas (Meijer *et al.*, 2012). The findings presented in this chapter related to mortality are consistent with these results. The authors proposed that the impact of neighbourhoods on individual mortality may be related to the fact that people living in the same area affect each other's health behaviours through shared norms and values. This may also extend to the physical environment, where the inhabitants of an area may impact upon the presence of services, businesses and activities in the area that are health promoting. There may also be differences between rural and urban areas which all have implications on population density, exposure to pollution, access to open and green spaces and safety. No measures related to the social and environmental characteristics of the EAs were available for inclusion in this analysis, so while conceptually all these factors may be contributing to the relationship seen between area deprivation and health, even when considering individual characteristics, further research is required to understand the exact dynamics at play, though the findings of this study serve as a necessary springboard for future analysis.

Secondly, the results presented in this chapter provide an analysis of individual level socioeconomic predictors and health. While this was not the specific aim of the analysis, the inclusion of individual predictors allowed for an assessment of individual level social determinants of health. The fact that this analysis was conducted on census data means that it is assumed that it included the entire population and not a sample, as would be the case for traditional surveys. Furthermore, the analysis also included the individual level linking of census data to register-based mortality data, which allowed for the comparison of the analysis of social determinants of health using outcomes measured using administrative data and self-report. To the author's knowledge, this is the first attempt in Malta to use census data to analyse individual level social determinants of health and the first attempt to link register-based mortality data to individual level socioeconomic predictors for health inequalities analysis. Previous analyses of social determinants of health have been restricted to sample surveys and when analysis was done in relation to administrative data, such as mortality rates, it was restricted to predictors available in the register – mainly age, sex and district of residence (Deguara, England and Azzopardi Muscat, 2017).

For all three health outcomes, when adjusting for all other covariates, employment status and literacy consistently presented a risk gradient, with higher odds seen in the least advantageous category compared to the most advantageous category. For literacy, the highest odds were for prevalence of a self-reported mental health condition, with the odds of those who were illiterate having reported a mental health condition being 2.84 times higher compared to those who were literate. The odds ratios for the prevalence of a self-reported chronic health condition and mortality within five years of the census were lower and nearly equal, at 1.66 and 1.65, respectively. The relationship between literacy and poorer health outcomes have been well documented. Most studies are cross-sectional, like the analysis presented here, and therefore it is difficult to establish causality, however there are several possible reasons why being illiterate leads to poorer outcomes. For mental health specifically, illiteracy may lead to stigma and social exclusion and limit access to employment opportunities. This may impact upon an individual's feelings of self-esteem and self-confidence which all may affect their psychological well-being (Hunn, Teague and Fisher, 2023). There may also be some overlap between literacy and certain emotional and behavioural difficulties in childhood, such as attention deficit hyperactivity disorder (Maughan and Carroll, 2006). For health outcomes in general, literacy may impact upon individuals' health seeking behaviour, participation in screening and prevention activities, and treatment compliance (DeWalt *et al.*, 2004). The relationship between general literacy and poorer health outcomes may partly be related to poorer health literacy amongst those who are illiterate (Nutbeam and Lloyd, 2021). A 2011 health literacy survey conducted in Malta amongst those aged 18 and over found that 45.8% of the population had limited (problematic or inadequate) health literacy based on the general health literacy index. The study found that 56.3% of those who reported a long term illness had limited health literacy, however this estimate was not adjusted for age and sex, therefore there may be confounding present as the same study found that 54.4% of those aged 76 and over also had limited health literacy (Office of the Commissioner for Mental Health, 2014).

The association between employment status and poorer health outcomes has also long been established. The two processes that may be influencing this relationship are health selection, where healthier people are more likely able to obtain and retain employment, and the positive health benefits of employment and job security. Therefore the causal relationship may take two opposing directions, poorer health may impact a person's employment status or employment status may cause poorer health (Bartley, Ferrie and Montgomery, 2006; Avendano and Berkman, 2014). Analysing cross-sectional self-reported health outcomes makes it difficult to ascertain the direction of the relationship seen between employment status and morbidity. It is likely that the association observed is a result of a combination of both pathways. Those reporting a chronic or mental health condition may have been more likely to be unemployed or inactive due to their health status. For the prevalence of a self-reported chronic condition, there was no difference between the unemployed and employed categories suggesting there was no increased odds of having reported a chronic health condition between those who were employed or unemployed, when adjusting for all other covariates. On the other hand, there was an increase in odds in those who were inactive compared to the employed category. Since the inactive category contains persons who were unable to work due to a health condition or disability, the increased odds seen here may be related to health selection. For the prevalence of a mental health condition, both categories, unemployed and inactive had increased odds when compared to the employed category. This may suggest that both processes are at play. While it is possible that persons with a mental health condition are more likely to be inactive or unemployed and unable to maintain employment due to their mental health condition, especially considering the very large odds ratio found in this analysis when comparing the inactive category to the employed category (OR 22.31), it is also possible that those who are inactive or unemployed have poorer mental health due to their employment status. Being unemployed and experiencing job insecurity may lead to increased stress. Being excluded from employment also means that individuals do not experience the nonfinancial benefits of work, such as opportunities for social contact, a structured daily routine, and a sense of belonging and identity. The relationship found between employment status and mortality may partly be due to health status. The two self-reported health outcomes in this study were not included as predictors in the model, so it is unclear how the relationship between employment and mortality may be attenuated had health status been considered. It is also possible that the increased odds of mortality seen by employment may be related to the impact of job loss and job insecurity on increased mortality because of specific causes of death, such as suicide and deaths brought on because of certain risk-taking behaviours such as drug and alcohol use. Since cause of death data was not available, it is not clear if this contributed to the increased odds for certain employment status categories.

The results for the two other individual level socioeconomic predictors, occupation skill level and highest level of education achieved, were less consistent across the three health outcomes. Occupation skill level was a significant individual level predictor for mortality and the prevalence of a mental health condition, but not for the prevalence of a chronic health condition. For mortality, the odds of dying within five years of the census was higher for individuals in occupations classified as both low and medium skill level when compared to the high skill level category. For the prevalence of a mental health condition, a higher odds ratio was only seen for the low skill category. Similarly, as for employment status, the relationship between occupation skill level and morality may partly be due to health status, but also related to the employment conditions of lower skilled occupations. Persons in lower skilled occupations may be more likely to experience accidents at the workplace due to the nature of their employment. When considering mental health, it is possible that those with a mental health condition face barriers to access employment due to the stigma of having a mental health illness and therefore have limited upward job mobility. However, it is also plausible that those in low skilled occupations are exposed to stress due to precarious work conditions, hazardous working conditions and greater financial strain due to low wages which may all have an impact upon mental health status. Results for highest education level achieved are inconclusive. Education level had no impact upon mortality within five years of the census. For self-reported mental health, having a lower education level increased the odds of reporting a mental health condition when compared to the highest education category. For chronic health, being in the lowest educational category seemed to have a protective effect with individuals in this category having lower odds of having reported a chronic health condition compared to those in the highest education category. Though additional analysis was conducted to try to explain this contradictory result, this was also inconclusive.

Finally, patterns exhibited for the key demographic features of the study population, namely age, sex, and citizenship, were generally consistent across outcomes. For all three health outcomes, increased odds were seen with increasing age when adjusting for all area and individual level predictors, as would be expected. Increased odds were seen across all health outcomes when comparing males to females. While this pattern was expected for mortality and self-reported chronic health, it is the inverse of what was expected for mental health conditions since women generally report poorer mental well-being when compared to men (Cabezas-Rodríguez, Utzet and Bacigalupe, 2021). Results from the European Health Interview Survey (EHIS) conducted in Malta in 2014 found that prevalence of depressive symptoms in females aged 18 to 64 was higher than that for males, at 5.6% and 4.0% respectively (Eurostat, 2014), therefore these results contrast with the pattern seen in other local studies. There is however a methodological difference between the EHIS and the census. While the EHIS is an individual sample survey, the census is a household survey where it is common for a single household member, usually the reference person, to complete the questionnaire for all members of the household. According to the 2011 census report, within enumerated private households, the majority of reference persons, 74.1%, were male (National Statistics Office, 2014b). It is unclear if this in any way influenced the results seen by sex for self-reported mental health condition. With respect to citizenship, there were no differences between those with Maltese citizenship and those without Maltese citizenship for mortality five years after the census. For the two self-reported health outcomes, being foreign seemed to be protective. This pattern was only seen for the self-reported outcomes and not mortality. It is possible that this is explained by reporting bias with underreporting of self-reported health conditions amongst the foreign population, but it is also possible this is a healthy migrant effect.

It must be acknowledged that the analysis presented here does have some limitations. Firstly, the modelling strategy, a priori, excluded the inclusion of interaction terms and random slopes. This decision was taken to limit the complexity of the models. Since the aim of the analysis was to assess whether the area-based deprivation index remained significant even with the inclusion of individual predictors, complex models were deemed out of scope. The inclusion of interaction terms may have provided a better understanding of the associations found in the analysis, while excluding random slopes in the model assumed that the relationship between the predictors and outcomes did not vary across the EAs. Future research could delve deeper into the results presented here by including such interaction terms and random slopes. Secondly, the relationships seen for some of the predictors seem contradictory to what was expected, this is specifically the case for the relationship between highest educational level achieved and prevalence of self-reported chronic health condition, and sex and prevalence of self-reported mental health condition. It is unclear if the patterns exhibited are real associations or possible issues related to the methodology used in the census. Self-reported health outcomes may face issues of response bias which are difficult to quantify. Since most of the predictors in the self-reported outcomes were consistent with the mortality outcome, which was based on an administrative source, it can be assumed that the health outcomes based on self-report used in this analysis were adequate measures of health status. Finally, categorisation of data fields was decided before the request for data was conducted. This was a requirement to ensure that data could be processed within the confidentiality rules employed by the NSO with respect to micro-data and in fact avoided the need for any record suppression. While, as much as possible, categorisation of the fields was done based on standard classifications that had already been used for data outputting and analysis in Malta, it is possible that the analysis may have benefited from different categories or possibly more detailed breakdown of certain fields. Despite these limitations, results from the analysis are comparable to those found in other similar analysis, and show that even when including individual socioeconomic predictors, the area-based deprivation index provided additional information with respect to the patterning of health inequalities in Malta. The fact that the analysis was conducted on census data means that the results presented here cover the entire population within the target group. To the author's knowledge this is the first attempt to use census data to analyse individual and area level social determinants of health in Malta.

Chapter 7: Discussion

7.1 Introduction

Researchers have presented a wealth of evidence showing that socioeconomic inequalities in health do exist, and improvements in population health outcomes may require that health systems address other risk factors apart from clinical or lifestyle (Marmot, 2001). The acknowledgment of the importance of health inequalities, has helped to highlight the need for more research in this field (World Health Organisation, 2008). However, measuring and understanding socioeconomic differences in health through regular monitoring, may not be easy. Much of the research in health inequalities has made use of self-reported data from cross-sectional health surveys which may suffer from low response, response bias and may be expensive and time consuming to conduct on a regular basis. Surveys may also underrepresent or exclude certain subsets of the population. While countries may have alternative data sources, such as health registers and administrative sources, these may not routinely collect any data on individual level socioeconomic characteristics.

The adoption of small area-based deprivation indices may enhance the measurement and monitoring of health inequalities by increasing the opportunities for the use of health data sources available within health systems (Hosseinpoor and Bergen, 2016). Such indices can be linked to any data source where the same geographical breakdown is available, thus facilitating the analysis of any outcome measure by area level deprivation. While these indices may be prone to ecological fallacy, they have proven useful tools in the study of health inequalities, especially when data on individual level socioeconomic characteristics are not available. Many well-established small area-based deprivation indices have been developed using data from the census, which is usually conducted decennially. While the decennial nature of the census is one of the main limitations of using it to develop such an index, the census considers the entire enumerated population, and its core set of questions generally remain stable over time.

The context of the research presented in this thesis is the Maltese Islands, referred to as Malta, an archipelago located in the middle of the Mediterranean Sea, sandwiched between Europe and North Africa. The two inhabited islands in the archipelago are divided into 68 official localities, 14 in Gozo and the rest in Malta. Despite its small size, Malta has a long history, evidenced by the presence of prehistoric temples. Its location has also led it to be seen as a valuable strategic asset and was, among others, a colony of the British up until 1964. Malta joined the European Union (EU) in 2004 which led to the opening of its borders and considerably less of the insularity normally associated with small island states. The population of Malta has grown substantially since 2013, following a change in government and in turn a change in economic policy that encouraged immigration of foreigners to fill gaps in the labour market (Grech, 2016a). As a result, the share of the foreign population increased rapidly over a ten-year period, from 5.6% at the end of 2013, to 25.3% by the end of 2022. At the end of 2022, the total population of Malta stood at 542,051, while the population density for a country area of approximately 315km^2 was 1,720 inhabitants per km^2 (National Statistics Office, 2023j). Malta is the smallest EU Member State in terms of land area but also has the highest population density. The National Statistics Office (NSO) established six geographical districts for the spatial analysis of the 1967 census. By and large, these districts have been retained as they were conceived in 1967 and have been adopted as the regular geographical breakdown used for regional spatial analysis of most of the official statistics disseminated by the NSO (National Statistics Office, 2017). Since these were conceived over fifty years ago, they are not expected to consider the major demographic and social shifts that have occurred since then.

While thirty years ago, public health researchers viewed Malta as relatively socially homogeneous, lacking the regional disparities in health outcomes seen in larger countries (Agius, 1990), this view has changed and there has been an increased recognition of the need for health inequalities research. Notwithstanding this, most recent research has been restricted to cross-sectional surveys. While Malta has several well-established country level health registers and administrative sources, these tend to lack data on the more traditional socioeconomic indicators such as education, income, and employment status. Spatial analysis of such data has attempted to serve as a proxy for more traditional socioeconomic analysis using individual predictors, however this has generally been conducted using the six districts established by the NSO, leading to limited utility of the results because of the large size and heterogeneity of these districts.

Since 2005, the NSO has implemented an annual cross-sectional survey within private households focused on income, poverty, social exclusion and living conditions, the EU-Social Income Living Condition (SILC). Indicators on material and social deprivation and risk of poverty or the at-risk-of-poverty rate (ARP) are estimated annually. While these indicators have been produced for nearly 20 years and are comparable throughout the EU, they are not without limitations. The items collected in the survey focus primarily on aspects of monetary deprivation. The indicators calculated from these items are dichotomous, categorised based on set thresholds. Such dichotomous measures of deprivation do not allow for the analysis of deprivation gradients. Indicators estimated from this survey are ultimately based on a sample, and marginalised or hidden population groups, who may be more deprived, may be excluded from the sampling frame all together. Non-response may also be higher in persons who are more deprived (Goodman and Gatward, 2008). Crucially, estimates are limited by small counts, which has restricted spatial analysis of the main indicators to the district level.

Within the context of a small country, where resources are limited and there is a push to reduce data collection burden from surveys, the routine and comprehensive study of health inequalities needs to look at alternatives to bridge information gaps. A small area-based deprivation index could be a valuable tool to bridge this gap. Such an index provides the flexibility for researchers to use available health data, which is routinely updated, provided some information on location is stored within it. The availability of the census allows for the development of a small area-based index that considers the entire population and can be updated with each new round in the census cycle. The census is a well-established exercise, which is legally mandated and conducted every ten years and captures information on the entire resident population through full enumeration. This thesis therefore aimed to develop and validate a small area-based deprivation index for Malta, using census data, and test the application of this index to the analysis of health inequalities using both administrative and self-reported health outcome data. This work will be summarised in Sections 7.2, 7.3 and 7.4. This chapter will also discuss the strengths and limitations of this research (Section 7.5), as well as the implications for policy and practice (Section 7.6) and implications for research (Section 7.7).

7.2 Development and validation of a small area-based deprivation index for Malta

Chapter 3 presented the analysis conducted to develop and validate a small area-based deprivation index for Malta. While it has been acknowledged in other parts of this thesis that several small area-based deprivation indices have been developed using other sources apart from the census, most notably administrative registers, the census was considered the best source for this research given the current issues with quality and coverage of administrative registers in Malta. While data was available to the researcher at the enumeration area (EA) level, and an EA level index was developed, the final index was intended to be produced for the 68 localities. The choice of locality as a small area for the final index, was a practical one. Enumeration areas are developed for operational use in the census and are updated each census round. Crucially, they are only used in the census and are not easily mapped to other external data sources. Since the aim of this thesis was to develop an index that could be used by researchers for inequalities analysis beyond the work presented here, locality was deemed the most appropriate option to maximise future usability of the index, as information on locality is generally available in health registers and administrative sources.

As discussed in this thesis, the concept of deprivation is not new, and researchers have been attempting to conceptualise it for many years. The general consensus, is that being deprived is relative, in that, to classify someone as deprived it must be relative to the standard level of resources, opportunities and social conditions available to most members within the society at the time in which deprivation is being measured (Townsend, 1987). Considering the relative nature of deprivation, an index developed for Malta could not simply be the replication of another index developed elsewhere but needed to be based on some form of empirical assessment of what factors within the Maltese context make a person deprived.

A priori, however, there is some consensus in the literature that several broad domains encompass the factors at the individual and household level that measure deprivation, mainly education, employment, income, living conditions, household characteristics, migration, and other demographic factors. Considering these domains, previous items used by other researchers, as well as national level indicators, nineteen (19) initial items were operationalised for possible inclusion in the index. It must be acknowledged that the items and domains which could be included were restricted by what was collected in the census questionnaire. For example, the census in Malta does not directly collect information on individual or household income, therefore items specifically related to income could not be included.

Given the relative nature of deprivation, it can be expected that with time, the relevance of certain items to the measurement of deprivation may change. To try and enhance, as much as possible, the stability of the index with time, especially for future updating, data from two census rounds were analysed in parallel when developing the index – 2005 and 2011. While the 2021 census was still ongoing at the time of the analysis and therefore data arising from this census could not be used to develop the index, the 2021 census questionnaire had been published and it was reviewed, and ultimately certain decisions related to the final selection of items to include in the index were also based on what was potentially available from the 2021 census, once data would be available to researchers.

While some researchers have selected items for their index based on desk research or through the pooling of opinions on the relevance of items from external individuals, deemed as experts, the most common method used to select and combine items for a deprivation index has been factor reduction, mainly principal component analysis (PCA) (Allik *et al.*, 2020). PCA was chosen as the method used to select and combine the items for the small area-based deprivation index developed in this thesis. This choice was not only made because this method has been applied most often, but also because a data-driven method was deemed the most appropriate technique to use to attempt to develop an index that was reasonably objective. Also, the intention was to develop an index that was concise, containing a limited number of items that were deemed, when combined, to measure deprivation most effectively. The choice of the principal component emerging from the PCA and thus exclusion of items that did not load on this component, was expected to produce the most parsimonious index.

Before computing the final index, sensitivity analysis was conducted to assess the impact of selected changes in the analytical process on the index output. This specifically focused on the impact of excluding one item that initially loaded on the principal component and the exclusion of areas with small population sizes due to possible extreme values in the items, since they were estimated as proportions. The final index contained five items covering the domains of employment, education and living conditions. The factor loadings emerging from the final PCAs for 2005 and 2011 separately, were used to weight the items before they were summed to produce a deprivation score for each geographical area, for each census year. Weighting was conducted to account for the fact that the impact of individual items on the composite measurement of deprivation, may vary. The index scores were grouped into quintiles of deprivation. Enumeration areas and localities for the two years were ranked by deprivation score and population, and quintiles were produced by segmenting the ranked scores based on a proportionate distribution of the population.

As far as was possible, validation of the final index was conducted, however the most common validation for such indices is predictive validity, specifically the ability of the index to present a gradient by deprivation in the health outcome being analysed (Allik *et al.*, 2020) – this will be discussed in sections 7.3 and 7.4. Apart from this analysis, additional work was conducted to test the performance of the index at the enumeration area and locality levels.

While the final index was developed at the locality level, since enumeration areas cannot be linked to other sources outside the census, it was acknowledged that an index developed at the enumeration area would be expected to perform better than that at the locality level due to the smaller sizes of enumeration areas. While this analysis showed that the enumeration area index did present generally steeper gradients by deprivation for mortality five years from the census, the patterns exhibited using the locality index were comparable.

7.3 Application to all-cause mortality and cancer incidence

Chapter 4 and Chapter 5 presented the application of the final locality level deprivation indices for 2005 and 2011 to the analysis of all-cause mortality and cancer incidence. In both analyses, age-standardised rates using the European Standard Population (ESP) were selected as outcomes for analysis. Age-standardised rates are a widely used outcome measure in public health because they are calculated using readily accessible health outcome data and population counts; and are easy to understand and interpret. Furthermore, age-standardisation allows for adjustment based on the age structure of the population, which produces estimates that are comparable across different categories, such as geographical areas and time periods. For the mortality analysis, estimates were produced for the total, as well as for premature mortality (0-69) and old-age mortality (70+). The cancer incidence rates were estimated for the top six cancer sites in terms of cancer incidence. Due to small counts, both analyses were conducted on data grouped over five-year periods. Analysis was conducted using the 2005 and 2011 indices.

To quantify the relative socioeconomic gradient, if any, for all-cause mortality and cancer incidence, a summary measure of the linear association between deprivation and the age-standardised rates was estimated. The Relative Index of Inequality (RII) was used to measure the inequality gradient. While premature mortality rates declined between the two periods, the linear gradient for premature mortality by deprivation quintile, increased. This suggests that the declines in mortality risk were not equal across deprivation quintiles. This was most notable for males in the most deprived areas, where the premature mortality rate in the most recent period increased. Similar patterns in premature mortality have been noted by other researchers, and this has been attributed to increases in mortality related to lifestyle factors such as alcohol and drug use, and mental health related causes in more deprived areas (Leyland *et al.*, 2007; Norman *et al.*, 2011). Results for old-age mortality were less consistent.

Old-age mortality rates estimated for the third quintile were outside of the linear trend. While exclusion of mortality in the very-old age group (85+) attenuated the results when using the 2011 index, estimates for males in quintile 3 remained out of the trend line when using the 2005 index. Additional analysis on individual items showed that it is possible that some of the items included in the composite index do not predict old-age mortality as well as premature mortality when using the 2005 index.

The relationship between deprivation and cancer incidence was not consistent across cancer sites and periods. For female breast cancer and colon-rectum cancer, the RII estimates were negative in the two periods, notwithstanding the wide confidence intervals, this suggests that the risk of incidence is higher in the lowest deprivation quintile when compared to the most deprived quintile. The results for bronchus-lung cancer incidence present the strongest deprivation gradient in this analysis, especially for the most recent period. This deprivation gradient in bronchus-lung cancer incidence has been documented by other researchers. Inequalities in smoking patterns are said to be the major contributory factor for the higher risk associated with the more deprived areas (Kuznetsov *et al.*, 2011; Riaz *et al.*, 2011; Li *et al.*, 2015; Mihor *et al.*, 2020). Occupational exposures may also contribute to the increased risk (Menvielle *et al.*, 2010).

7.4 Combining area deprivation and individual predictors of health

The final analysis conducted in this thesis is presented in Chapter 6 and sought to use multi-level modelling to analyse whether the area-based measure of deprivation still contributed to the analysis of health inequalities when individual level characteristics were included. Three binary health outcomes were selected for the analysis – mortality within five years of enumeration in the census, and two self-reported health outcomes taken from the census data itself, prevalence of a chronic health condition and prevalence of a mental health condition. Since census data were used to conduct this analysis, the enumeration index could be used. Analysis was restricted to the 2011 data. This analysis highlighted that the small area-based deprivation index still contributed to the analysis of health inequalities even when adjusting for individual level socioeconomic predictors. For all three health outcomes, a gradient was still seen with increasing deprivation, though the odds were attenuated when including individual predictors. These results are in line with those presented in similar analyses (Skapinakis *et al.*, 2005; Meijer *et al.*, 2012; Fone *et al.*, 2014). The strongest relationships were maintained for the most deprived quintile, for all three outcomes, suggesting that something specific to these areas had an impact on health beyond the impact of the individual predictors. While it was beyond the scope of this analysis, there are several possible reasons why neighbourhoods may impact upon individual health outcomes. People living in the same area may affect each other's health behaviours through shared norms and values. This may also extend to the physical environment, where the inhabitants of an area may impact upon the presence of services, businesses and activities in the area that are health promoting. Areas that are more deprived may have lower social cohesion, trust, and social capital (Visser *et al.*, 2021). There may also be differences between rural and urban areas which all have implications on population density, exposure to pollution, access to open and green spaces, and safety.

The use of individual socioeconomic predictors in the analysis presented in Chapter 6 allowed for an assessment of individual level social determinants of both self-reported and register based health outcomes in Malta using the census. For all three health outcomes, when adjusting for all other covariates, employment status and literacy consistently presented a risk gradient, with higher odds seen in the least advantageous category compared to the most advantageous category. The results for the two other individual level socioeconomic predictors, occupation skill level and highest level of education achieved, were less consistent across the three health outcomes. For all three health outcomes, as would be expected, increased odds were seen with increasing age when adjusting for all area and individual level predictors. Increased odds were also seen across all health outcomes when comparing males to females. With respect to citizenship, there were no differences between those with Maltese citizenship and those without Maltese citizenship for mortality risk five years after the census. For the two self-reported health outcomes, being foreign seemed to be a protective factor.

7.5 Strengths and limitations

To the author's knowledge, this is the first attempt to develop a small area-based deprivation index for Malta using census data. While such indices are not new and the techniques applied in the development of this index have been used elsewhere, the research does have several strengths. The index described in this thesis was developed using two census datasets, something that, to the author's knowledge has not been done previously. Furthermore, since the questionnaire for the 2021 census had already been published at the time of analysis, it was also taken into consideration when finalising the items to be included in the index. This approach aimed to enhance the consistency of the index across time to maximise comparability, and allow for future updating with the 2021 data, without having to exclude any items because they were no longer available.

Sensitivity analysis was done to assess the impact of selected changes in the analysis process on the index and validation of the final index was conducted, as far as was possible. Explicit validation of small area-based deprivation indices is not commonly conducted by researchers when the index is created, this may be because not many methods have been outlined specifically to conduct such validation (Pampalon *et al.*, 2014; Allik *et al.*, 2020). While testing the deprivation index against health outcomes, its predictive validity, might be one of the best and most common approaches to validate an index, an attempt was made to apply other criteria for validation, including internal consistency, content validity and convergence validity with other external measures of deprivation.

To test for a health gradient and in turn the predictive validity of the index created in this thesis, two administrative health outcomes were analysed – all cause mortality and cancer incidence. The testing of two rather than one health outcome strengthened the validity of the index as, despite some possible limitations due to small counts, results using both indices were consistent. Also, to the author's knowledge, this was the first attempt to analyse social determinants of health related to mortality and cancer incidence in Malta beyond previous analysis using district as a proxy for socioeconomic status. Finally, multi-level modelling was used to analyse the added value, if any, of the area-based deprivation index when using it in combination with individual level socioeconomic predictors. Using the 2011 census data, and three separate health outcomes, the analysis showed that even when adjusting for individual predictors, area-based deprivation still presented a health gradient. Since census data was used, analysis could be conducted using the enumeration area index. To the author's knowledge, this was the first attempt to use census data for detailed health inequalities analysis in Malta.

Despite the strengths outlined above, the research presented in this thesis does have a few limitations. As acknowledged previously, using area-based deprivation indices in health inequalities analysis may be prone to ecological fallacy, since an area level measure of deprivation is being used as a proxy for individual predictors. Notwithstanding this limitation, this research presented several use cases which provided consistent results showing that the index is useful for health inequalities analysis. Beyond that, even when including the index with individual level predictors, the area index was still a valuable predictor. The use of a composite index of deprivation rather than a single item measure also allows for a meaningful coverage of the whole population in a way that single items may not be able to. For example, items on employment are generally restricted to the working age population while measures of education may mean different things for different cohorts of the population because of changes in the education system over time. Finally, despite the limitation of ecological fallacy, the area-based index developed here allows for extensive applicability to health inequalities analysis, as the only information required in health data sources to use the indices is locality, something that is generally available as part of the basic information collected in health registers and administrative sources.

The use of census data for the development of a deprivation index is limited by the fact that the census is only updated every ten years. While it is acknowledged that the census becomes outdated the further away one moves from the census year, it is currently the best source available for the development of such an index in Malta. While the smallest area available from the census was the enumeration area, since this cannot be linked to other external sources, the final index had to be developed at the locality level. While comparative analysis between the enumeration area and locality level indices showed that the enumeration area index predicted slightly stronger gradients using the relative index of inequality (RII), the results using the locality index were still comparable.

Population estimates by locality produced by the NSO are not presently disseminated disaggregated by age, with only total counts by sex available at the locality level. The distribution by age of the mid-year population by locality was required as denominator data for the analysis presented in Chapter 4 and Chapter 5, therefore this had to be approximated using the age and sex breakdowns available from the censuses. This method assumed that the age distribution by locality for the years around the censuses had remained fixed and this may not have been applicable for all years or all localities. However, since the distributions used for analysis were 5-year age groups, the differences over time were expected to have had less impact than would have been the case for single years of age.

Despite grouping outcome data into five-year periods to be able to analyse larger counts, some of the analysis was still based on relatively small numbers per quintile which led to large margins of error around the results. This was especially for the cancer incidence data. This made it difficult to draw conclusions on the association between deprivation and cancer incidence. This is an issue unique to small countries that must be acknowledged when interpreting findings.

Finally, some of the results emerging from the analysis presented in this thesis were inconsistent with what was expected. This was specifically for the performance of indices in patterning old age mortality by deprivation presented in Chapter 4, and as well as the behaviour of the education variable as an individual level predictor in the analysis presented in Chapter 6. Estimates for all cause old age mortality (70+) peaked in quintile 3, rather than in the most deprived category, meaning this quintile behaved outside of the deprivation gradient. For education, when included in the multi-level model, low level education as an individual predictor appeared to be a protective factor for prevalence of self-reported chronic health conditions. Additional analysis was conducted, as far as was possible, to attempt to explain these findings.

7.6 Implications for policy and practice

As discussed previously, more prominence has been given to social determinants of health by researchers and policy makers in Malta. This is clearly evidenced by the content of recent published policies and strategies issued by the Ministry for Health which contain, in some cases, specific sections or actions related to addressing socioeconomic inequalities. The national strategy on obesity published in 2012 highlighted the possibility of a higher prevalence of overweight and obesity in lower socioeconomic groups (Superintendence of Public Health Ministry for Health, 2012). The national cancer plan also acknowledges that there are variations in survival outcomes between different social strata (Ministry for Health Malta, 2017). The two most recent strategies take it further by putting social determinants as the main area to address to improve health outcomes. The mental health strategy presents social determinants as the major factors leading to mental health disorders over the life course (Ministry for Health Malta, 2019), while the national health system strategy recognises the widening health gap in Malta and that reducing this gap is the governments priority (Ministry for Health Malta, 2022).

While there has been clear recognition of the need to address social determinants in health planning and service provision within the health system, there is very limited local research available on which to develop them. In fact, when reviewing the evidence on which the above policies and strategies were developed, most of the quoted data and studies were international research. Local evidence was restricted to patterning of self-reported health outcomes by education or income from a national cross-sectional health survey conducted on a sample of the population every five years, or spatial patterning of health outcomes, such as mortality, using the same six districts discussed previously. This made it very difficult for those who developed the strategies to outline highly targeted interventions to address the social determinants specifically impacting Malta. For example, in the national obesity strategy, an action was outlined to set up community initiatives specifically targeting lower socioeconomic groups, however it was not clear what was meant by "lower socio-economic groups" and how they would be targeted (Superintendence of Public Health Ministry for Health, 2012). Similarly, the national cancer plan identified the need to implement targeted actions specifically to reduce socioeconomic and regional inequalities in access to and uptake of screening services, however it was not made clear who would be targeted and why (Ministry for Health Malta, 2017).

The national health system strategy highlighted the priority given to health inequalities by the government through the creation of a unit specifically focused on addressing social determinants of health, as well as the implementation of a national survey in 2019, by this unit, specifically on social determinants. However, it is not clear, since its inception, what the specific work plan is for this unit and even though the survey was conducted in 2019, until now, results have not been published. The strategy presented actions to be implemented to address health inequalities such as targeted outreach to groups with low uptake of preventive services, focussing health services to ensure that persons who are at risk of falling behind are followed up, and training clinicians on how to identify and manage health inequalities in their practice. As in the other strategies, however, it was not clear who these groups are, how they were identified and ultimately targeted (Ministry for Health Malta, 2022).

The potential implications of the small area-based deprivation index developed in this research for future evidenced-based policy and strategy making are substantial. The index allows for more extensive and continuous patterning of health inequalities for a variety of outcomes, which will be discussed in further detail in the next section. While this will allow for policy making to be based on data related specifically to the local scenario, the fact that the index identifies areas that are deemed as deprived, creates the potential for actions that are currently generic in nature to be more targeted to areas. Such targeted action and strategies are be expected to lead to more cost-effectiveness, in the future, where resources are allocated to the areas where they are expected to have the most impact in reducing the equality gaps. A robust health inequality monitoring system would also ensure that the implementation and results of these targeted actions can be monitored.

7.7 Implications for future research

The strategies and policies discussed above have a common thread running through them – the recognition that there is an urgent need for more research in health inequalities in Malta. In fact, all the policies and strategies referenced here include an action specifically related to increasing research in health inequalities to fill in the large gaps in national evidence. Unfortunately, there seems to be an inherent reliance on conducting this research through traditional surveys, which are cost and resource intensive and seem to take a very long time to be disseminated. This may be because there are no perceived current alternatives available. The potential applications for the small area-based deprivation index developed in this research with respect to patterning of health inequalities in Malta, are widespread. This is especially true with respect to the potential analysis of administrative and health registers which generally all have data related to locality of residence. This opens the possibility of analysing such outcomes as uptake of preventive and national screening services, clinical performance indicators in hospitals such as 30-day case fatality rates and readmission rates, cause specific mortality and survival rates and uptake of specific health services.

As was presented in this thesis, the index is not only valuable as a tool when no individual based socioeconomic data is available, but it is also valuable as an additional predictor along with individual level data. Future research may also delve deeper into the specific characteristics of the most deprived areas that may be contributing to continued, and in some cases, increasing poorer health outcomes, even when controlling for individual factors. While the index was developed to bridge the evidence gap specifically in health inequalities, the index has potential applications in any domain of research interested in studying patterning of outcomes related to deprivation. The index could also be of use in research planning including survey sampling. Deprivation can potentially be used as a criterion in sampling schemes to ensure representative coverage of all deprivation quintiles in survey research.

Finally, the index developed through this research can be updated with the 2021 census data once this becomes available to researchers. The 2021 census data will provide an update which is expected to align with the considerable demographic change experienced in Malta over the last ten years. The 2021 census also brings with it potential advances in spatial analysis of data not currently available in previous censuses. While enumeration areas were developed in the same way as previous rounds of the census, and data disseminated so far has been outputted using the traditional spatial breakdown of locality or districts (National Statistics Office, 2023b), the 2021 census data has been geocoded. This novel aspect of this round of the census is a result of an EU wide regulation which mandates all EU Member States to geocode census data to a $1km^2$ grid as from 2021. While currently geocoding of other data sources apart from the census, is relatively non-existent, if geocoding of data becomes more widespread, nationally, there is the potential for future development of an index to a $1km^2$ grid allowing for dissemination at a smaller area. As was shown when comparing analysis using the enumeration area and locality indices, using smaller areas improves the ability of the deprivation index to outline gradients in health outcomes by deprivation.

7.8 Conclusion

The contribution of this thesis to health inequalities research and evidence-based policy and strategy making in Malta, is clear. While area-based deprivation indices are not new and the methods applied here are not innovative, this thesis still contributes meaningfully to health inequalities research, in general, despite its focus being on the Maltese context. Firstly, the breath of the use cases presented show that there is continued value in small area-based deprivation indices for health inequalities research, despite their limitations. The analysis presented here has shown applications using both administrative data and self-reported data, as well as the application of the index combined with individual predictors. Though traditionally these indices have been used as proxies for individual predictors when they are absent, this research has shown that they are also valuable when combined with individual predictors.

While recently, small area-based indices have been developed using other sources apart from the census, mainly administrative registers, this thesis has shown that the census is still a useful source to develop an index, despite its decennial nature. The research presented here is an example of the creation of a small area-based deprivation index in a small island state. While small countries may be viewed as not benefitting from regional or small area analysis, due to their small size, this research has shown that patterning of health outcomes to small geographic areas are important, even in such cases. Indeed, the availability of sub-national and smaller geographical estimates of health indicators across countries varies considerably, which limits the ability to measure comparative population health across small areas (Murray *et al.*, 2022). The research presented here adds to the current evidence and improves the ability for Malta to compare itself internationally with respect to small area health inequalities analysis.

Appendix A

Summary of domains and items included in the census small area-based deprivation indices.

Index	Domains	Items included in index
Jarman underprivileged area index (UPA)	Household type	Proportion of elderly living alone Proportion of persons in one-parent families
	Employment	Proportion of unskilled (SEI l) Proportion of unemployed
	Demography	Proportion of children under five years of age
	Migration	Proportion of those who have moved house during the previous year Proportion of persons of minority ethnic origin (New Commonwealth and Pakistan).
	Living conditions	Proportion of persons living in overcrowded households
Townsend Index	Employment	Proportion of those aged 16 and over who are unemployed
	Living conditions	Proportion of households not owner occupied Proportion of households that are over-crowded. Proportion of households not owning a car
Carstairs score	Living conditions	Overcrowding Car ownership
	Employment	Low social class Male unemployment
Index of Relative Social Deprivation (IRSD)	Household type	Proportion of one parent families with dependent offspring only Proportion of people aged 15 and over who are separated or divorced
	Income	Proportion people with stated annual household equivalised income at approx. 1st and 2nd deciles
	Employment	Proportion of families with children under 15 years of age who live with jobless parents Proportion of employed people classified as 'labourers' Proportion of people (in the labour force) unemployed Proportion of employed people classified as Machinery Operators and Drivers Proportion of employed people classified as Low Skill Community and Personal Service Workers
	Living conditions	Proportion of occupied private dwellings with no internet connection Proportion of occupied private dwellings paying rent less than $166 per week (excluding $0 per week) Proportion of occupied private dwellings with no cars Proportion of occupied private dwellings requiring one or more extra bedrooms (based on Canadian National Occupancy Standard)
	Education	Proportion of people aged 15 years and over whose highest level of education is Year 11 or lower Proportion of people aged 15 years and over who have no educational attainment

Index	**Domains**	**Items included in index**
		Proportion of people who do not speak English well
	Health	Proportion of people aged under 70 who have a long–term health condition or disability and need assistance with core activities
Socioeconomic Factor Index (SEFI)	Living conditions	Mean value of dwelling
	Household type	Proportion of female parent households with children
	Employment	Proportion of female labour force participation aged 15+ Proportion unemployed age 45 – 54 Proportion unemployed aged 15-24
New Zealand Index of Deprivation (NZDep)	Household type	Proportion of persons aged <65 living in a single parent family Proportion of persons separated or divorced and aged 60 or over Proportion of persons separated or divorced and aged 60 or over
	Income	Proportion on means tested benefit and aged 18-59 Proportion of persons below threshold for equivalised household income
	Living conditions	Proportion of persons without access to a car and aged 18 or over Proportion of persons not in owned house Proportion of persons below threshold for equivalised occupancy
	Education	Proportion without qualifications and aged 18-59
	Employment	Proportion of persons unemployed and aged 18-59
Singh Area Deprivation Index	Household type	Proportion of single-parent households with children aged < 18 years
	Education	Proportion aged ≥25 y with <9 y of education Proportion aged ≥25 y with at least a high school diploma
	Employment	Proportion employed persons aged ≥16 y in white-collar occupations Proportion civilian labor force population aged ≥16 y unemployed
	Income	Median family income, $ Income disparity (ratio of low income to high income households) Proportion families below poverty level
	Living conditions	Median home value, $ Median gross rent, $ Median monthly mortgage, $ Proportion owner-occupied housing units Proportion of households with more than 1 person per room, Proportion of households without a motor vehicle Proportion of households without a telephone Proportion of occupied housing units without complete plumbing
Tello et al	Education	Individuals with elementary school level Individuals with university qualification
	Employment	People employed in the industry sector Civil servants or people employed in the tertiary sector Unemployment rate
	Household type	Individuals married Individuals separated or divorced or widowed Single-parent families
	Living conditions	Rented accommodation
Messer et al	Education	Proportion earning less than a high school education
	Employment	Proportion of males in management and professional occupations Proportion unemployed

Index	Domains	Items included in index
	Income	Proportion of households in poverty Proportion of households on public assistance Percent of households earning $30,000 per year estimating poverty
	Living conditions	Proportion of households that are overcrowded
	Household type	Proportion of female headed households with dependents
Stimpson et al	Education	Proportion low education (<9 years) Proportion high education (college educated)
	Employment	Proportion in managerial/professional occupations of all with reported occupations Percent of persons in labour force who are unemployed
	Income	Median family income Ratio of income for highest and lowest quintiles Percent of persons for whom poverty status is determined who live in households with income ≤100Proportion of the federal poverty level by census definition
	Living conditions	Median house value Median gross rent Percent of households with no telephone Proportion of households with no plumbing
Winkleby et al	Education	Proportion of residents with low educational status (<10 years of formal education)
	Employment	Proportion residents unemployed (excluding full- time students, those completing compulsory military service, and early retirees)
	Income	Proportion of residents with low income (defined as less than 50Proportion of individual median income) Proportion residents who are social welfare recipients
Vancouver Area Neighbourhood Deprivation Index (VANDIX)	Education	Proportion of residents without high school completion Proportion of residents with a university degree
	Employment	Ratio of those 15 years working or seeking work to the total population Unemployment rate of population aged 15 years and over
	Living conditions	Proportion of persons owning their home
	Household type	Proportion of lone parent families among all census families
	Income	Average 2000 income
Havard et al	Employment	Blue-collar workers in the labour force People in the labour force with insecure jobs People in the labour force with stable jobs Unemployed people in the labour force People in the labour force unemployed for more than 1 year
	Living conditions	Primary residences that are houses or farms Primary residences that are multiple dwelling units Households without a car Households with two or more cars Mean number of people per room Primary residences with more than one person per room Non-owner-occupied primary residences
	Income	Subsidised housing among all primary residences Median income per consumption unit (in euros per year)

Index	Domains	Items included in index
	Household type	Single-parent families
	Education	People aged 15 years or older with general or vocational maturity certificates People aged 15 years or older with at least a lower tertiary education People aged 15 years or older who did not go beyond an elementary education
	Migration	Foreigners in the total population
Sanchez-Cantalejo et al	Employment	Proportion of unemployed Proportion of manual labourers as a share of total workers
	Education	Proportion of illiterate persons in the population aged 10 and over
Lian et al	Employment	Proportion of unemployed males aged 20 years or more Proportion of unemployed females aged 20 years or more
	Living conditions	Proportion of households with no car
	Income	Proportion of households on public assistance Proportion of households with low income (1999) Proportion of population below federal poverty line
	Household type	Proportion of female-headed households with dependent children
	Migration	Proportion of non-Hispanic black
	Education	Proportion of population 25 to 64 years old with high school education or lower
Choi et al	Employment	Proportion of population men 15 - 64 who are jobless and actively seeking jobs Proportion of reference persons aged 15-64 who have elementary occupations
	Living conditions	Proportion of households not owner occupied Proportion of households without access to a car Proportion of households that are not an apartment Proportion of households living below the minimum standard
	Demography	Proportion of individuals who are older than 65 years
	Household type	Proportion of single-person households Proportion of population aged 15 and over who are divorced/bereaved Proportion of households with a female reference person
	Education	Proportion of population aged ≥25 years with <12 years education Proportion population aged ≥5 years that speaks Spanish and speaks English very well
Torres-Cintrón et al	Employment	Proportion of employed civilian population aged ≥16 years in management, professional, and related occupations Proportion of civilian labour force aged ≥16 years unemployed
	Living conditions	Proportion of occupied housing units without telephone Proportion of occupied housing units with no vehicle available
	Income	Proportion of population with 1999 income below poverty level
	Education	Proportion of households headed by a person with primary education or less
Panczak et al	Employment	Proportion of households headed by a person in manual or unskilled occupations
	Living conditions	Mean number of persons per room Median rent in Swiss Francs per square metre

Index	Domains	Items included in index
	Education	Proportion of adults over 25 years with less than a high school education
Rossen	Employment	Proportion of men over 16 years who are unemployed
	Household type	Proportion of female-headed households with children
	Income	Proportion of households receiving public assistance Median household income Proportion of families below the Federal Poverty Threshold (FPT)
	Education	Low educational status (<10 years of formal education)
Bender et al	Employment	Proportion of persons not in work (e.g. students, unemployed)
	Income	Proportion of persons belonging to the lowest income quartile
	Education	Proportion of people with degree below elementary school Proportion of people receiving no fundamental education Illiteracy rate
Multi-criteria deprivation index for the city of Quito (MDIQ)	Employment	Proportion of the population that works in unpaid jobs
	Living conditions	Proportion of households with four or more persons per dormitory Proportion of households without access to drinking water from the public system Proportion of households without access to the sewage system Proportion of households without access to the public electricity grid Proportion of households without garbage collection service
	Health	Proportion of the population that have a long-term disability (for more than one year) Proportion of the population that has no public social/health insurance Distance (metres) to the nearest primary healthcare service
	Education	Average years of education for people over 6 years of age Illiteracy rate among people over 15 years of age Proportion of people over 6 years of age not completing junior high school
Weng et al	Employment	Proportion of blue-collar workers
	Living conditions	Proportion of households without fixed housing Proportion of households without tapping water pipe Proportion of households without bathing facilities Proportion of households without kitchens or toilets
	Household type	Proportion of adult female living alone Proportion of old people living alone Proportion of non-registered population
	Income	Proportion of low-income households
	Education	Proportion of persons 25 years and older without a high school diploma
Palmetto Small-Area Deprivation Index (SADI)	Living conditions	Proportion of housing units with no vehicle available
	Income	Proportion of noninstitutionalized population below the federal poverty level
	Education	Proportion of persons aged over 15 years with less than primary or primary education
	Household type	Proportion of couples who are married with 3 or more children

Index	Domains	Items included in index
Aungkulanon et al	Employment	Proportion of persons aged over 15 years who are unemployed Ratio of the population aged under 15 or over 60 to the population aged between 15 and 60
	Living conditions	Proportion of households with more than 1 person per bedroom Proportion of households without access to the Internet Proportion of households without access to phone Proportion of households with no television Proportion of households with no refrigerator Proportion of households without car/motorbike Proportion of households not owning their dwelling
	Household type	Proportion of couples who are married with 3 or more children Proportion of elderly persons (aged 60 and over) living alone Proportion of females aged 15 and over who are separated, divorced, or widowed
	Rural-urban location	Proportion of households in non-municipal areas
	Health	Proportion of persons with visible disabilities
	Migration	Proportion of persons with a different address 1 year before the census Proportion of non-Thai citizens
	Education	Proportion of the population that does not have any level of formal education or instruction
Powell-Wiley et al	Employment	Proportion working adults not in an executive, managerial, or professional occupation
	Household type	Proportion single mothers with children
	Income	Log-transformed median household income Proportion receiving welfare Proportion below the poverty level Proportion households not receiving dividends, interest, or rental income
	Education	Proportion adults >=25 years old without a high school diploma Proportion adults >=25 years old without a Bachelor's degree
	Living conditions	Proportion of households without a telephone Proportion non-owner occupied units Log-transformed median home value
Wang et al	Employment	Proportion of people over 16 years of age working in low-income industries Proportion of people over 16 years of age losing working ability
	Living conditions	Proportion of households without indoor facilities of water, sanitary toilet, kitchen, or shower
	Rural-urban location	Proportion of people with rural hukou (chosen to reflect discrimination against rural people).

Appendix B

Spearman correlations for EA item Z-scores: 2005

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1.CROWDED	1	.085**	.469**	.252**	.472**	.435**	.504**	.564**	.577**	-.207**	-.152**	.368**	.353**	.229**	.189**	.107**	.481**	-.049	.382**
2.ELDERLY_ALONE		1	.101**	.021	.098**	.098**	.132**	.131**	.100**	.286**	.262**	.273**	.392**	.223**	.130**	.035	.338**	.420**	.311**
3.UNEMPLOYED_M			1	.361**	.897**	.457**	.491**	.557**	.538**	-.160**	-.135**	.364**	.414**	.219**	.165**	.076*	.434**	.068*	.320**
4.UNEMPLOYED_F				1	.700**	.240**	.224**	.276**	.311**	-.068*	-.034	.144**	.195**	.069*	.041	.053	.211**	.009	.126**
5.UNEMPLOYED_T					1	.451**	.472**	.548**	.547**	-.139**	-.110**	.335**	.410**	.189**	.139**	.078*	.429**	.060	.299**
6.ELEM_OCC						1	.658**	.747**	.569**	-.323**	-.303**	.382**	.383**	.242**	.163**	-.111**	.292**	-.025	.309**
7.ILLITERATE							1	.850**	.639**	-.354**	-.283**	.577**	.484**	.329**	.205**	-.117**	.401**	-.006	.498**
8.NO_QUAL								1	.726**	-.388**	-.355**	.558**	.513**	.343**	.206**	-.157**	.404**	.003	.462**
9.ESL									1	-.290**	-.172**	.411**	.367**	.279**	.190**	.074*	.434**	-.036	.330**
10.MT_CITIZEN										1	.538**	-.115**	-.003	-.059	-.040	.208**	.115**	.529**	-.016
11.HOUSE_STABILITY											1	-.111**	-.006	-.062*	.019	.291**	.071*	.442**	-.048
12.BASIC_SERVICE												1	.546**	.450**	.279**	-.070*	.429**	.147**	.546**
13.APP_POV_1													1	.535**	.306**	-.044	.497**	.311**	.534**
14.APP_POV_2														1	.485**	-.030	.318**	.152**	.368**
15.APP_POV_3															1	-.009	.176**	.056	.197**
16.SINGLE_PARENTS																1	.192**	.101**	-.009
17.RENTAL																	1	.206**	.529**
18.VACANT																		1	.213**
19.STATUS_REPAIR																			1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Spearman correlations for EA item Z-scores: 2011

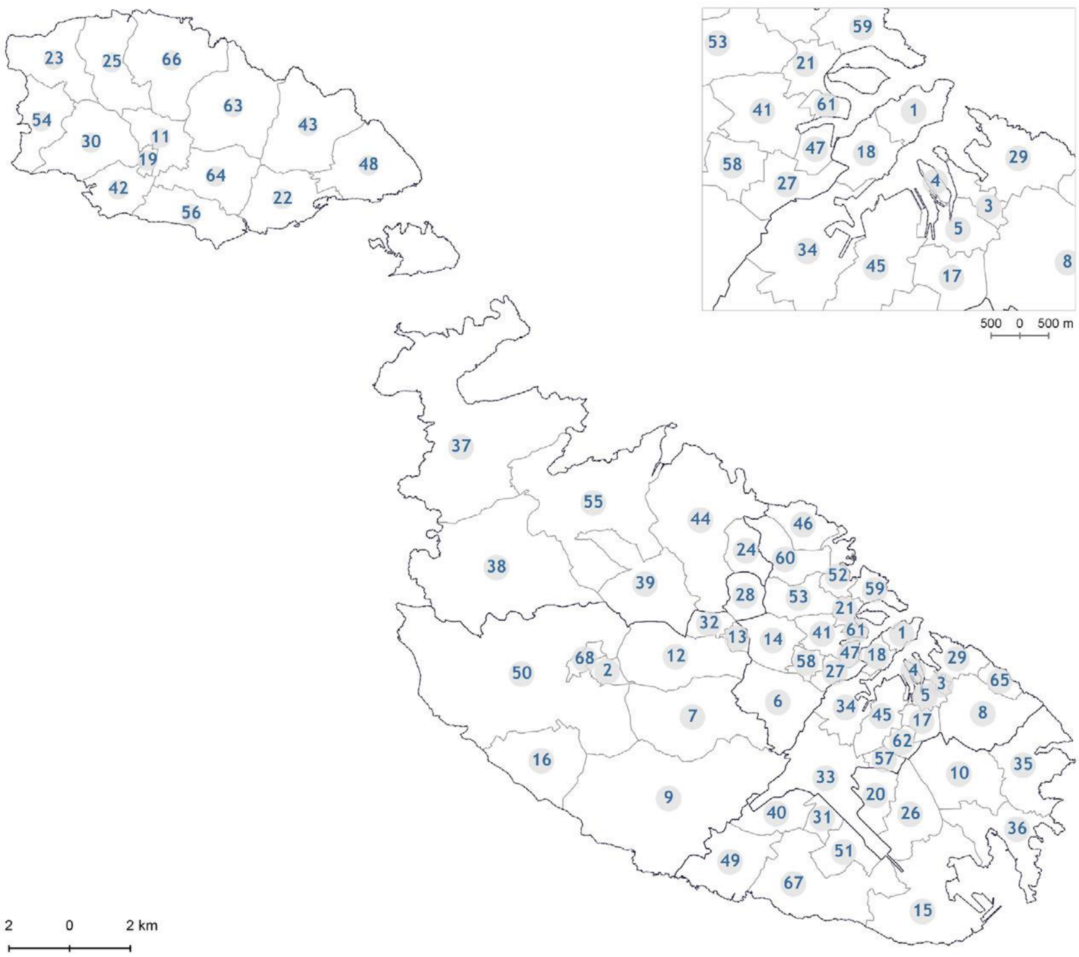
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1.CROWDED	1	.135**	.433**	.266**	.443**	.480**	.432**	.401**	.530**	-.140**	-.086**	.309**	.309**	.230**	.120**	.224**	.437**	-.021	.315**
2.ELDERLY_ALONE		1	.226**	.227**	.287**	.142**	.138**	.223**	.176**	.297**	.144**	.318**	.468**	.276**	.099**	.091**	.416**	.445**	.293**
3.UNEMPLOYED_M			1	.336**	.874**	.494**	.446**	.430**	.528**	-.057	-.088**	.362**	.428**	.281**	.146**	.129**	.472**	.093**	.283**
4.UNEMPLOYED_F				1	.713**	.238**	.171**	.146**	.307**	.142**	.106**	.152**	.251**	.165**	.066*	.275**	.342**	.212**	.174**
5.UNEMPLOYED_T					1	.471**	.405**	.381**	.526**	.039	-.009	.340**	.443**	.286**	.140**	.214**	.516**	.192**	.282**
6.ELEM_OCC						1	.728**	.735**	.650**	-.383**	-.357**	.398**	.410**	.328**	.165**	.004	.280**	-.090**	.295**
7.ILLITERATE							1	.795**	.607**	-.463**	-.419**	.503**	.467**	.398**	.233**	-.079*	.282**	-.123**	.413**
8.NO_QUAL								1	.553**	-.415**	-.464**	.467**	.521**	.404**	.232**	-.181**	.327**	-.080*	.425**
9.ESL									1	-.213**	-.197**	.371**	.416**	.307**	.173**	.173**	.444**	-.041	.314**
10.MT_CITIZEN										1	.626**	-.134**	.052	-.089**	-.074*	.227**	.291**	.601**	-.086**
11.HOUSE_STABILITY											1	-.171**	-.081**	-.140**	-.081**	.307**	.172**	.427**	-.111**
12.BASIC_SERVICE												1	.481**	.446**	.281**	-.003	.322**	.098**	.432**
13.APP_POV_1													1	.473**	.213**	.044	.508**	.298**	.451**
14.APP_POV_2														1	.397**	.039	.291**	.079*	.359**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
15.APP_POV_3															1	.037	.166**	.016	.216**
16.SINGLE_PARENTS																1	.264**	.154**	.066*
17.RENTAL																	1	.267**	.395**
18.VACANT																		1	.054
19.STATUS_REPAIR																			1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix C



Key

1 Valletta	18 Floriana	35 Marsaskala	52 St Julian's
2 Mdina	19 Fontana	36 Marsaxlokk	53 San Ġwann
3 Vittoriosa	20 Gudja	37 Mellieħa	54 San Lawrenz
4 Senglea	21 Gżira	38 Mġarr	55 St Paul's Bay
5 Cospicua	22 Għajnsielem and Comino	39 Mosta	56 Ta' Sannat
6 Ħal Qormi	23 Għarb	40 Mqabba	57 Santa Luċija
7 Ħaż-Żebbuġ	24 Ħal Għargħur	41 Msida	58 Santa Venera
8 Ħaż-Żabbar	25 Għasri	42 Munxar	59 Tas-Sliema
9 Siġġiewi	26 Ħal Għaxaq	43 Nadur	60 Swieqi
10 Żejtun	27 Ħamrun	44 Naxxar	61 Ta' Xbiex
11 Victoria	28 Iklin	45 Paola	62 Ħal Tarxien
12 Ħ'Attard	29 Kalkara	46 Pembroke	63 Xagħra
13 Ħal Balzan	30 Ta' Kerċem	47 Tal-Pieta'	64 Xewkija
14 Birkirkara	31 Ħal Kirkop	48 Qala	65 Xgħajra
15 Birżebbuġa	32 Ħal Lija	49 Qrendi	66 Żebbuġ (Gozo)
16 Ħad-Dingli	33 Ħal Luqa	50 Rabat (Malta)	67 Żurrieq
17 Fgura	34 Marsa	51 Ħal Safi	68 Mtarfa

Appendix D

Average age for the localities based on the 2005 census and the 2011 census

Locality	Total			Males			Females		
	2005	2011	Difference	2005	2011	Difference	2005	2011	Difference
Mdina	47.1	52.4	5.4	44.3	49.2	5.0	49.2	55.0	5.8
Ta' Xbiex	42.1	47.1	5.1	41.7	46.6	4.9	42.4	47.6	5.2
Iklin	33.5	37.9	4.4	33.1	37.5	4.5	34.0	38.4	4.3
San Ġwann	36.0	39.3	3.4	35.9	39.2	3.3	36.1	39.5	3.4
Gudja	37.2	40.4	3.2	36.6	39.9	3.3	37.7	40.8	3.1
Għajnsielem	37.7	40.9	3.2	37.5	40.8	3.3	38.0	41.0	3.0
Mellieħa	37.3	40.4	3.1	36.9	39.9	3.0	37.7	41.0	3.3
Attard	36.2	39.3	3.1	35.4	38.1	2.7	37.0	40.4	3.4
Munxar	36.1	39.1	3.1	35.0	38.4	3.4	37.2	39.9	2.8
Kerċem	37.9	41.0	3.1	36.5	39.9	3.5	39.4	42.0	2.6
Santa Luċija	40.5	43.5	3.0	39.4	42.7	3.4	41.7	44.3	2.6
Birgu	42.2	45.2	3.0	40.8	44.2	3.4	43.7	46.3	2.5
Mtarfa	31.6	34.5	3.0	30.1	33.4	3.3	33.0	35.6	2.6
San Pawl il-Baħar	36.8	39.7	2.9	36.5	39.4	2.9	37.2	40.0	2.8
Pembroke	30.6	33.4	2.8	30.7	33.1	2.5	30.6	33.8	3.2
Luqa	45.4	48.2	2.8	42.0	44.2	2.2	48.5	51.8	3.3
Fgura	35.8	38.6	2.7	35.5	38.2	2.7	36.2	38.9	2.7
Rabat	41.6	44.2	2.6	38.9	41.9	2.9	44.0	46.4	2.3
Naxxar	35.4	38.0	2.6	35.3	37.7	2.4	35.5	38.3	2.8
Safi	34.6	37.2	2.6	33.5	35.8	2.3	35.7	38.6	2.9
Siġġiewi	35.9	38.5	2.6	35.1	37.7	2.7	36.8	39.3	2.5
Kalkara	37.3	39.9	2.5	36.6	38.8	2.1	38.0	41.0	3.0
Swieqi	35.0	37.6	2.5	34.8	37.3	2.4	35.3	37.9	2.6
Isla	41.2	43.7	2.5	39.7	42.3	2.6	42.7	45.1	2.4
Nadur	40.6	43.1	2.4	39.4	41.9	2.5	41.8	44.2	2.4
Pieta`	40.6	43.0	2.4	38.9	41.2	2.3	42.1	44.6	2.5
Mosta	36.2	38.6	2.4	35.4	37.5	2.1	37.0	39.8	2.7
Dingli	34.9	37.2	2.3	34.7	36.9	2.2	35.1	37.6	2.5
Qormi	38.2	40.5	2.3	37.1	39.3	2.2	39.3	41.7	2.4
Kirkop	34.9	37.2	2.3	33.4	36.2	2.9	36.4	38.1	1.7
Marsaxlokk	35.9	38.2	2.3	35.7	37.9	2.2	36.1	38.5	2.4
Marsaskala	33.2	35.4	2.3	33.3	35.3	2.1	33.1	35.5	2.5
Xewkija	39.0	41.3	2.3	37.6	40.0	2.5	40.4	42.5	2.1
Mqabba	34.9	37.2	2.2	34.1	36.0	1.9	35.7	38.3	2.6
Tarxien	37.1	39.2	2.2	36.3	38.4	2.1	37.8	40.0	2.2
Bormla	39.4	41.5	2.1	37.8	39.3	1.5	40.9	43.7	2.8
Rabat	42.2	44.3	2.0	40.4	42.5	2.1	44.0	46.0	2.0
Xagħra	39.0	41.0	2.0	37.9	40.0	2.1	40.0	42.0	2.0

Locality	Total			Males			Females		
	2005	2011	Difference	2005	2011	Difference	2005	2011	Difference
Floriana	46.4	48.4	2.0	45.0	46.3	1.4	47.8	50.3	2.5
Żabbar	37.0	38.9	1.9	36.2	38.1	2.0	37.8	39.6	1.8
Xgħajra	31.9	33.8	1.9	33.0	34.7	1.7	30.7	32.8	2.1
San Ġiljan	41.8	43.7	1.9	40.5	42.9	2.4	43.0	44.4	1.4
Żebbuġ	36.5	38.4	1.9	35.5	37.4	1.9	37.6	39.4	1.8
Żurrieq	37.3	39.1	1.8	36.5	38.3	1.7	38.0	39.9	2.0
Santa Venera	38.6	40.3	1.8	37.4	39.3	1.9	39.6	41.3	1.6
Għaxaq	36.2	37.9	1.7	35.8	37.4	1.6	36.6	38.4	1.9
Msida	39.8	41.5	1.7	37.9	39.4	1.5	41.7	43.7	2.0
Lija	38.7	40.4	1.7	36.9	39.4	2.5	40.4	41.4	1.0
Żejtun	40.0	41.7	1.7	38.5	40.0	1.6	41.6	43.3	1.8
Ħamrun	43.6	45.2	1.6	41.9	43.4	1.5	45.2	47.0	1.8
Qala	40.9	42.5	1.6	38.3	41.0	2.8	43.6	44.0	0.4
San Lawrenz	39.2	40.8	1.6	38.5	40.7	2.2	39.9	40.9	1.0
Sannat	38.2	39.8	1.6	37.8	39.6	1.8	38.7	40.0	1.3
Birkirkara	38.2	39.7	1.5	37.1	38.8	1.7	39.3	40.7	1.4
Sliema	46.6	48.2	1.5	44.1	46.2	2.1	48.9	49.9	1.0
Marsa	43.0	44.6	1.5	41.4	42.8	1.4	44.7	46.4	1.7
Gżira	41.4	42.9	1.5	39.7	41.1	1.4	43.1	44.7	1.6
Qrendi	38.1	39.6	1.5	37.4	39.1	1.7	38.8	40.0	1.2
Mġarr	35.9	37.3	1.4	35.6	37.3	1.8	36.2	37.2	1.1
Żebbuġ	38.1	39.5	1.4	37.2	38.4	1.1	38.9	40.6	1.7
Fontana	38.3	39.7	1.4	36.9	38.5	1.6	39.6	40.8	1.2
Għargħur	35.9	37.2	1.4	35.6	37.4	1.8	36.2	37.1	0.8
Valletta	44.2	45.5	1.2	41.6	43.4	1.7	46.7	47.4	0.8
Għarb	39.6	40.8	1.2	39.0	40.2	1.2	40.1	41.4	1.3
Balzan	43.2	44.4	1.2	40.8	41.8	1.0	45.4	46.9	1.5
Paola	42.8	43.8	1.0	40.8	42.0	1.2	44.8	45.8	1.0
Birżebbuġa	36.0	36.5	0.5	35.6	35.5	-0.1	36.4	37.7	1.3
Għasri	42.0	42.5	0.5	38.9	38.6	-0.3	45.1	46.4	1.3

References

Adams, R. J. *et al.* (2009) 'Effects of area deprivation on health risks and outcomes: a multilevel, cross-sectional, Australian population study.', *International journal of public health*, 54(3), pp. 183–192. doi: 10.1007/s00038-009-7113-x.

Agius, F. (1990) 'Health and social inequities in Malta.', *Social science & medicine (1982)*. England, 31(3), pp. 313–318. doi: 10.1016/0277-9536(90)90278-z.

Algren, M. H. *et al.* (2015) 'Health-risk behaviour in deprived neighbourhoods compared with non-deprived neighbourhoods: a systematic literature review of quantitative observational studies', *PloS one*. Public Library of Science San Francisco, CA USA, 10(10), p. e0139297.

Allik, M. *et al.* (2016) 'Developing a new small-area measure of deprivation using 2001 and 2011 census data from Scotland.', *Health & Place*. Urban Big Data Centre, University of Glasgow, 7 Lilybank Gardens, Glasgow G12 8RZ, Scotland: Elsevier Science, 39, pp. 122–130. doi: 10.1016/j.healthplace.2016.03.006.

Allik, M. (2018) *SocEpi*, *GitHub*. Available at: https://github.com/m-allik/SocEpi (Accessed: 26 August 2023).

Allik, M. *et al.* (2020) 'Creating small-area deprivation indices: a guide for stages and options.', *Journal of epidemiology and community health*, 74(1), pp. 20–25. doi: 10.1136/jech-2019-213255.

Amir, L. H. and Donath, S. M. (2008) 'Socioeconomic status and rates of breastfeeding in Australia: evidence from three recent national health surveys', *Medical Journal of Australia*. Wiley Online Library, 189(5), pp. 254–256.

Amram, O. *et al.* (2015) 'Socio economic status and traumatic brain injury amongst pediatric populations: a spatial analysis in greater vancouver', *International journal of environmental research and public health*. MDPI, 12(12), pp. 15594–15604.

Andersen, A. F. *et al.* (2008) 'Life-course socio-economic position, area deprivation and Type 2 diabetes: findings from the British Women's Heart and Health Study.', *Diabetic medicine : a journal of the British Diabetic Association*, 25(12), pp. 1462–1468. doi:

10.1111/j.1464-5491.2008.02594.x.

Auluck, A. *et al.* (2016) 'Socio-economic deprivation: a significant determinant affecting stage of oral cancer diagnosis and survival', *BMC cancer*. BioMed Central, 16(1), pp. 1–10.

Aungkulanon, S. *et al.* (2017) 'Area-level socioeconomic deprivation and mortality differentials in Thailand: results from principal component analysis and cluster analysis', *International Journal for Equity in Health*, 16(1), p. 117. doi: 10.1186/s12939-017-0613-z.

Australian Bureau of Statistics (1998) '1996 Census of Population and Housing. Socio-economic Indexes for Areas'. Australian Bureau of Statistics Canberra.

Avendano, M. and Berkman, L. F. (2014) 'Labor markets, employment policies, and health', *Social epidemiology*. Oxford University Press New York, pp. 182–233.

Bartley, M., Ferrie, J. and Montgomery, S. M. (2006) 'Health and labour market disadvantage: unemployment, non-employment, and job insecurity', *Social determinants of health*. Oxford University Press Oxford, 2, pp. 78–96.

Bell, N. *et al.* (2007) 'Towards the construction of place-specific measures of deprivation: a case study from the Vancouver metropolitan area', *The Canadian Geographer / Le Géographe canadien*, 51(4), pp. 444–461. doi: https://doi.org/10.1111/j.1541-0064.2007.00191.x.

Bell, N. and Hayes, M. V (2012) 'The Vancouver Area Neighbourhood Deprivation Index (VANDIX): a census-based tool for assessing small-area variations in health status.', *Canadian journal of public health = Revue canadienne de sante publique*. Canada, 103(8 Suppl 2), pp. S28-32.

Bellis, M. A. *et al.* (2016) 'The alcohol harm paradox: using a national survey to explore how alcohol may disproportionately impact health in deprived individuals', *BMC Public Health*, 16(1), p. 111. doi: 10.1186/s12889-016-2766-x.

Bender, A. M. *et al.* (2015) 'Neighborhood deprivation is strongly associated with participation in a population-based health check.', *PloS one*, 10(6), p. e0129819. doi: 10.1371/journal.pone.0129819.

Bleich, S. N. *et al.* (2012) 'Health inequalities: trends, progress, and policy', *Annual review of public health*. Annual Reviews, 33, pp. 7–40.

Blumenshine, P. *et al.* (2010) ‘Socioeconomic Disparities in Adverse Birth Outcomes: A Systematic Review’, *American Journal of Preventive Medicine*, 39(3), pp. 263–272. doi: https://doi.org/10.1016/j.amepre.2010.05.012.

Breeze, E. *et al.* (2005) ‘Area deprivation, social class, and quality of life among people aged 75 years and over in Britain.’, *International journal of epidemiology*, 34(2), pp. 276–283. doi: 10.1093/ije/dyh328.

Butler, D. C. *et al.* (2010) ‘Use of measures of socioeconomic deprivation in planning primary health care workforce and defining health care need in Australia’, *Australian Journal of Rural Health*. Wiley Online Library, 18(5), pp. 199–204.

Cabezas-Rodríguez, A., Utzet, M. and Bacigalupe, A. (2021) ‘Which are the intermediate determinants of gender inequalities in mental health?: A scoping review’, *International Journal of Social Psychiatry*. Sage Publications Sage UK: London, England, 67(8), pp. 1005–1025.

Cabrera-Barona, P. *et al.* (2015) ‘A multi-criteria spatial deprivation index to support health inequality analyses.’, *International journal of health geographics*, 14, p. 11. doi: 10.1186/s12942-015-0004-x.

Carstairs, V. and Morris, R. (1991) *Deprivation and Health in Scotland*. Aberdeen University Press. Available at: https://books.google.com.mt/books?id=H2NyQgAACAAJ.

Cass, A. *et al.* (2001) ‘Social disadvantage and variation in the incidence of end-stage renal disease in Australian capital cities’, *Australian and New Zealand journal of public health*. Elsevier, 25(4), pp. 322–326.

Census Office Malta (1968) *Malta Census 1967 - Preliminary report*. Valletta.

Chandola, T. and Conibere, R. (2015) ‘Social Exclusion, Social Deprivation and Health’, in Wright, J. D. (ed.) *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Second Edi. Oxford: Elsevier, pp. 285–290. doi: https://doi.org/10.1016/B978-0-08-097086-8.14087-5.

Chang, K. *et al.* (2023) ‘Ultra-processed food consumption, cancer risk and cancer mortality: a large-scale prospective analysis within the UK Biobank.’, *EClinicalMedicine*. England, 56, p. 101840. doi: 10.1016/j.eclinm.2023.101840.

Chaparro, M. P. *et al.* (2018) 'Neighborhood deprivation and biomarkers of health in Britain: the mediating role of the physical environment.', *BMC public health*, 18(1), p. 801. doi: 10.1186/s12889-018-5667-3.

Choi, M. H. *et al.* (2011) 'Deprivation and mortality at the town level in Busan, Korea: an ecological study', *Journal of preventive medicine and public health = Yebang Uihakhoe chi*. The Korean Society for Preventive Medicine, 44(6), pp. 242–248. doi: 10.3961/jpmph.2011.44.6.242.

Chyn, E. and Katz, L. F. (2021) 'Neighborhoods matter: Assessing the evidence for place effects', *Journal of Economic Perspectives*. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418, 35(4), pp. 197–222.

Cubbin, C. *et al.* (2008) 'Is neighborhood deprivation independently associated with maternal and infant health? Evidence from Florida and Washington.', *Maternal and child health journal*, 12(1), pp. 61–74. doi: 10.1007/s10995-007-0225-0.

Cuschieri, S. *et al.* (2017) 'The Effects of Socioeconomic Determinants on Hypertension in a Cardiometabolic At-Risk European Country', *International Journal of Hypertension*. Edited by F. Veglio. Hindawi, 2017, p. 7107385. doi: 10.1155/2017/7107385.

Cuschieri, S., Calleja, N. and Mamo, J. (2022) 'Health Inequities Exist in Europe: Are Spatial Health Inequities Present in the Small State of Malta?', *SAGE Open*. SAGE Publications Sage CA: Los Angeles, CA, 12(1), p. 21582440221082124.

Deguara, M., England, K. and Azzopardi Muscat, N. (2017) 'A review of the impact of social determinants on health in Malta', *The Synapse*. Medical Portals Ltd., Volume 16(3), pp. 6–9.

DeWalt, D. A. *et al.* (2004) 'Literacy and health outcomes: a systematic review of the literature', *Journal of general internal medicine*. Springer, 19, pp. 1228–1239.

Diez Roux, A. V (2018) 'Neighborhoods and Health'. Edited by D. T. Duncan and I. Kawachi. Oxford University Press. doi: 10.1093/oso/9780190843496.001.0001.

Directorate for Health Information and Research Malta (2023) *National Cancer Register*. Available at: https://deputyprimeminister.gov.mt/en/dhir/Pages/Registries/cancers.aspx (Accessed: 7 April 2023).

Egli, V. *et al.* (2020) 'Deprivation matters: understanding associations between neighbourhood deprivation, unhealthy food outlets, unhealthy dietary behaviours and child body size using structural equation modelling.’, *Journal of epidemiology and community health*, 74(5), pp. 460–466. doi: 10.1136/jech-2019-213159.

Elgar, F. J. *et al.* (2015) ‘Socioeconomic inequalities in adolescent health 2002–2010: a time-series analysis of 34 countries participating in the Health Behaviour in School-aged Children study’, *The Lancet*. Elsevier, 385(9982), pp. 2088–2095.

Ellwood, R. P. *et al.* (2004) ‘Relationship between area deprivation and the anticaries benefit of an oral health programme providing free fluoride toothpaste to young children.’, *Community dentistry and oral epidemiology*, 32(3), pp. 159–165. doi: 10.1111/j.1600-0528.2004.00150.x.

European Commission (2015) ‘METHODOLOGICAL GUIDELINES AND DESCRIPTION OF EU-SILC TARGET VARIABLES 2015 operation (Version november 2014)’, 065(november 2014). Available at: https://circabc.europa.eu/sd/a/2aa6257f-0e3c-4f1c-947f-76ae7b275cfe/DOCSILC065 operation 2014 VERSION reconciliated and early transmission October 2014.pdf.

European Parliament (2008) *Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses*. Strasbourg. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1412688957286&uri=CELEX:32008R0763.

Eurostat (2014) *Current depressive symptoms by sex, age and educational attainment level, European Health Interview Survey*. Available at: https://ec.europa.eu/eurostat/databrowser//product/view/HLTH_EHIS_MH1E (Accessed: 13 August 2023).

Eurostat (2023a) *Early school leavers from education and training, Statistics explained*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Early_leavers_from_education_and_training#:~:text=In 2022%2C an average of 9.6 %25 of young people aged,Romania (see Figure 1). (Accessed: 2 July 2023).

Eurostat (2023b) *EU labour force survey - methodology, Statistics explained*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_force_survey_-

_methodology#Other_concepts_and_definitions (Accessed: 17 July 2023).

Eurostat (2023c) *Fertility statistics, Statistics explained*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Fertility_statistics&stable=1 (Accessed: 2 July 2023).

Eurostat (2023d) *International Standard Classification of Education (ISCED), Statistics explained*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED) (Accessed: 17 July 2023).

Eurostat (2023e) *People at risk of poverty or social exclusion in 2022, News article*. Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/DDN-20230614-1#:~:text=In 2022%2C 95.3 million people,household with very low work (Accessed: 25 August 2023).

Eurostat (2023f) *Unemployment statistics and beyond, Statistics explained*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Unemployment_statistics_and_beyond (Accessed: 2 July 2023).

Fone, D. *et al.* (2007) 'Does social cohesion modify the association between area income deprivation and mental health? A multilevel analysis.', *International journal of epidemiology*, 36(2), pp. 338–345. doi: 10.1093/ije/dym004.

Fone, D. *et al.* (2014) 'Effect of neighbourhood deprivation and social cohesion on mental health inequality: a multilevel population-based longitudinal study.', *Psychological medicine*, 44(11), pp. 2449–2460. doi: 10.1017/S0033291713003255.

Fone, D. L. *et al.* (2003) 'Meningococcal disease and social deprivation: a small area geographical study in Gwent, UK.', *Epidemiology and infection*, 130(1), pp. 53–58.

Frohlich, N. and Mustard, C. (1996) 'A regional comparison of socioeconomic and health indices in a Canadian province.', *Social science & medicine (1982)*. England, 42(9), pp. 1273–1281. doi: 10.1016/0277-9536(95)00220-0.

Goodman, A. and Gatward, R. (2008) 'Who are we missing? Area deprivation and survey participation.', *European journal of epidemiology*, 23(6), pp. 379–387. doi: 10.1007/s10654-008-9248-0.

Government of Malta (1948) *Chapter 118 Census Act*. Malta. Available at: https://legislation.mt/eli/cap/118/eng.

Government of Malta (1957) *Notification of Cancer Act*. Malta. Available at: https://legislation.mt/eli/cap/154/19571212/eng.

Government of Malta (1993) *Chapter 363: Local Government Act*. Malta. Available at: https://legislation.mt/eli/cap/363/eng/pdf.

Government of Malta (2008) *Chapter 422 Malta Statistics Authority Aact*. Available at: https://legislation.mt/eli/cap/422/eng.

Government of Malta (2021) 'A Once-Only Grant for Former Malta Drydocks Employees who were Exposed to Asbestos', *The Malta Government Gazette No. 20,743*, 30 November, pp. 12587–12590. Available at: https://www.gov.mt/en/Government/DOI/Government Gazette/Documents/2021/11/Government Gazette - 30th November.pdf.

Gray, R. *et al.* (2008) 'Social inequalities in preterm birth in Scotland 1980-2003: findings from an area-based measure of deprivation.', *BJOG: An International Journal of Obstetrics & Gynaecology*. National Perinatal Epidemiology Unit, University of Oxford, Oxford, UK. ron.gray@npeu.ox.ac.uk: Wiley-Blackwell, 115(1), pp. 82–90. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=105860940&site=ehost-live.

Grech, A. (2016a) 'Assessing the economic impact of foreign workers in Malta', *Quarterly Review*, pp. 29–44.

Grech, A. (2016b) 'The Possible Impact of Pension Age Changes on Malta ' s Potential Output', (April), pp. 1–17. Available at: https://www.centralbankmalta.org/file.aspx?f=21344#:~:text=Until 2012%2C the pension age,final rise scheduled for 2026.

Hales, S. *et al.* (2003) 'Social deprivation and the public health risks of community drinking water supplies in New Zealand.', *Journal of epidemiology and community health*, 57(8), pp. 581–583.

Havard, S. *et al.* (2008) 'A small-area index of socioeconomic deprivation to capture health inequalities in France', *Social Science and Medicine*, 67(12), pp. 2007–2016. doi:

10.1016/j.socscimed.2008.09.031.

Hosseinpoor, A. R. and Bergen, N. (2016) ‘Area-based units of analysis for strengthening health inequality monitoring.’, *Bulletin of the World Health Organization*. Switzerland, 94(11), pp. 856–858. doi: 10.2471/BLT.15.165266.

Hosseinpoor, A. R., Bergen, N. and Schlotheuber, A. (2015) ‘Promoting health equity: WHO health inequality monitoring at global and national levels.’, *Global health action*. United States, 8, p. 29034. doi: 10.3402/gha.v8.29034.

Huisman, M. *et al.* (2013) ‘Socioeconomic inequalities in mortality rates in old age in the World Health Organization Europe region’, *Epidemiologic reviews*. Oxford University Press, 35(1), pp. 84–97.

Hunn, L., Teague, B. and Fisher, P. (2023) ‘Literacy and mental health across the globe: a systematic review’, *Mental Health and Social Inclusion*. Emerald Publishing Limited.

Ingraham, N. E. *et al.* (2020) ‘Racial/Ethnic Disparities in Hospital Admissions from COVID-19 and Determining the Impact of Neighborhood Deprivation and Primary Language.’, *medRxiv : the preprint server for health sciences*. doi: 10.1101/2020.09.02.20185983.

International Labour Office (2023) *International Standard Classification of Occupations (ISCO), Concepts and definitions*. Available at: https://ilostat.ilo.org/resources/concepts-and-definitions/classification-occupation/ (Accessed: 17 July 2023).

Isaksen, I. M. and Dankel, S. N. (2023) ‘Ultra-processed food consumption and cancer risk: A systematic review and meta-analysis’, *Clinical Nutrition*, 42(6), pp. 919–928. doi: https://doi.org/10.1016/j.clnu.2023.03.018.

Janghorbani, M., Jones, R. B. and Nelder, R. (2006) ‘Neighbourhood deprivation and excess coronary heart disease mortality and hospital admissions in Plymouth, UK: an ecological study.’, *Acta cardiologica*, 61(3), pp. 313–320. doi: 10.2143/AC.61.3.2014834.

Jarman, B. (1983) ‘Identification of underprivileged areas.’, *British medical journal (Clinical research ed.)*, 286(6379), pp. 1705–1709. doi: 10.1136/bmj.286.6379.1705.

Jeong, B.-G. *et al.* (2006) ‘[The relationship between regional material deprivation and the standardized mortality ratio of the community residents aged 15-64 in Korea].’, *Journal of* *preventive medicine and public health = Yebang Uihakhoe chi*, 39(1), pp. 46–52.

Jobsplus (2023) *Employed Third Country Nationals by Major Occupation Group (2012-22)*. Available at: https://jobsplus.gov.mt/resources/publication-statistics-mt-mt-en-gb/labour-market-information/foreigners-data#title3.4 (Accessed: 26 August 2023).

Jubber, I. *et al.* (2023) 'Epidemiology of Bladder Cancer in 2023: A Systematic Review of Risk Factors', *European Urology*. Elsevier.

Krewski, D. *et al.* (2009) 'Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality.', *Research report (Health Effects Institute)*. United States, (140), pp. 5–36.

Kurani, S. *et al.* (2020) 'Place, poverty and prescriptions: a cross-sectional study using Area Deprivation Index to assess opioid use and drug-poisoning mortality in the USA from 2012 to 2017.', *BMJ open*, 10(5), p. e035376. doi: 10.1136/bmjopen-2019-035376.

Kuznetsov, L. *et al.* (2011) 'Associations between regional socioeconomic deprivation and cancer risk: Analysis of Population-based Cancer Registry data from Bavaria, Germany', *Preventive Medicine*, 53(4), pp. 328–330. doi: https://doi.org/10.1016/j.ypmed.2011.07.021.

Kwate, N. O. A. *et al.* (2009) 'Inequality in obesigenic environments: fast food density in New York City.', *Health & place*. England, 15(1), pp. 364–373. doi: 10.1016/j.healthplace.2008.07.003.

Larsen, K., Rydz, E. and Peters, C. E. (2023) 'Inequalities in Environmental Cancer Risk and Carcinogen Exposures: A Scoping Review', *International Journal of Environmental Research and Public Health*. doi: 10.3390/ijerph20095718.

Leyland, A. H. *et al.* (2007) 'Cause-specific inequalities in mortality in Scotland: two decades of change. A population-based study', *BMC public health*. Springer, 7, pp. 1–12.

Leyland, A. H. and Groenewegen, P. P. (2020) *Multilevel modelling for public health and health services research: health in context*. Springer Nature.

Li, X. *et al.* (2015) 'Neighborhood Deprivation and Lung Cancer Incidence and Mortality: A Multilevel Analysis from Sweden', *Journal of Thoracic Oncology*, 10(2), pp. 256–263. doi: https://doi.org/10.1097/JTO.0000000000000417.

Lian, M. *et al.* (2011) 'Geographic variation in colorectal cancer survival and the role of small-area socioeconomic deprivation: a multilevel survival analysis of the NIH-AARP Diet and Health Study Cohort', *American journal of epidemiology*. 2011/08/11. Oxford University Press, 174(7), pp. 828–838. doi: 10.1093/aje/kwr162.

Lian, M. *et al.* (2014) 'Neighborhood socioeconomic deprivation, tumor subtypes, and causes of death after non-metastatic invasive breast cancer diagnosis: a multilevel competing-risk analysis.', *Breast cancer research and treatment*, 147(3), pp. 661–670. doi: 10.1007/s10549-014-3135-z.

Lopez-De Fede, A. *et al.* (2016) 'Comparison of small-area deprivation measures as predictors of chronic disease burden in a low-income population.', *International journal for equity in health*, 15, p. 89. doi: 10.1186/s12939-016-0378-9.

Lortet-Tieulent, J. *et al.* (2019) 'Reducing social inequalities in cancer: evidence and priorities for research'. International Agency for Research on Cancer.

Luben, R. *et al.* (2019) 'Residential area deprivation and risk of subsequent hospital admission in a British population: the EPIC-Norfolk cohort.', *BMJ open*, 9(12), p. e031251. doi: 10.1136/bmjopen-2019-031251.

Lumme, S. *et al.* (2015) 'A Monte Carlo method to estimate the confidence intervals for the concentration index using aggregated population register data', *Health Services and Outcomes Research Methodology*. Springer, 15, pp. 82–98.

Macdonald, L. *et al.* (2018) 'Do "environmental bads" such as alcohol, fast food, tobacco, and gambling outlets cluster and co-locate in more deprived areas in Glasgow City, Scotland?', *Health & place*. England, 51, pp. 224–231. doi: 10.1016/j.healthplace.2018.04.008.

Mackenbach, J. P. *et al.* (2003) 'Widening socioeconomic inequalities in mortality in six Western European countries', *International journal of epidemiology*. Oxford University Press, 32(5), pp. 830–837.

Mackenbach, J. P. *et al.* (2008) 'Socioeconomic inequalities in health in 22 European countries', *New England journal of medicine*. Mass Medical Soc, 358(23), pp. 2468–2481.

Mackenbach, J. P. (2012) 'The persistence of health inequalities in modern welfare states: the explanation of a paradox.', *Social science & medicine (1982)*. England, 75(4), pp. 761–769. doi: 10.1016/j.socscimed.2012.02.031.

Mackenbach, J. P. (2014) 'The persistence of health inequalities in modern welfare states: the role of health behaviours', *Eurohealth*. World Health Organization. Regional Office for Europe, 20(2), pp. 6–9.

Mackenbach, J. P., Meerding, W. J. and Kunst, A. E. (2011) 'Economic costs of health inequalities in the European Union', *Journal of Epidemiology & Community Health*. BMJ Publishing Group Ltd, 65(5), pp. 412–419.

Marmarà, D., Marmarà, V. and Hubbard, G. (2017) 'Health beliefs, illness perceptions and determinants of breast screening uptake in Malta: a cross-sectional survey.', *BMC public health*. England, 17(1), p. 416. doi: 10.1186/s12889-017-4324-6.

Marmot, M. (2001) 'From Black to Acheson: two decades of concern with inequalities in health. A celebration of the 90th birthday of Professor Jerry Morris.', *International journal of epidemiology*. England, 30(5), pp. 1165–1171. doi: 10.1093/ije/30.5.1165.

Marmot, M. *et al.* (2012) 'WHO European review of social determinants of health and the health divide', *The Lancet*. Elsevier, 380(9846), pp. 1011–1029.

Marmot, M. G. *et al.* (1991) 'Health inequalities among British civil servants: the Whitehall II study', *The Lancet*. Elsevier, 337(8754), pp. 1387–1393.

Maughan, B. and Carroll, J. (2006) 'Literacy and mental disorders', *Current Opinion in Psychiatry*. LWW, 19(4), pp. 350–354.

Mehta, B. *et al.* (2020) 'Community Deprivation Index and Discharge Destination after Elective Hip Replacement.', *Arthritis care & research*. doi: 10.1002/acr.24145.

Meijer, M. *et al.* (2012) 'Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies', *Social science & medicine*. Elsevier, 74(8), pp. 1204–1212.

Menvielle, G. *et al.* (2010) 'Occupational exposures contribute to educational inequalities in lung cancer incidence among men: Evidence from the EPIC prospective cohort study.', *International journal of cancer*. United States, 126(8), pp. 1928–1935. doi: 10.1002/ijc.24924.

Messer, L. C. *et al.* (2006) ‘The development of a standardized neighborhood deprivation index’, *Journal of Urban Health*. United States, 83(6), pp. 1041–1062. doi: 10.1007/s11524-006-9094-x.

Mihor, A. *et al.* (2020) ‘Socioeconomic inequalities in cancer incidence in Europe: a comprehensive review of population-based epidemiological studies’, *Radiology and oncology*, 54(1), pp. 1–13.

Ministry for Health Malta (2017) *The National Cancer Plan for the Maltese Islands 2017-2021*. Available at: https://health.gov.mt/wp-content/uploads/2023/04/The_National_Cancer_Plan_for_the_Maltese_Islands_2017–2021_EN.pdf.

Ministry for Health Malta (2019) ‘Building Resilience; Transforming Services: A Mental Health Strategy for Malta 2020 - 2030’, (JULY 2019), p. 52. Available at: https://deputyprimeminister.gov.mt/en/Documents/National-Health-Strategies/Mental_Health_Strategy_EN.pdf.

Ministry for Health Malta (2022) ‘A National Health Systems Strategy for Malta 2023-2030 Investing successfully for a Healthy Future’, (December 2022), p. 80.

Ministry of Housing Communities and Local Government (2019) ‘The English Indices of Deprivation 2019: Technical report’, (September).

Murray, E. T. *et al.* (2022) ‘Measuring the health of people in places: A scoping review of OECD member countries’, *Health and Place*. Elsevier Ltd, 73, p. 102731. doi: 10.1016/j.healthplace.2021.102731.

Musu, M. (2015) ‘Strengthening the Pension System A Strategy for an Adequate and Sustainable Maltese Pension System Pensions Strategy Group’, (June), pp. 1–131. Available at: https://family.gov.mt/wp-content/uploads/2021/05/Pensions-Report.pdf.

National Cancer Register (2023a) *All malignant cancers (excluding non-melanoma skin cancers) 2007 - 2019*. Gwardamangia, Malta. Available at: https://deputyprimeminister.gov.mt/en/dhir/Documents/Cancer/2020/cancers pdfs/All cancers 2019.pdf.

National Cancer Register (2023b) *Cancer of the colorecturm (2007-2019)*. Gwardamangia, Malta. Available at: https://deputyprimeminister.gov.mt/en/dhir/Documents/Cancer/2020/cancers pdfs/Colorectal Cancer 2019.pdf.

National Cancer Register (2023c) *Cancer of the female breast (2007 - 2019)*. Gwardamangia, Malta. Available at: https://deputyprimeminister.gov.mt/en/dhir/Documents/Cancer/2020/cancers pdfs/Cancer of the female breast 2019.pdf.

National Cancer Register (2023d) *Lung cancer (2007 - 2019)*. Gwardamangia, Malta. Available at: https://deputyprimeminister.gov.mt/en/dhir/Documents/Cancer/2020/cancers pdfs/Lung 2019.pdf.

National Cancer Register (2023e) *Prostate cancer (2007-2019)*. Gwardamangia, Malta. Available at: https://deputyprimeminister.gov.mt/en/dhir/Documents/Cancer/2020/cancers pdfs/Prostate 2019.pdf.

National Statistics Office (2001) *Demographic Review 2000*.

National Statistics Office (2007) 'Census of Population 2005 Volume 1 Population'.

National Statistics Office (2009) *Census of Population 2005 Volume 2 Dwellings*. Available at: http://www.nso.gov.mt/docs/CensusVol2.pdf.

National Statistics Office (2014a) *Census 2011 metadata, National Reference Metadata in Euro SDMX Metadata Structure (ESMS)*. Available at: https://ec.europa.eu/eurostat/cache/metadata/EN/cens_11r_esmscs_mt.htm (Accessed: 17 July 2023).

National Statistics Office (2014b) 'Census of the population and housing 2011: Final report', pp. 1–365. Available at: https://nso.gov.mt/en/publicatons/Publications_by_Unit/Documents/01_Methodology_and_Research/Census2011_FinalReport.pdf.

National Statistics Office (2017) *Regional Statistics Malta2017 Edition*. Available at: https://nso.gov.mt/wp-content/uploads/Regional-Statistics-MALTA-2017-Edition.pdf.

National Statistics Office (2019) 'World Refugee Day : 20 June 2019', (June), pp. 1–14. Available at: https://nso.gov.mt/en/News_Releases/Documents/2021/06/News2021_109.pdf.

National Statistics Office (2020) ‘EU-SILC 2019 : Material Deprivation and Monetary Poverty’, (December). Available at: https://nso.gov.mt/en/News_Releases/Documents/2020/12/News2020_202.pdf.

National Statistics Office (2023a) *Census in Malta, Sources and methods*. Available at: http://www.nso.gov.mt/site/page.aspx?pageid=577 (Accessed: 2 July 2023).

National Statistics Office (2023b) *Census of Population and Housing 2021: Final Report: Population, migration and other social characteristics (Volume 1)*. Valletta. Available at: https://nso.gov.mt/themes_publications/census-of-population-and-housing-2021-final-report-population-migration-and-other-social-characteristics-volume-1/.

National Statistics Office (2023c) *Districts (Local Administrative Unit), Metadata*. Available at: https://metadata.nso.gov.mt/classificationdetails.aspx?id=Districts (Local Administrative Unit).

National Statistics Office (2023d) *End of year population estimates by locality, sex and broad citizenship category: 2021 - 2022, Selected indicators*. Available at: https://nso.gov.mt/population/ (Accessed: 25 August 2023).

National Statistics Office (2023e) *EU-SILC 2021: Well-being, Social and Health Indicators, News release*. Available at: https://nso.gov.mt/eu-silc-2021-well-being-social-and-health-indicators/ (Accessed: 2 July 2023).

National Statistics Office (2023f) *EU-SILC 2022: Salient Indicators, News release*.

National Statistics Office (2023g) *Labour Force Survey: Q4/2022, News release*. Available at: https://nso.gov.mt/labour-force-survey-q4-2022/ (Accessed: 26 August 2023).

National Statistics Office (2023h) *Statistical concepts - Dependent child, Metadata*. Available at: https://metadata.nso.gov.mt/ConceptDetails.aspx?id=466 (Accessed: 16 July 2023).

National Statistics Office (2023i) *Statistical concepts - Working age population, Metadata*. Available at: https://metadata.nso.gov.mt/ConceptDetails.aspx?id=2118 (Accessed: 16 July 2023).

National Statistics Office (2023j) *World Population Day: 11 July 2023, News release*. Available at: https://nso.gov.mt/world-population-day-11-july-2023/#:~:text=NSO Malta

%7C World Population Day%3A 11 July 2023 - NSO Malta (Accessed: 15 August 2023).

National Statistics Office (2024) *Population and migration: 2012-2022 (including intercensal revisions), News release*. Available at: https://nso.gov.mt/intercensal-population-revisions-2012-2021/ (Accessed: 27 January 2024).

Neasham, D. *et al.* (2001) 'Stillbirth and neonatal mortality due to congenital anomalies: temporal trends and variation by small area deprivation scores in England and Wales, 1986–96.', *Paediatric & Perinatal Epidemiology*, 15(4), pp. 364–373. Available at: http://10.0.4.22/j.1365-3016.2001.0379a.x.

Norman, P. *et al.* (2011) 'Rising premature mortality in the UK's persistently deprived areas: Only a Scottish phenomenon?', *Social science & medicine*. Elsevier, 73(11), pp. 1575–1584.

Nutbeam, D. and Lloyd, J. E. (2021) 'Understanding and responding to health literacy as a social determinant of health', *Annu Rev Public Health*, 42(1), pp. 159–173.

OECD (2021) *Potential years of life lost, OECD Data*. Available at: https://data.oecd.org/healthstat/potential-years-of-life-lost.htm#:~:text=The calculation of Potential Years,used in OECD Health Statistics). (Accessed: 26 August 2023).

Office of the Commissioner for Mental Health (2014) *Health Literacy Survey Malta 2014*. Available at: https://deputyprimeminister.gov.mt/en/CommMentalHealth/Documents/health-literacy/health-literacy-survey-malta-2014.pdf.

Pampalon, R. *et al.* (2014) 'Validation of a deprivation index for public health: a complex exercise illustrated by the Quebec index.', *Chronic diseases and injuries in Canada*. Canada, 34(1), pp. 12–22.

Panczak, R. *et al.* (2012) 'A Swiss neighbourhood index of socioeconomic position: development and association with mortality', *Journal of Epidemiology and Community Health*, 66(12), pp. 1129 LP – 1136. doi: 10.1136/jech-2011-200699.

Park, E.-J. *et al.* (2010) 'Area deprivation, individual socioeconomic position and smoking among women in South Korea.', *Tobacco control*, 19(5), pp. 383–390. doi: 10.1136/tc.2009.032763.

Powell-Wiley, T. M. *et al.* (2019) 'The relationship between neighborhood socioeconomic deprivation and telomere length: The 1999-2002 National Health and Nutrition Examination

Survey’, *SSM - population health*. Elsevier, 10, p. 100517. doi: 10.1016/j.ssmph.2019.100517.

Powell-Wiley, T. M. *et al.* (2020) ‘The relationship between neighborhood socioeconomic deprivation and telomere length: The 1999-2002 National Health and Nutrition Examination Survey.’, *SSM - population health*, 10, p. 100517. doi: 10.1016/j.ssmph.2019.100517.

Raaschou-Nielsen, O. *et al.* (2013) ‘Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE).’, *The Lancet. Oncology*. England, 14(9), pp. 813–822. doi: 10.1016/S1470-2045(13)70279-1.

Riaz, S. P. *et al.* (2011) ‘Lung Cancer Incidence and Survival in England: An Analysis by Socioeconomic Deprivation and Urbanization’, *Journal of Thoracic Oncology*, 6(12), pp. 2005–2010. doi: https://doi.org/10.1097/JTO.0b013e31822b02db.

Rossen, L. M. (2014) ‘Neighbourhood economic deprivation explains racial/ethnic disparities in overweight and obesity among children and adolescents in the U.S.A.’, *Journal of epidemiology and community health*, 68(2), pp. 123–129. doi: 10.1136/jech-2012-202245.

Salmond, C., Crampton, P. and Sutton, F. (1998) ‘NZDep91: A New Zealand index of deprivation.’, *Australian and New Zealand journal of public health*. Australia, 22(7), pp. 835–837. doi: 10.1111/j.1467-842x.1998.tb01505.x.

Sánchez-Cantalejo, C., Ocana-Riola, R. and Fernández-Ajuria, A. (2008) ‘Deprivation index for small areas in Spain.’, *Social Indicators Research*. Springer Science & Business Media B.V., 89(2), pp. 259–273. Available at: http://10.0.3.239/s11205-007-9114-6.

Schraw, J. M. *et al.* (2020) ‘Area deprivation is associated with poorer overall survival in children with acute lymphoblastic leukemia.’, *Pediatric blood & cancer*. United States, 67(9), p. e28525. doi: 10.1002/pbc.28525.

Singh, G. K. (2003) ‘Area Deprivation and Widening Inequalities in US Mortality, 1969–1998’, *American Journal of Public Health*. American Public Health Association, 93(7), pp. 1137–1143. doi: 10.2105/AJPH.93.7.1137.

Skapinakis, P. *et al.* (2005) ‘Mental health inequalities in Wales, UK: multi-level investigation of the effect of area deprivation.’, *The British journal of psychiatry : the journal* *of mental science*, 186, pp. 417–422. doi: 10.1192/bjp.186.5.417.

Smith, D. *et al.* (2019) 'The breast cancer paradox: A systematic review of the association between area-level deprivation and breast cancer screening uptake in Europe.', *Cancer epidemiology*, 60, pp. 77–85. doi: 10.1016/j.canep.2019.03.008.

Soobader, M. *et al.* (2001) 'Using aggregate geographic data to proxy individual socioeconomic status: does size matter?', *American journal of public health*, 91(4), pp. 632–636. doi: 10.2105/ajph.91.4.632.

Stavrou, E. P. *et al.* (2009) 'Adenocarcinoma of the oesophagus: incidence and survival rates in New South Wales, 1972–2005', *Medical journal of Australia*. Wiley Online Library, 191(6), pp. 310–314.

Stimpson, J. P. *et al.* (2007) 'Neighborhood Deprivation is associated with lower levels of serum carotenoids among adults participating in the Third National Health and Nutrition Examination Survey.', *Journal of the American Dietetic Association*, 107(11), pp. 1895–1902. doi: 10.1016/j.jada.2007.08.016.

Superintendence of Public Health Ministry for Health, the E. and C. C. (2012) 'A Healthy Weight for Life: A National Strategy for Malta (2012–2020)'. Superintendence of Public Health Malta.

Tello, J. E. *et al.* (2005) 'A census-based socio-economic status (SES) index as a tool to examine the relationship between mental health services use and deprivation.', *Social science & medicine (1982)*. England, 61(10), pp. 2096–2105. doi: 10.1016/j.socscimed.2005.04.018.

The Scottish Government (2020) 'SIMD 2020 technical notes'. Available at: https://www.gov.scot/publications/simd-2020-technical-notes/.

Tobias, M. I. and Cheung, J. (2003) 'Monitoring health inequalities: life expectancy and small area deprivation in New Zealand.', *Population health metrics*, 1(1), p. 2.

Torres-Cintrón, M. *et al.* (2012) 'Using a socioeconomic position index to assess disparities in cancer incidence and mortality, Puerto Rico, 1995-2004.', *Preventing chronic disease*, 9, p. E15.

Townsend, P. (1987) 'Deprivation', *Journal of social policy*. Cambridge University Press, 16(2), pp. 125–146.

Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and deprivation: inequality and the North*. Taylor & Francis.

Tuliani, T. A. *et al.* (2017) ‘Impact of Area Deprivation Index on Coronary Stent Utilization in a Medicare Nationwide Cohort.’, *Population Health Management*. 1 Loma Linda University Medical Center , Loma Linda, California.: Mary Ann Liebert. doi: 10.1089/pop.2016.0086.

UNECE (2019) *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses, Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses*. doi: 10.18356/48165223-en.

Unit of Social Determinants of Health (no date) *National Survey on Social Determinants of Health*. Available at: https://deputyprimeminister.gov.mt/en/sdh/Pages/research.aspx (Accessed: 1 October 2021).

Urwin, M. *et al.* (1998) ‘Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation.’, *Annals of the rheumatic diseases*, 57(11), pp. 649–655.

Visser, K. *et al.* (2021) ‘Neighbourhood deprivation effects on young people’s mental health and well-being: A systematic review of the literature’, *Social Science & Medicine*. Elsevier, 270, p. 113542.

Walker, B. B., Schuurman, N. and Hameed, S. M. (2014) ‘A GIS-based spatiotemporal analysis of violent trauma hotspots in Vancouver, Canada: identification, contextualisation and intervention’, *BMJ open*. British Medical Journal Publishing Group, 4(2), p. e003642.

Wang, E. *et al.* (2020) ‘Social Determinants of Pregnancy-Related Mortality and Morbidity in the United States: A Systematic Review.’, *Obstetrics and gynecology*. United States, 135(4), pp. 896–915. doi: 10.1097/AOG.0000000000003762.

Wang, Z. *et al.* (2020) ‘Construction of an area deprivation index for 2869 counties in China: a census-based approach.’, *Journal of epidemiology and community health*. England. doi: 10.1136/jech-2020-214198.

Weng, M. *et al.* (2016) ‘Area deprivation and liver cancer prevalence in shenzhen, china: A spatial approach based on social indicators.’, *Social Indicators Research*. Su, Shiliang, School of Resource and Environmental Sciences, Wuhan University, No. 129 Luoyu Rd, Wuhan, China: Springer. doi: 10.1007/s11205-016-1358-6.

Wijayaratna, S. *et al.* (2021) 'Socioeconomic status and risk factors for complications in young people with type 1 or type 2 diabetes: a cross-sectional study', *BMJ Open Diabetes Research and Care*. BMJ Specialist Journals, 9(2), p. e002485.

Winkleby, M., Sundquist, K. and Cubbin, C. (2007) 'Inequities in CHD incidence and case fatality by neighborhood deprivation.', *American journal of preventive medicine*, 32(2), pp. 97–106. doi: 10.1016/j.amepre.2006.10.002.

World Health Organisation (2008) *Closing the gap in a generation: health equity through action on the social determinants of health: Commission on Social Determinants of Health final report*. World Health Organization.

World Health Organisation (2020) 'Spotlight on adolescent health and well-being. Findings from the 2017'. World Health Organization. Regional Office for Europe.

World Health Organisation (2022) *Fact sheets: Cancer*. Available at: https://www.who.int/news-room/fact-sheets/detail/cancer (Accessed: 7 April 2023).

World Health Organisation (2023) *Premature mortality from noncommunicable disease, The Global Health Observatory: Indicator Metadata Registry List*. Available at: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/3411 (Accessed: 26 August 2023).

Young, G. S. *et al.* (2012) 'Differential exposure to hazardous air pollution in the United States: a multilevel analysis of urbanization and neighborhood socioeconomic deprivation.', *International journal of environmental research and public health*, 9(6), pp. 2204–2225. doi: 10.3390/ijerph9062204.

Znaor, A. *et al.* (2013) 'Cancer incidence and mortality patterns in South Eastern Europe in the last decade: Gaps persist compared with the rest of Europe', *European Journal of Cancer*, 49(7), pp. 1683–1691. doi: https://doi.org/10.1016/j.ejca.2012.11.030.