Socioeconomic Inequalities in Health

Cecilia Potente

University of Oxford

Nuffield College

A thesis presented for the degree of Doctor of Philosophy in Sociology

Trinity Term 2019

Word count: 41,353

Socioeconomic Inequalities in Health

Cecilia Potente

University of Oxford, Nuffield College

A thesis presented for the degree of Doctor of Philosophy in Sociology

Trinity Term 2019

Abstract

The aim of this thesis is to advance the understanding of how the relationship between health and socioeconomic status changes over the life-course and over time. Three different studies which address several important aspects of the socioeconomic inequalities in health are presented. First, the existence of a socioeconomic gradient in functional health at the end of life in England is investigated. The objective is to contribute to the empirical evidence concerning the existence of a disappearing association between socioeconomic status and health in old age, by analysing whether socioeconomic status is associated with disability trajectories in the final years of life. The second study examines how the changing demographic composition of different educational groups over time can help to explain the widening educational gap in self-reported health in the United States, connecting family demography with health inequalities. The third study explores whether, and to what extent, education has a causal effect on cancer outcomes in England and Wales. This thesis contributes to the existing literature in at least three ways. First, it conceptualises the last years of life as a life stage where socioeconomic inequalities in health might emerge. Differences in the disability-related pathways preceding death were found only wealth. Second, it highlights the increasingly important role of demographic factors and marital status in particular, when studying health inequalities over time. Indeed, the findings suggested an increasingly relevant contribution of marital status over time for educational inequalities in health in the United States, despite marriage representing only a small part of the explanation. Furthermore, it investigated for the first time the causal effect of education on survival after cancer diagnosis, providing evidence for the absence of a causal impact of an additional year of compulsory schooling on cancer incidence and survival after two educational reforms.

Acknowledgements

First and foremost, I would like to thank my supervisor, Christiaan Monden, for his excellent mentorship and guidance. His constant encouragement and enthusiasm for my ideas has been precious for the development of this thesis. I am very grateful for the many hours he has spent discussing ideas with me and reading drafts. He has been a great source of learning throughout the DPhil. Moreover, Christiaan always instilled me confidence and enthusiasm in the value of my research. I also am grateful to Francesco Billari, Melinda Mills and John Ermisch for their brilliant help in improving my thesis at different points. Nicola Barban and Patrick Prag for helping me in the development of my ideas. I would also thank the academic community at Department of Sociology and in Nuffield who gave me support, inspiration, and help in these years. In particular, I would like to thank David Kirk, Maria Sironi, Felix Tropf, Rob Gruijters, Tina Baier, Zachary VanWinkle, Stine Mollegard, Lucie Kalousova, Paula Sheppard, Per Engzell and Julia Behrman.

I would like to thank also ESRC, Clarendon Fund and Scatcherd European Scholarship, who kindly provided support for my doctoral research. Nuffield College was invaluable source of support. I believe that the interdisciplinary approach is essential for a complete and far-reaching research in the social sciences, and Nuffield was the perfect setting for my research. The Nuffield environment has been tremendously inspiring for developing a rigorous and cutting-edge research. In particular I would like to thank the senior tutor Eleni Kechagia-Ovseiko for the precious and constant support in my professional development. Finally, I would like to thank the Celsius Team and Longitudinal Study Team at Office for National Statistics for their kind and precious support, and in particular Rachel and Wei have been exceptional in their help during the different phases of the project.

The number of friends to thank is countless. Friendship has been a precious gift in Oxford. First, I would like to thank my Oxford friends. I thank Christian for being such patient, kind and caring friend, Marta A. for the amazing moments shared together and her affectionate friendship, Rey for his ability to understand things beyond surface and his amazing soup, Alice for being a loving friend and her three-course dinners, Pietro for the punchy discussions in the kitchen and his warm heart, Pia for her sunny attitude and affectionate friendship, Marta G. for being supportive and caring and for pushing me to go to the gym, Shan for her sweet friendship, Luigi for his loyal friendship, Andrea for her contagious smile, Alejandro for his joyful cheerfulness, Tom for the interesting discussions about physics, Effrosyni for being such a loving friend and an amazing window office mate, Matthias and Tabea for their graceful support, Simon for his caring friendship, Claudia for her advices and ability to listen, Ilona for her great sense of humour and loving support, Magda for her sweetness and cheerful smile. I would like to thank also Sophia, Merce, Fr. Dancho and Fr. Keith for their support and encouragement.

Great friends have been precious fellow travellers in my DPhil adventure. Our friendship made this doctorate in something much more than just a degree. I am grateful for this time shared with you and for those to come. In particular, I thank Nicolò for being such a supportive warm-hearted friend and our interesting discussions, Anna B. for her loving encouragement and her ability to make me see things always in a positive light, Isabel and Felix for many moments shared together in Abbey Road and for the many more that will come in Zurich, Rosie for being such an amazing and cheerful friend, Giacomo for his humour and the nice time spent together, Justin for the good discussions about politics, science and religion, Xuejie for her lovely support. I would like to thank also Meg and Julie for proofreading with interest this thesis. I thank the people with whom I shared many of these years in Nuffield beyond, in particular Artur, Jeff, Giovanni, Roberto, Apsara, Elisabetta D.C., Kelvin, Gareth, Rasa, Kinga, Riley and Liz. Finally, my gratitude goes to those who are accompanying me in my new adventure in Zurich, Johanna, Elizabeth, Wenjia, Alex and Justin.

Finally, I would like to thank my dear friends, despite the distance they have been such a great source of joy: Anna M. for her amazing and positive energy, Irene D. for her kindness and generosity, Irene F. for being such a good listener and affectionate friend. Also, my friends Laura, Beatrice, Sofia, Flavio, Yhidad, Ilaria S., Giulia D.P., Simone, Donatella and the amazing friends of Dolomiti adventures which have been precious support and source of great fun. My family has always been bright light in my life; without their support, encouragement and enthusiasm I would not have never been able to achieve this doctorate. In particular, I would like to thank my father, Luigi, for transmitting me passion for knowledge and enthusiasm about my academic work. He always remembered me to put interest and joy in studying since school is a “temple of knowledge”. I thank my mother, Maria Luisa, for her constant support, kind love and her good humour when things got harder. My brother, Gianluigi, has always been a pillar of support encouraging me to improve my skills, overcome the challenges with enthusiasm and be proud of my work. I would like to thank also Ilaria for offering me always positive thought and a sunny smile. My aunt Roberta, my uncles Gianni and Giuseppe and my cousin Alessandro have been precious support and source of joy in these years. Finally, I would like to thank my grandparents Margherita, Matilde, Antonio e Armando. Their unconditional love stays with me.

Table of Contents

Chapter 1

Introduction

1.1 Background

In the past century, the population has aged considerably. Simultaneously, educational expansion and economic development have changed the socioeconomic circumstances of many people. Consequently, socioeconomic inequality in health and mortality has become increasingly relevant. Therefore, the main objective of this doctoral thesis was to connect the changing socioeconomic landscape with inequalities in health and mortality, by investigating socioeconomic differences from both a micro-level and macro-level perspective. Moreover, family processes are also included as relevant aspects to study when analysing socioeconomic inequality in health. The following research questions were addressed in three different empirical studies: *How does the relationship between health and socioeconomic status (SES) change over the life-course and over time? What are the mechanisms driving these processes?*

This thesis contributes to three issues in the study of socioeconomic determinants of health, firstly examining the socioeconomic gradient in a specific part of the life-course, secondly investigating the role of demographic structure within different educational groups, and finally estimating the causal effect of education on an emerging morbidity such as cancer. Although these are quite different aspects of the socioeconomic gap in health and mortality, they are all extremely relevant and central in the context of population ageing and increasing differences among SES groups. This thesis investigates trends at the population-level in socioeconomic inequalities over time and decomposes them at the individual-level to better understand what is observed at the macro-level. In the first empirical paper, the focus is on socioeconomic differences associated with certain pathways towards death. The second study examines the role of marriage in the rising educational differences over cohorts in self-reported health in the United States. The third empirical analysis addresses how macro-policy changes can affect the relationship between education and health.

Chapter 2 examines the association between SES and disability in the last years of life in England, conceptualising the final years of life in terms of a life-course stage and analysing how SES is related to disability in this phase of life. The associations between socioeconomic factors and health-related trajectories in the last years of life is of central interest. Do socioeconomic inequalities exist in the disability pathways before death? Understanding where inequalities in health are located is essential to better design policy interventions as well as to guide health professionals. Moreover, this chapter contributes to the literature by examining how different socioeconomic indicators relate to functional limitations in the final years of life.

Chapter 3 and 4 focus on one particular aspect of SES: education. The third chapter aims to assess the role of demographic characteristics which correlate with education over time. Changes in demographic factors, such as marital status within different educational groups, could contribute to educational differentials in health. Therefore, one mechanism through which SES could shape health outcomes is explored, marital status composition within groups. The main idea is that socioeconomic inequalities in health are strongly connected to the demographic structure of different groups. Rising educational inequality, especially in the United States, calls for in-depth analyses about its origin and development over time. This chapter provides a perspective on how changes in marital composition of different educational groups affect health inequality over time in the United States.

Finally, Chapter 4 attempts to separate association from causation in the relationship between education and an increasingly prevalent morbidity, cancer. Population ageing is linked to a higher incidence and mortality from cancer, with socioeconomic status being especially associated with higher survival after cancer diagnosis. The main contribution of this work is the test of the causal effect of education on cancer survival for the first time. Moreover, it provides a comprehensive overview regarding the impact of one additional year of compulsory schooling on different cancer outcomes in England and Wales. The causal effect was identified by using two quasi-natural experiments which raised the compulsory schooling leaving age in 1947 and 1972. Therefore, the exploration of the causal link between education and cancer outcomes integrates previous work which focussed only on associations.

The research questions were developed to address three different debates. First, the socioeconomic gradient in health varies during the life-course, but it is unclear whether there is a disappearing association in old age. The first empirical work addresses this aspect by investigating the SES gradient in functional limitations in the final years of life. Second, this thesis calls for greater attention to the measurement of SES, in particular, various socioeconomic indicators have different implications and are related to several demographic characteristics. Third, this thesis emphasises how family demography can be connected to health inequalities over time. The changes over time in SES differences in health need to be explained, also examining the family structure within different educational groups.

1.2 Socioeconomic status and health in old age: a disappearing association?

Socioeconomic status is an important factor for an individuals' health. Link and Phelan (1995) point out the extensive impact of socioeconomic conditions on health. In fact, socioeconomic stratification influences access to resources and has an abiding link with disease development. More advantaged socioeconomic groups on average benefit from better health conditions and longer life expectancy. The strength of this positive association varies during the life-course, but there is no consensus regarding the description of the direction of changes (Chen, Yang, & Liu, 2010). In particular, the literature describes the varying association between socioeconomic status and health in old age in terms of divergence (Ross & Wu, 1996), stability (Clark & Maddox, 1992), and divergence until middle adulthood and then convergence (Beckett, 2000; House, Lantz, & Herd, 2005).

The divergence observed in empirical studies is theoretically conceptualised as the manifestation of *cumulative dis/advantage*, meaning the increasing gap in health over the age for different socioeconomic groups, starting from a modest discrepancy in previous periods of life (Dannefer, 2003). Several reasons can be considered as theoretical explanations for the convergence in health outcomes in old age (Robert & House, 1994), but despite the extensive developments in the study of the topic, there is no consensus concerning the role of socioeconomic status in later life health from a theoretical as well as empirical point of view. For example, different conclusions have been reached using cross-sectional and longitudinal data. In particular, some studies on longitudinal data support *cumulative disadvantage* theory (Lynch, 2003; Willson, Shuey, & Elder, 2007), whereas most of those based on cross-sectional data reach other conclusions, such as stability or convergence in the social gradient over age (House et al., 1994; Schöllgen, Huxhold, & Tesch-Römer, 2010).

The absence of consensus in the literature concerning theoretical explanation and empirical evidence motivated the choice of the first research question. Therefore, the purpose of the first study was to investigate whether socioeconomic differences in disability and functional limitations exist at the end of life contributing to the SES gap in health. Likewise, SES was measured using different indicators, because the intersection of several conditions often represents the best description of the reality.

Today, in the context of an increasing share of an older population, it is essential to study how socioeconomic status and health are linked in the last years of life. People aged 65 and older are expected to account for 16% of the total population by 2050, about three times the number observed in 2010 (WHO, 2011), which will lead to an increasing pressure on healthcare and social security systems to ensure adequate care for this growing portion of the population. Therefore, it is important to study the process of ageing in various strata of the society to better understand the effect of socioeconomic status in the later stage of life. Policy responses depend on the explanations for health inequalities (Bartley, 2004), so understanding inequalities in different strata might be informative for policy makers, helping them to design effective policies targeted to specific segments of the population.

1.3 Theoretical frameworks on the role of socioeconomic status for health inequalities

Numerous theoretical frameworks have been proposed to explain socioeconomic differences in health and mortality, which provide insight on several various aspects of these inequalities. For the purpose of this thesis, three different perspectives are discussed since they can be helpful interpretative lenses for a better understanding of health inequalities in the population. The three main approaches in the literature are the *fundamental cause*, *social determinants of health* and *life-course* perspective. However, it is important not to take these perspectives as exclusive explanations for SES differences, since there are many more theories that can be used (Mackenbach, 2012).

The *fundamental cause* approach is a useful interpretative theory for understanding the mechanisms of endurance of health inequalities over time. Link and Phelan (1995) argued that socioeconomic disparities persist, even when the generating mechanisms change because SES provides a series of resources which enable individuals to enjoy better health. The *social determinants of health* perspective suggests the importance of social circumstances in affecting individuals' health and advocates the primary role of policy makers in mitigating these inequalities within and across countries. Moreover, it highlights that social circumstances can affect health through multiple concurrent pathways: material, behavioural and psychosocial (Marmot, 2005a, 2005b). Finally, the *life-course* perspective emphasises the role of different life stages for the development of health inequalities. In particular, an important focus of this research is on how conditions early in life can contribute to socioeconomic differentials later in life (Davey Smith & Lynch, 2004; Potischman, Troisi, & Vatten, 2004).

Despite the existence of an important link between SES and health, variation is present depending on the socioeconomic indicator used. Therefore, it is important to select the appropriate socioeconomic indicator depending on life-course phase analysed and the type of outcomes. Previous studies have shown the challenges of measuring SES, especially in old age (Davey Smith et al., 1998; Geyer, Hemström, Peter, & Vågerö, 2006; Grundy & Holt, 2001). Definitive answers are not reached, but the socio-historical context deserves particular attention when examining different socioeconomic indicators.

Therefore, in this thesis, a combination of different socioeconomic indicators were used to study health inequalities over the life-course and over time. The second chapter focuses on three different SES indicators with the aim of giving an overall picture of the SES inequalities in functional limitations at the end of life. Wealth emerged as a strong predictor of socioeconomic differences in disability trajectories at the end of life. It is important to reflect on why wealth is related to later-life functioning. One possible hypothesis is that wealth captures to a better extent the exposure to positive circumstances over the life-course and can be regarded as a measure of the *cumulative advantage*. However, these results might not be generalisable to different historical contexts since the relevance of other SES indicators, such as education, has increased over time.

The third and fourth chapters focus on education as a measure of SES. Educational expansion in the last century has allowed many more people to access higher levels of education, so education has emerged as increasingly relevant marker of socioeconomic conditions. Despite the increasing access to education, educational inequalities in health and mortality are increasing, especially in the United States (Montez & Zajacova, 2013), so it is crucial to understand what the drivers of these changes in the relationship between education and health. Moreover, policy interventions targeted to increase the level of schooling in the society need to be evaluated to better inform policy makers on how to improve population health.

1.4 Health inequalities: the role of family

Demographic behaviours are tightly linked to socioeconomic circumstances. However, in the studies of health inequalities, family demography does not appear to be completely integrated. The *diverging destinies* hypothesis suggested by McLanahan

(2004) focussed on the consequences of changes in family structure for children. However, the consequences of different family arrangements for health in older adults has been examined to a much lower extent. For this reason, this thesis examined whether some of the demographic behaviours connected to the second demographic transition can be associated with trends in population health.

Education is increasingly connected to changing mechanisms of assortative mating, marriage and family stability (Härkönen & Dronkers, 2006; Kalmijn, 2013; Schwartz, 2013). Rising divorce and declining marriage rates as well as increasing cohabitation have changed the family landscape (Sweeney, 2002). These changes call for increasing attention to the relationship between education and marital status when looking at health inequalities (Kravdal, Grundy, & Keenan, 2018). More educated individuals have over time an increasing prevalence of marriage as well as higher marital stability (Goldstein & Kenney, 2001), but few studies have been able to connect this with the increasing inequalities in health and mortality by educational group.

Education and health are also strongly associated (Elo & Preston, 1996; Goesling, 2007; Mirowsky & Ross, 2003). Nonetheless, there is no definitive answer about the causal effect of education on health (Galama, Lleras-Muney, & van Kippersluis, 2018; Montez & Friedman, 2015). Moreover, in the United States, there is an increasing gap in health and mortality between individuals at the top and the bottom of the educational distribution (Meara, Richards, & Cutler, 2008; Montez & Zajacova, 2013). The changing relationship between education and health over time calls for a greater attention on the mechanisms through which this association endures. Marriage prevalence and marital stability could represent a new emerging channel for the endurance of education as *fundamental cause* of health inequalities (Link & Phelan, 1995).

Consistent relationships between marital status and mortality/health have been found in the literature (Grundy & Tomassini, 2010; Manzoli, Villari, M Pirone, & Boccia, 2007; Murphy, Grundy, & Kalogirou, 2007). However, it is difficult to distinguish between causal and selection effect (Lillard & Panis, 1996; Murray, 2000; Waldron, Hughes, & Brooks, 1996), particularly to gather causal evidence of a protective effect of marital status for health despite the presence of a clear association (Carr & Springer, 2010). Recent evidence has also became more critical about whether gains from entering marital unions exist or simply if divorce leads to worse health conditions (Kalmijn, 2017).

1.5 Data and methods

Various sources of data were used in this thesis. The English Longitudinal Study of Ageing (ELSA) was used in chapter 2 to provide a representative sample of the English population aged 50 and above. The overall length of the study is 12 years, with an increasing potential for longitudinal analysis and causal pathways in old age. This data contains measures concerning different dimensions of life in old age, including health (also, biomarkers are introduced in more recent waves) and disability, economic conditions, retirement information, social networks, civic and cultural participation and well-being.

Chapter 3 utilises repeated cross-sectional data from the National Health Interview Survey (NHIS), which collected information concerning health and psychosocial indicators, demographic and socioeconomic characteristics, access to health services and measurement of biometric characteristics, such as height, weight, and blood pressure since 1963. Other specific topics are included in the survey at regular intervals to track changes over time. NHIS provides an overview of health conditions for people living in United States for every year. For the present analysis the sample is restricted to individuals aged between 30 and 70 years old.

Finally, the data source used in chapter 4 was the Office of National Statistics Longitudinal Study (ONS LS) for England and Wales. ONS LS contains census linked information concerning vital events for 1% of the population of England and Wales from 1971 to 2011. Events registered are births, deaths, immigrations, emigrations, and cancer. Data are also linked to death records which allows examination of cancer survival. There are around 500,000 sample members who were followed longitudinally across different censuses (1971–2011), of which, 135,000 experienced a cancer event and there is information concerning their diagnosis and disease survival.

Moreover, the three empirical chapters use a variety of advanced quantitative methods. First, repeated measures latent class analysis was used to model trajectories of functional limitation in the last years of life. Through latent class analysis, the most common disability-related pathways leading to death are selected and analyzed (Linzer & Lewis, 2011; Magidson & Vermunt, 2004). Latent class analysis allows the modelling of the pattern of disability in later life without imposing a functional form to the data.

The non-linear decomposition method (Fairlie, 1999) was used to calculate the proportion of the educational gap in health attributable to marital status across cohorts. Decomposition methods are extremely useful to break down different components of an observed phenomenon. Therefore, they are particularly helpful for connecting micro-level dynamics to macro-level patterns. In chapter 3, Fairlie's decomposition allowed for a better understanding of raising educational inequalities in health by taking into account the changing marital composition.

Finally, chapter 4 employed a regression discontinuity design with categorical outcomes (Xu, 2017) to analyse the causal effect of education on cancer outcomes. It represents the first empirical application of the method which allows for the joint estimation of the treatment effects for different categories.

1.6 Structure of the thesis

The research questions of this thesis are: *How does the relationship between health and socioeconomic status change over the life-course and time? What are the mechanisms driving these processes?* Starting from these overarching questions, this thesis studied different aspects of the relationship between socioeconomic status and health and mortality. Chapter 2 analysed the existence of socioeconomic gradient in health at the end of life. Chapter 3 investigated the changing gap among different socioeconomic groups over time, focusing on the intersection of socioeconomic status with marital status. Chapter 4 examined whether there is an emergence of new educational inequalities in health, analysing the causal effect of education on cancer incidence, survival and mortality.

Chapter 2 studied the socioeconomic differential in health at the individual-level, contributing to the empirical evidence concerning the existence of a disappearing association between socioeconomic status and health in old age. It examined whether socioeconomic status is associated with disability-related trajectories at the end of life. Chapter 3 decomposed the gap in self-reported health among different educational categories over cohorts, with particular focus on the analysis of marital status composition of different socioeconomic groups over time. The demographic structure of these groups can help to explain the widening socioeconomic gap over time. Finally, chapter 4 analysed whether, and to what extent, education has a causal effect on cancer outcomes in England and Wales using a quasi-natural experiment.

In conclusion, this thesis has contributed the literature in at least three ways. First, it calls for a greater integration of the life-course analysis in the study of the relationship between SES and health. End-of-life health pathways are long being disregarded in favour of age-standardised health or disability trajectories. The final part of life can be conceptualised as a life stage which could differ among advantaged and disadvantaged socioeconomic groups. Second, it highlights the importance of the role of demographic factors when studying health inequalities over time, and in particular to consider the association of SES with other family processes. The analysis revealed an increasingly relevant role of marital status for educational inequalities in health, despite still having a modest contribution on the health differentials. Finally, this thesis provides weak evidence for a causal effect of an additional year of compulsory education on cancer incidence and survival in England and Wales. Previous literature has suggested that lower educational attainment is connected with higher survival, but no evidence was offered so far in support of or against a causal effect.

References

Bartley, M. (2004). *Health Inequality: An Introduction to Theories, Concepts and Methods*. Cambridge: Polity Press.

Beckett, M. (2000). Converging health inequalities in later life–an artifact of mortality selection? *Journal of Health and Social Behavior*, *41*(1), 106–119.

Carr, D., & Springer, K. W. (2010). Advances in families and health research in the 21st Century. *Journal of Marriage and Family*, *72*(3), 743–761.

Chen, F., Yang, Y., & Liu, G. (2010). Social change and socioeconomic disparities in health over the life course in China: a cohort analysis. *American Sociological Review*, *75*(1), 126–150.

Clark, D. O., & Maddox, G. L. (1992). Racial and social correlates of age-related changes in functioning. *Journal of Gerontology*, *47*(5), S222–S232.

Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: cross-fertilizing age and social science theory. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *58*(6), S327–S337.

Davey Smith, G., Hart, C., Hole, D., MacKinnon, P., Gillis, C., Watt, G., … Hawthorne, V. (1998). Education and occupational social class: which is the more important predictor of mortality risk? *Journal of Epidemiology and Community Health*, *52*(3), 153–160.

Davey Smith, G., & Lynch, J. (2004). Life course approaches to socioeconomic differentials in health. In D. Kuh, Y. Ben Shlomo, & S. Ezra (Eds.), *A Life Course Approach to Chronic Disease Epidemiology* (Vol. 6, pp. 77–115). Oxford University Press.

Elo, I. T., & Preston, S. H. (1996). Educational differentials in mortality: United States, 1979-85. *Social Science Medicine*, *42*(1), 47–57.

Fairlie, R. W. (1999). The absence of the african-american owned business: an analysis of the dynamics of self-employment. *Journal of Labor Economics*, *17*(1), 80–108.

Galama, T. J., Lleras-Muney, A., & van Kippersluis, H. (2018). *The effect of education on health and mortality: a review of experimental and quasi-experimental evidence* (NBER Working Paper 24225). Cambridge, MA: National Bureau of Economic Research.

Geyer, S., Hemström, O., Peter, R., & Vågerö, D. (2006). Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. *Journal of Epidemiology and Community Health*, *60*(9), 804–810.

Goesling, B. (2007). The rising significance of education for health? *Social Forces*, *85*(4), 1621–1644.

Goldstein, J. R., & Kenney, C. T. (2001). Marriage delayed or marriage forgone? New cohort forecasts of first marriage for U.S. women. *American Sociological Review*, *66*(4), 506–519.

Grundy, E., & Holt, G. (2001). The socioeconomic status of older adults: how should we measure it in studies of health inequalities? *Journal of Epidemiology and Community Health*, *55*, 895–904.

Grundy, E., & Tomassini, C. (2010). Marital history, health and mortality among older men and women in England and Wales. *BMC Public Health*, *10*, 554.

Härkönen, J., & Dronkers, J. (2006). Stability and change in the educational gradient of divorce. A comparison of seventeen countries. *European Sociological Review*, *22*(5), 501–517.

House, J. S., Lantz, P. M., & Herd, P. (2005). Continuity and change in the social stratification of aging and health over the life course: evidence from a nationally representative longitudinal study from 1986 to 2001/2002 (Americans' Changing Lives Study). *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(Special Issue 2), S15–S26.

House, J. S., Lepkowski, J. M., Kinney, A. M., Mero, R. P., Kessler, R. C., & Herzog, A. R. (1994). The social stratification of aging and health. *Journal of Health and Social Behavior*, *35*(3), 213.

Kalmijn, M. (2013). The educational gradient in marriage: a comparison of 25 European countries. *Demography*, *50*(4), 1499–1520.

Kalmijn, M. (2017). The ambiguous link between marriage and health: a dynamic reanalysis of loss and gain effects. *Social Forces*, *95*(4), 1607–1636.

Kravdal, Ø., Grundy, E., & Keenan, K. (2018). The increasing mortality advantage of the married: the role played by education. *Demographic Research*, *38*(1), 471–512.

Lillard, L. A., & Panis, C. W. A. (1996). Marital status and mortality: the role of health. *Demography*, *33*(3), 313–327.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*, *35*(Extra Issue: Forty Years of Medical Sociology: The State of the Art and Directions for the Future), 80–94.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: an R package for polytomous variable latent class analysis. *Journal of Statistical Software*, *42*(10), 1–29.

Lynch, S. M. (2003). Cohort and life-course patterns in the relationship between education and health: a hierarchical approach. *Demography*, *40*(2), 309–331.

Mackenbach, J. P. (2012). The persistence of health inequalities in modern welfare states: the explanation of a paradox. *Social Science and Medicine*, *75*(4), 761–

769.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 176–200).

Manzoli, L., Villari, P., M Pirone, G., & Boccia, A. (2007). Marital status and mortality in the elderly: a systematic review and meta-analysis. *Social Science & Medicine*, *64*(1), 77–94.

Marmot, M. G. (2005a). Social determinants of health inequalities. *Lancet*, *365*(9464), 1099–1104.

Marmot, M. G. (2005b). *Status Syndrome: How Your Social Standing Directly Affects Your Health*. London: Bloomsbury.

McLanahan, S. (2004). Diverging destinies: how children are faring under the second demographic transition. *Demography, 41*(4), 607–627.

Meara, E. R., Richards, S., & Cutler, D. M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981-2000. *Health Affairs*, *27*(2), 350–360.

Mirowsky, J., & Ross, C. E. (2003). *Education, Social Status and Health*. New York, NY, USA: Aldine de Gruyter.

Montez, J. K., & Friedman, E. M. (2015). Educational attainment and adult health: under what conditions is the association causal? *Social Science and Medicine, 127*, 1–7.

Montez, J. K., & Zajacova, A. (2013). Explaining the widening education gap in mortality among U.S. white women. *Journal of Health and Social Behavior, 54*(2), 166–182.

Murphy, M., Grundy, E., & Kalogirou, S. (2007). The increase in marital status differences in mortality up to the oldest age in seven European countries, 1990-

99. *Population Studies*, *61*(3), 287–298.

Murray, J. E. (2000). Marital protection and marital selection: evidence from a historical-prospective sample of American men. *Demography*, *37*(4), 511–521.

Potischman, N., Troisi, R., & Vatten, L. (2004). A life course approach to cancer epidemiology. In D. Kuh, Y. Ben Shlomo, & S. Ezra (Eds.), *A Life Course Approach to Chronic Disease Epidemiology* (pp. 260–280). Oxford University Press.

Robert, S. A., & House, J. S. (1994). Social structures, quality of life, and aging. In R. Abeles, H. Gift, & M. G. Ory (Eds.), *Aging and Quality of Life* (pp. 253–274). New York: Springer Publishing Company.

Ross, C. E., & Wu, C.-L. (1996). Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior*, *37*(1), 104–120.

Schöllgen, I., Huxhold, O., & Tesch-Römer, C. (2010). Socioeconomic status and health in the second half of life: findings from the German Ageing Survey. *European Journal of Ageing*, *7*(1), 17–28.

Schwartz, C. R. (2013). Trends and variation in assortative mating: causes and consequences. *Annual Review of Sociology*, *39*(1), 451–470.

Sweeney, M. M. (2002). Two decades of family change: the shifting economic foundations of marriage. *American Sociological Review*, *67*(1), 132–147.

Waldron, I., Hughes, M. E., & Brooks, T. L. (1996). Marriage protection and marriage selection–Prospective evidence for reciprocal effects of marital status and health. *Social Science and Medicine*, *43*(1), 113–123.

WHO. (2011). *Global Health and Aging*. *NIH Publication no 117737*. Retrieved from http://links.jstor.org/sici?sici=0095-9006(196024)1:4%3C273:HAA%3E2.0.CO;2-C

Willson, A. E., Shuey, K. M., & Elder, G. H. (2007). Cumulative advantage processes as mechanisms of inequality in life course health. *American Journal of Sociology*, *112*(6), 1886–1924.

Xu, K. L. (2017). Regression discontinuity with categorical outcomes. *Journal of Econometrics*, *201*(1), 1–18.

Chapter 2

Disability pathways preceding death in England by socioeconomic status

The paper is in collaboration with Christiaan Monden and is published in *Population Studies*.

Abstract

The role of socioeconomic status (SES) in the last years of life is an under-researched aspect of health inequalities. This study examines disability patterns preceding death using data from the English Longitudinal Study of Ageing. We use repeated measures latent class analysis to identify the most common pathways preceding death in terms of walking ability and limitations in activities of daily living. Three pathways emerge: one characterized by consistently low disability; a second by a constant high level of functional limitations; and a third by medium impairment. We examine how different SES indicators predict belonging to each disability pathway. Conditional on income, higher wealth is associated with a lower likelihood of belonging to the high disability pathway. Contrary to our expectations, we find no educational gradient in the pathways preceding death. Health inequalities in the last years of life seem to exist especially between individuals with different levels of wealth.

Keywords: ageing; death; old age; socioeconomic status; latent class; disability; time to death

2.1 Introduction

While the general association between socioeconomic status and mortality/morbidity is well-established—people with higher education, greater wealth, and higher incomes live longer and in better health—much remains to be understood about the role of socioeconomic status (SES) across the life-course. The importance of SES for health in the last years of life is one area where much is still to be learned. Most analyses of socioeconomic differences in health during later life are designed around chronological age; that is, they follow people in certain age groups regardless of their eventual age at death. Such designs can inform us about disability patterns among people aged in their 70s or 80s, for example. However, we lack a detailed understanding of actual disability patterns in the last few years before death. What health pathways are most common in the last years of life? Moreover, we know little about the socioeconomic differences in what we refer to as *pathways to death*.

Pathways to death are relevant in at least three ways. First, they have intrinsic meaning for individuals and their loved ones. Second, they are important from a healthcare policy perspective, insofar as the last years of life are the most expensive in terms of inpatient and social care costs (Zweifel, Felder, & Meiers, 1999; Cutler & Meara, 2001; Bardsley, Georghiou, & Dixon, 2010). Third, we can use these pathways to further our understanding of theories about socioeconomic variation in health and disability across the life-course.

Our study has two main goals. First, we explore the disability pathways leading up to death for older adults in England. We follow respondents from around eight years before they die up until their death, and map their pathways in terms of ability to walk one-quarter of a mile and severity of difficulties in activities of daily living. The second aim is to understand whether, and to what extent, these pathways are shaped by socioeconomic factors. We use data from the first six waves of the English Longitudinal Study of Ageing (ELSA). The analysis proceeds in two steps. First, we use repeated measures latent class analysis to identify the different impairment pathways preceding death. Then we study how different socioeconomic indicators help to predict the probability of experiencing a certain disability pathway.

Our study makes three contributions to the literature on SES and health in later life. The first is our design to study disability patterns in the last years of life. Several methods have been suggested for dealing with non-random attrition due to death and differential survivorship by health status in conventional longitudinal studies of ageing; these include imputation of expected values, joint models, and exclusion of those who die (Diehr & Patrick, 2003; Rajan & Leurgans, 2010; Murphy et al., 2011). Our design represents a different way to examine the confounding effect of death in later-life trajectories by focusing only on individuals who die and studying disability patterns in their last years of life.

Our second contribution lies in assessing the explanatory power of traditional SES indicators with respect to disability patterns in the last years of life. SES might play a critical role in shaping disability pathways in the last years of life, but few studies have tried to examine this particular socioeconomic gradient in the years before death. Moreover, previous studies have suggested that SES differences in health and disability widen over the life course (Ross & Wu, 1996; Power, Matthews, & Manor, 1998), while other studies have suggested that age narrows health disparities (Beckett, 2000; Dupre, 2007). In particular, some studies have emphasized that selective mortality may be responsible for the decreasing association between SES and health in old age (Willson, Shuey, & Elder, 2007). By examining SES differences in the last years of life, we contribute to the empirical investigation of the changing association between SES and health in old age. We also take into account heterogeneity in socioeconomic indicators. It has long been acknowledged that several mechanisms link the various measures of socioeconomic position to health and mortality to different extents (Robert & House, 1996; Torssander & Erikson, 2010; Cutler, Lleras-Muney, & Vogl, 2011).

Third, this study contributes to a greater integration of life course analysis in the study of the relationship between SES and health. Health-related pathways in the last years of life have been disregarded in favour of age-standardized trajectories in the literature. The final part of life can be conceptualized as a life stage that could differ among advantaged and disadvantaged socioeconomic groups. In the study of ageing, one of the emerging questions is how chronological age is useful as a life stage marker (Morgan & Kunkel, 2016). Lowsky et al. (2014, p. 640) have noted: 'For a surprisingly large segment of the older population, chronological age is not a relevant marker for understanding, measuring, or experiencing healthy aging'. Biological age and life stages have emerged as potential ways forward for going beyond the limits of considering chronological age as the only life stage indicator (Simm et al., 2008). Simultaneously, in public health and gerontology literature, there has been an increasing interest in time to death as an indicator of ageing (Freedman, Ondrich, Seplaki, & Spillman, 2015). Therefore, our study conceptualizes the final years of life in terms of a life course stage and analyses how SES is related to this phase of life.

2.1.1 Trajectories in the last years of life

Glaser and Strauss (1968) were the first to draw attention to the longitudinal paths at the end of life. Although they took the perspective of caregivers and medical managers, their work showed an early interest in disability patterns at the end of life. Most studies of health trajectories in later life so far have focused on age- and cohort-related health patterns (Lynch, 2003; Zimmer, Martin, Nagin, & Jones, 2012; Rohlfsen & Kronenfeld, 2014). Death has often only been considered in terms of attrition bias (Dodge, Du, Saxton, & Ganguli, 2006; Liang et al., 2010; Taylor, 2010).

There has been more interest in longitudinal pathways to death in studies on cognitive functioning, where the debate is whether these trajectories at the end of life exhibit a steady and gradual decline or a more abrupt curvilinear drop (Wilson et al., 2003; MacDonald, Hultsch, & Dixon, 2011; Gerstorf & Ram, 2013). Only a small number of studies have explicitly examined physical health trajectories preceding death (Bradley, Fried, Kasl, & Idler, 2000; Schoeni, Freedman, & Wallace, 2003; Gill, Gahbauer, Han, & Allore, 2010). Gerstorf and colleagues (2013) compared the age and time-to-death perspectives for changes in six health domains (cognitive, sensory, physical, health, social, and perceived control). From their study it emerged that a mortality-related perspective described changes better than an age-based scale. They showed that less advantaged socio-demographic circumstances were correlated with worse health conditions at the baseline, but they did not find SES to be associated with differential rates of decline. However, their analysis did not differentiate between socioeconomic indicators (see for instance Herd, Goesling, & House, 2007); instead they used a composite index. Liao et al. (1999) explored educational and racial differentials in morbidity and disability in the last years of life using a United States (US) sample for 1986–90. They found that more educated individuals were better off in terms of activity limitations, chronic conditions, bed days, and hospital days in the two years before death. These findings strongly suggest that health trajectories leading up to death vary by educational level. Other studies have examined actual health trajectories to death for various indicators of morbidity and disability, but they have not related the observed patterns to socioeconomic differences or have only examined short periods before death (Guralnik, LaCroix, Branch, Kasl, & Wallace, 1991; Lunney, Lynn, Foley, & Lipson, 2003; Gill et al., 2010; Klijs, Mackenbach, & Kunst, 2010).

Schoeni and colleagues (2003) explored socioeconomic differences in disability status before death across education and poverty groups in the US. However, due to the limited availability of longitudinal data they used a synthetic approach to estimation, using repeated cross-sectional data from the National Health Interview Survey. Therefore, they were not able to highlight actual individual trajectories preceding death, which ELSA allows us to do. They concluded that socioeconomic differences in disability present before death fade in proximity to death.

2.1.2 Socioeconomic status and health in later life

Previous research has analysed mostly health conditions and physical barriers as determinants of the onset and progression of disability trajectories in old age (Stuck et al., 1999; Fried, Young, Rubin, & Bandeen-Roche, 2001). Only a few studies have looked at the associations between socioeconomic factors and health-related trajectories (House, Lantz, & Herd, 2005; Herd et al., 2007), and very limited longitudinal research is available on the socioeconomic gradient in the last years of life.

Previous research in the US has found that education is associated with the onset of health problems, while income is related to their progression rate (House et al. 2005; Herd et al. 2007). Other SES indicators such as wealth and occupation have been studied less frequently. Nevertheless, Robert and House (1994) showed how liquid assets and homeownership predict differences in functional health better than both income and education.

We study the association between SES and disability patterns in the last years of life for three different socioeconomic indicators: education, income, and wealth. The patterns of impairment in the eight years before death reflect the life course circumstances that have shaped the history of each individual as well as how well individuals have coped with increasing impairment during the last years. Different socioeconomic indicators might affect patterns in different ways. Given the explorative nature of our study and the stepwise procedure for identifying patterns, it does not seem appropriate to formulate specific hypotheses for the socioeconomic factors; instead, we discuss possible mechanisms in the rest of this section.

Education has a well-documented positive impact on status attainment, which is beneficial for health through material and behavioural factors. Education may be both an important factor for accumulation of health advantage over the life course and a positive moderating element in the presence of health problems. Education seems to increase the ability to manage illness (Goldman & Lakdawalla, 2001) and take-up of new medical technologies is greater among highly educated people, which might help to mitigate unfavourable health conditions at the end of life (Lleras-Muney & Lichenberg, 2002).

Although income and wealth might seem to be very similar resources, their functions are heterogeneous and they assume different meanings, especially in the context of an older population (Wise, 1998; Poterba, Venti, & Wise, 2011, 2013). Income primarily supports daily expenditures and permits a certain standard of living. We can view income as reflecting the short-term and current component of a person's financial resources. Wealth, on the other hand, reflects the accumulation of advantages over the life course and it also assumes functions other than purchasing or expenditure. It can provide an insurance function for unexpected unfavourable events, such as health shocks or unforeseen economic hardships. Moreover, for older individuals, wealth can also have a bequest or inter vivos transfer purpose: individuals might want to accumulate financial resources in order to transfer them to their loved ones later on. On the one hand, income could represent a resource for management of disability, which could allow the purchase of rehabilitative programmes after the onset of impairment. Therefore, income might relate to the development of the pathway of disability. On the other hand, assets could reflect advantage over the life course, which might affect the level of disability (i.e., severity of impairment) in the last years of life.

2.2 Data and methods

2.2.1 Research design and data

We use data from Waves 0–5 of ELSA, carried out between 1998 and 2011 (Marmot et al., 2011). ELSA respondents were originally drawn from households participating in the cross-sectional Health Survey for England. This initial interview was called Wave 0, with the subsequent longitudinal waves being Waves 1–5. Our sample includes only respondents who died within the two years after either Wave 4 or Wave 5 and for whom information about the previous four waves is available. Therefore, the first subsample includes longitudinal information from Waves 1–4 and the second from Waves 2–5 (detailed construction is included in Panels B and C in Table 1). Mortality follow-up is obtained through the linkage of death records from the National Health Service Central Register Database. Since more than 95 per cent of the sample agreed to the linkage and the database contains high-quality data, the reliability of the deaths reported is very high (Banks et al., 2010). Time is coded as distance to death (t-1 stands for one period before death, t-2 for two periods before death, and so on). Consecutive waves are approximately two years apart. Our sample thus allows us to visualize patterns across four periods, that is, the last eight years before individuals' deaths.

This sample includes only individuals aged 65 and older at the time one period before death (t-1). Mean age at baseline (i.e., four periods before death) is 75 and 77 for men and women, respectively. There are 836 respondents who died in the two years after Wave 4 or Wave 5 for whom we have information at baseline. However, panel attrition means that many of them did not participate in all four waves. Table 1 presents the sample size achieved by including only those present in all four periods preceding death (444; that is 276 who died after Wave 4 and 168 who died after Wave 5). Our final analytic sample (without missing data for any of the socioeconomic variables) comprises 437 individuals of whom 214 (about 49 per cent) are male (Table 2). From Table 1 we observe that the groups of individuals who died (Panels B and C) do not differ from the overall sample (Panel A) in terms of retention rate. The retention rates between waves are slightly higher in the overall sample than among respondents who died (e.g., from Wave 3 to Wave 4, 83.6 per cent of individuals remain in the sample, but among those dying after Waves 4 or 5, the figure is only 78.0 per cent). Previous analyses of ELSA data showed that people lost to follow-up between Waves 1 and 5 were older, with higher prevalence of limiting long-standing illness, lower wealth, and lower educational attainment (Steptoe, Breeze, Banks, & Nazroo, 2013). Similarly, in this analysis, we can look at the characteristics of those lost at follow-up from period t–4 to period t–1 only among those who died. Table 2 shows that the decedents for whom we have information for all four periods are wealthier, more educated, and less likely to be married four periods before death than those leaving the sample between waves. No significant differences are observed in terms of disability four periods before death.

Therefore, our final sample of decedents may be slightly selected in terms of advantaged socioeconomic position—as was already shown for the overall longitudinal ELSA sample (Steptoe et al., 2013)—but not in terms of disability.

Table 1 Sample Size in Different Waves and Retention Rates in ELSA Waves 1–5

	Wave 1 (2002/03)	Wave 2 (2004/05)	Wave 3 (2006/07)	Wave 4 (2008/09)	Wave 5 (2010/11)
Panel A: whole sample					
Number of individuals	12,099	9,432	9,771	11,050	10,274
Number of individuals (for whom we have data on all previous waves)	12,099	9,324	7,613	6,367	5,633
Panel retention (whole sample; percentage)	–	*77.0*	*81.6*	*83.6*	*88.5*
Panel B: individuals who died after Wave 4					
Number of individuals (aged 65 and above in period t-1, for whom we have data on all previous waves)	507	418	354	276	–
Panel retention (percentage)	–	*82.4*	*84.7*	*78.0*	–
Panel C: individuals who died after Wave 5					
Number of individuals (aged 65 and above in period t-1, for whom we have data on all previous waves)	–	329	273	213	168
Panel retention (percentage)	–	–	*83.0*	*78.0*	*78.8*

Notes: Period t-1 is the time two years before death. Source: ELSA.

Table 2 Descriptive Statistics at the Baseline (Four Periods before Death) for Those who Died after Waves 4 and 5 of ELSA, Comparing Those in Our Sample (Present in all Waves) with Those Lost to Follow-up

	Present in all waves			**Lost to follow-up**		
Variable	Mean/ *percentage*	Standard deviation	Number	Mean/ *Percentage*	Standard deviation	Number
Male	*49*	–	437	*46*	–	384
Married	*44**	–	437	*51**	–	384
Age	76	7.8	437	76	7.2	384
Wealth	165,916*	184,264	437	141,673*	165,833	384
Income	195.9	135.8	437	190.5	114.0	384
Education	1.59*	0.76	437	1.49*	0.73	384
IADL	0.35	0.80	437	0.44	0.91	383
ADL	0.65	1.17	437	0.70	1.25	383
WALK	1.97	1.17	437	2.09	1.25	380

† $p<0.10$; *$p<0.05$; **$p<0.01$; ***$p<0.001$
Notes: Observations with missing data on SES are excluded. WALK = walking ability indicator, ADL = activities of daily living indicator, IADL = instrumental activities of daily living indicator.
Source: As for Table 1.

Our design helps to solve two recognized issues in ageing studies: *healthy survivor* bias and informative censoring due to death (Murphy et al., 2011). The first problem, the so-called healthy survivor effect (Murphy et al., 2011), refers to studies observing only the healthier individuals in the population at old age, because of age-truncated designs and attrition due to death (Kelley-Moore & Lin, 2012). Premature mortality, in particular, could cause selection in those observed reaching old age: for example, a cohort observed in old age could appear more robust because those with higher mortality risk have been selected out of the sample. Our study design does not suffer from this healthy survivor bias, since it includes only individuals who have died at different ages. In this way, we eliminate the bias caused by the selection happening in age-based groups (Hoffmann, 2008).

The second problem that our design avoids is the non-randomness of missing data due to death. Non-random attrition due to death could bias the study of longitudinal patterns of health and disability in old age. Missing data caused by participants' deaths are associated with the longitudinal outcomes under examination. Therefore, it is necessary to take into account the dependence between longitudinal outcomes and death. Our design considers only individuals who have died and follows their disability pathways preceding death. Studies that simply exclude those dying from the longitudinal analysis could produce biased results (Biering et al., 2015). Alternative strategies that help to deal with non-random missingness due to death, such as imputation of the expected values, transformation of the outcome variable, competing risk models, and joint models, suffer from other drawbacks (Murphy et al., 2011). First, the imputation of expected values that a person would report, if still alive, treats death as non-informative censoring and produces biased estimates. Second, incorporating death in the disability score by transforming the variable would result in an underestimation of the mean scores for the surviving population (Biering et al., 2015). Competing risk models—despite being able to deal with the informative censoring—do not prevent the healthy survivor effect and can produce biased estimates. The joint modelling of longitudinal outcome and time to death is one approach that could handle both healthy survivor bias and informative censoring. However, these models are computationally complex and the current software packages available have limited applicability to different problems (Tsiatis & Davidian, 2004; Diggle et al., 2008). By looking only at individuals who have died, we do not need to take care of the non-random attrition due to death. Our design therefore represents a different approach to the difficulties described in longitudinal studies of ageing and offers a viable solution of limited computational complexity.

2.2.2 Dependent variables

Being able to function independently is crucial for older people and represents an extremely important public health issue, because later-life functioning provides a valuable indicator of the effect of chronic diseases on quality of life and long-term care use in old age (Guralnik, Fried, & Salive, 1996). Chronic conditions, as well as the severity of diseases, greatly affect later-life functioning, making disability a valid and useful indicator of functional health. We use three different measures of the disability status of individuals. The first two are frequently used indicators of difficulties with the activities of daily living (Martin, Schoeni, & Andreski, 2010; Montez & Hayward, 2014). The Activities of Daily Living indicator (ADL) refers to the physical dimension of everyday activities. It includes six activities: walking across a room, dressing, bathing or showering, eating, getting in or out of bed, and using the toilet (including getting up and down) (Freedman, Martin, Schoeni, & Cornman, 2008). The Instrumental Activities of Daily Living indicator (IADL) refers to everyday tasks that are not necessarily physically intense but require a good level of cognition. Five questions are asked in order to measure difficulties in instrumental activities of daily living: the ability to prepare a hot meal, to shop for groceries, to make phone calls, to take medications, and to manage money, for example, paying bills and keeping track of expenses. Some studies have created a dichotomous variable for the presence of limitations (Cai & Lubitz, 2007) or used the absolute number of limitations to measure the degree of impairment (Zimmer, Liu, Hermalin, & Chuang, 1998). Yang and Lee (2009) divided the ADL score into four different levels of functional impairment. We take a similar approach and divide the ADL and IADL scores into three groups: no impairment (no limitations), moderate impairment (one limitation), and severe impairment (two or more limitations). The presence of at least one limitation is generally a useful criterion for determining disability, but this classification allows us to quantify not only the presence but also the overall severity of impairment. Our third indicator, WALK, represents the degree of difficulty respondents report with accomplishing a specific physical task: walking for one-quarter of a mile. We retain the original scoring system of the survey: *1* for *no difficulty*, *2* for *some difficulty*, *3* for *much difficulty*, and *4* for *unable to do this*.

These three indicators may highlight different degrees of impairment associated with a common underlying condition. Unlike mere walking ability, ADL and IADL difficulties are not only an assessment of physical capacity but also reflect environmental and social factors (Freedman and Martin, 1998). While ADL indicates difficulties in basic everyday activities, IADL constitutes a broader indicator of impairment that reflects environmental factors influencing the ability to live independently, as well as the perception of disability (Spillman, 2004). Moreover, difficulties with the instrumental activities of daily living are also related to cognitive functioning.

2.2.3 Socioeconomic variables

We use three indicators for SES: education, equivalized household income, and wealth. For the educational variable, information about the highest qualification attained comes from Wave 0. We assign individuals without any qualifications to the first educational category, those with secondary level qualifications or below to the second category, and those with higher education (below degree or degree level) to the third category. The individuals with foreign education (44 in total) are classified according to the age at which they finished full-time education (first educational category if they had never been in full-time education or were 14 or under on leaving; second educational category if the leaving age ranged between 15 and 17; and third educational category for those leaving at age 18 or over).

Data on income and wealth are derived from period t-4. Monetary income is reported as weekly equivalent income for the household. It is the sum of several possible sources of income: employment income, self-employment income, state benefit income, state pension income, private pension income, asset income, and other income. Wealth is measured as the net total wealth of the household, which is the sum of savings, investments, physical wealth, and housing wealth after financial and mortgage debt has been subtracted. Note that wealth can also assume a negative value if debts are larger than assets. We use the Organization for Economic Cooperation and Development (OECD) equivalence scale to take household size into account (OECD, 2013). Income and wealth are divided into quartiles based on the distributions among all living and deceased respondents in the sample within the age range of interest. All socioeconomic factors are included in the form of dummy variables, to allow for non-linear associations.

2.2.4 Methods

Our analysis proceeds in two steps. First, we apply repeated measures latent class analysis (RMLCA) to describe the most common pathways to death in our sample (Collins & Lanza, 2010) in terms of WALK, ADL, and IADL difficulties. Prior research has analysed how socioeconomic position influences the onset and progression of health problems, often by imposing a priori parametric forms to the patterns, for example, strict linearity of changes (Ferraro, 2011; Zimmer et al., 2012). However, Ferraro (2011) has pointed out that health changes tend to have an episodic and non-linear nature. Therefore, RMLCA seems to be a preferable method for this analysis, rather than latent class trajectory analysis (LCTA), because change over time is modelled without imposing any functional form (George, 2009). RMLCA is especially helpful in this case since the change over time could differ across classes: for example, if we expect one class with a smooth increase in disability, one with an abrupt increase, and one with more discontinuous changes, RMLCA will be able to identify them distinctly. By using RMLCA, the patterns of disability are modelled in such a way that they could take any form in each latent class. On the contrary, an LCTA approach would require the choice of a unique functional form for the growth in all classes.

Moreover, RMLCA models within-time and between-time changes together, while latent transition analysis (LTA) models latent classes at different points and then estimates the transition probabilities from one class to the other. Latent classes defined by RMLCA are characterized by individuals' values for each disability indicator over time, with the goal of finding the patterns that explain variation in disability across periods. On the contrary, LTA does not model one vector over time for each individual, but focuses more on the transitions between one class and the other across periods. In this study, we are interested in the patterns over time rather than state changes across consecutive periods, and this motivates the choice of the RMLCA approach.

Each disability outcome has four observed points corresponding to the periods before death, with three response categories based on severity (except for walking, where there are four categories). One of the main assumptions for the RMLCA is local independence, meaning that the four observed points are related only through the latent variable. We, therefore, assume that, conditional on latent class membership, the variables are unrelated over time. We estimate three separate models one for each disability outcome (WALK, ADL, and IADL). It is important to highlight that the latent class analysis approach is a probabilistic one, meaning that patterns are probabilistic and do not represent real entities (i.e., cases are similar to each other because their responses are generated by the same probability distribution). In order to simplify the interpretation, we apply labels to the classes identified, with the aim of characterizing the probabilistic patterns found. Latent classes are denoted as pathways or patterns in the following sections. However, each group identified cannot be considered completely homogenous. We use the Expectation–Maximization (EM) algorithm for optimization. The model was estimated 300 times and the maximum number of iterations for the estimation algorithm set to 100,000 to avoid local maxima.

The number of classes was chosen by fitting latent class models for each disability outcome, starting from one class and progressively adding more classes (Nylund, Asparouhov, & Muthén, 2007). The optimal number was determined by looking at different criteria such as the Bayesian Information Criterion (BIC) and likelihood ratio Chi-squared (Schwarz, 1978; Magidson & Vermunt, 2004). The BIC is the preferred criterion, because it takes into account the goodness of fit as well as the parsimony of the model. A less formal criterion widely used in the literature consists of looking at the change in likelihood ratio compared with the one-class model. According to this test, a sufficient number of classes is obtained when a significant change is no longer observable (Magidson & Vermunt, 2004).

In the second step, we use multinomial logistic models to investigate how socioeconomic characteristics—education, income, and wealth—are related to membership of the different pathways (i.e., to the latent classes obtained from our

RMLCA). In order to estimate the coefficient without bias, we apply the *one-step* technique in which the covariates' coefficients are estimated as part of the latent class model (Bolck, Croon, & Hagenaars, 2004). We also include the control variables age, sex, and marital status (i.e., if the individual was married four periods before death). We test the robustness of the results, differentiating those experiencing widowhood during the four periods analysed. Moreover, all SES variables are included simultaneously in the model. Similar results are obtained by including the three SES variables separately (results available on request). For all analyses we use the poLCA package in R (Linzer & Lewis, 2011).

2.3 Results

2.3.1 Pathways to death

Which disability pathways preceding death do we observe in the data? Table 3 presents model fit statistics, according to which we can decide the optimal number of classes. Following the criteria described in the Methods subsection, three latent classes (also called pathways) emerge as adequate for the WALK and IADL models. In the ADL model the optimal number of pathways identified is only two.

Table 3 Model Fit Statistics from Latent Class Analysis for Models with Different Numbers of Latent Classes, by Disability Indicator

Disability indicator	Number of latent classes	Number of parameters estimated	G^2	Df	AIC	BIC	Log-likelihood
WALK	1	12	980.9	243	4,437.1	4,486.0	-2,206.6
	2	25	396.3	230	3,878.5	3,989.3	-1,914.2
	3	38	245.7	217	3,753.9	3,908.7	-1,839.0
	4	51	190.1	204	3,724.3	3,932.1	-1,811.2

	5	64	158.8	191	3,719.0	3,980.0	-1,795.5
	6	77	136.8	178	3,723.0	4,037.0	-1,784.5
	7	90	118.1	165	3,730.4	4,097.0	-1,775.2
ADL	1	8	511.6	72	3,337.1	3,369.8	-1,660.6
	2	17	125.2	63	2,968.8	3,038.1	-1,467.4
	3	26	85.7	54	2,947.2	3,053.3	-1,447.6
	4	35	62.4	45	2,942.0	3,084.8	-1,436.0
	5	44	49.2	36	2,946.7	3,126.3	-1,429.4
	6	53	35.7	27	2,954.2	3,167.5	-1,422.6
	7	62	28.8	18	2,962.4	3,215.4	-1,419.2
IADL	1	8	575.3	72	2,974.4	3,007.0	-1,479.2
	2	17	180.2	63	2,597.2	2,666.6	-1,281.6
	3	26	102.1	54	2,537.2	2,643.2	-1,242.6
	4	35	59.8	45	2,512.9	2,655.7	-1,221.5
	5	44	42.1	36	2,513.1	2,692.7	-1,212.6
	6	53	32.1	27	2,521.1	2,737.4	-1,207.6
	7	62	24.8	18	2,531.8	2,784.8	-1,203.9

Notes: G^2 = likelihood ratio/deviance statistics, Df = Degrees of Freedom, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. WALK = walking ability indicator, ADL = activities of daily living indicator, IADL = instrumental activities of daily living indicator.
Source: Authors' calculations based on ELSA data.

The composition and prevalence of the disability pathways vary somewhat between the indicators, but three similar patterns can be identified. Figure 1 presents the probability compositions of impairment scores for each class. It visualizes these three pathways, with time to death on the x-axis and the probability of reporting different impairment scores on the y-axis. Each shading colour represents the probability of reporting a particular impairment score. Individuals belonging to the first pathway, which we denote 'Low Disability', experience a high probability of reporting no difficulty over the whole period (i.e., a high value for the probability of an impairment score of *1*). This group is characterized by low disability conditions in the four periods before death and a small, but slightly increasing, probability of moderate difficulties. A pattern of constant high disability ('High Disability') is evident in the second pathway; a high likelihood of severe impairment in all four periods before death is evident from the large probability of having two or more ADL and IADL limitations and from the inability to walk one-quarter of a mile or severe difficulty in doing so. The third pathway, 'Medium Disability', shows medium probability of impairment over the four periods (and also moderately increasing in the WALK model).

The prevalence of the first pathway, Low Disability, varies across the disability indicators. In the WALK model, only 35 per cent of the sample falls into this pathway, whereas this increases to 66 per cent for IADL. The prevalence of the High Disability pathway ranges between 27 per cent for WALK and 38 per cent for ADL. While the Medium Disability pathway is the most common pathway in the WALK model, it is the least prevalent in the IADL model: 38 per cent of the sample belong to this pathway for walking ability and only 11per cent for IADL limitations. A Medium Disability pathway does not emerge from the ADL model.

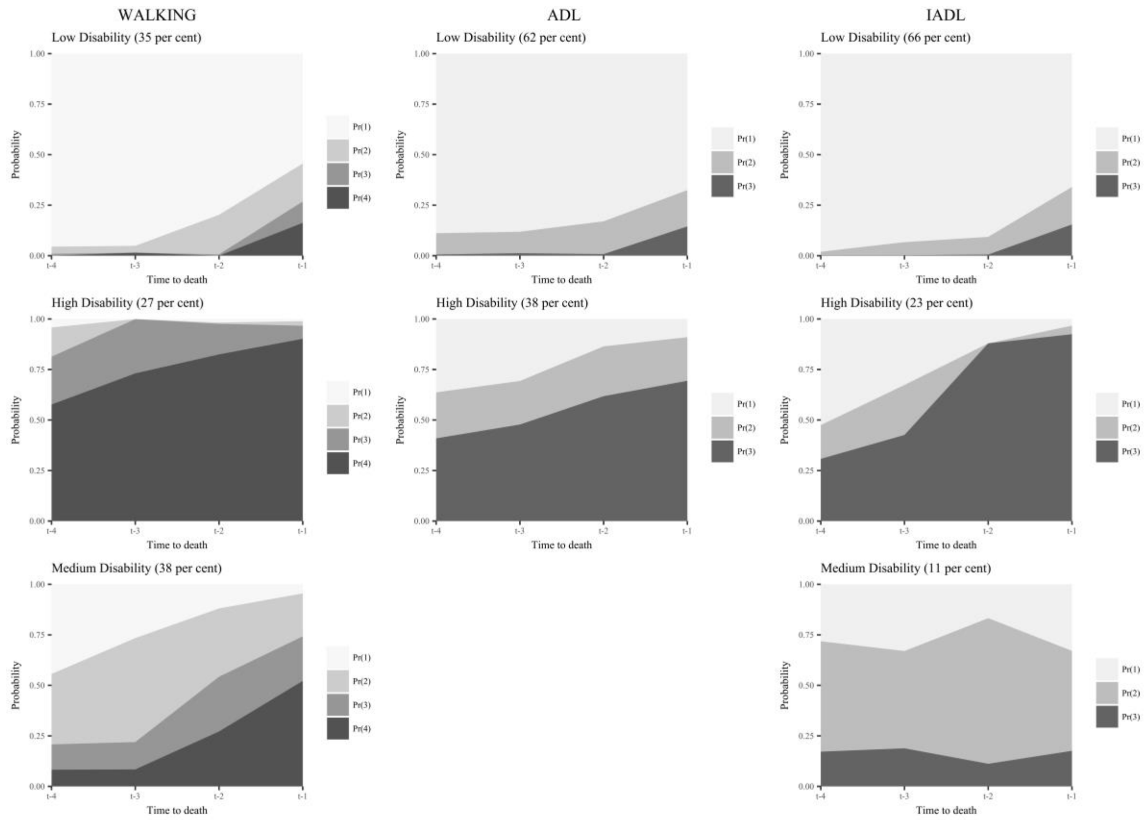

Figure 1 Probability composition of the pathways to death for WALK, ADL and IADL models.

Notes: WALK = walking ability indicator, ADL = activities of daily living indicator, IADL = instrumental activities of daily living indicator. P(1) represents the probability of an impairment score of 1 (which stands for *no difficulty*), Pr(2) the probability of an impairment score of 2 (which stands for *some difficulty* in WALK model and one limitation in ADL and IADL models), Pr(3) the probability of an impairment score of 3 (which stands for *much difficulty* in the WALK model and two or more limitations in ADL and IADL models), Pr(4) is the probability of an impairment score of 4 (which stands for *inability to do this* in the WALK model). t-1 stands for one period before death, t-2 for two periods before death and so on. Classes are derived from latent class models without covariates.

Table 4 Number of Individuals Classified in Different Pathways for Each Pair of Disa bility Indicators

		IADL		
		Low Disability	High Disability	Medium Disability
ADL	Low Disability	234	19	13
	High Disability	55	80	32
		WALK		
		Low Disability	High Disability	Medium Disability
ADL	Low Disability	141	21	104
	High Disability	13	95	59
		IADL		
		Low Disability	High Disability	Medium Disability
WALK	Low Disability	144	7	4
	High Disability	27	60	29
	Medium Disability	119	32	12

Notes: WALK = walking ability indicator, ADL = activities of daily living indicator, IADL = instrumental activities of daily living indicator. *Source:* Authors' calculations based on ELSA data.

Table 4 presents the joint distributions of class assignment for the three different models. The classification seems quite consistent across disability indicators, and especially so for the Low Disability pathways. For example, the majority of individuals in the Low Disability class in the IADL model are similarly assigned to this class in the other two models. However, a fair amount of dissimilarity among classifications exists, especially when looking at the Medium Disability pathways; this can be explained by the fact that the three disability indicators aim to measure different types of functional limitation. Therefore, the separate analysis of the three disability indicators offers a multi-faceted view of impairment status.

2.3.2 Pathways to death by socioeconomic status

How are these pathways associated with SES? Table 5 shows the relative risk ratios along with 95 per cent confidence intervals for income, wealth, and education, and for the three impairment indicators. We compare the likelihood of experiencing the High or Medium Disability pathways with the likelihood of experiencing the Low Disability one. Therefore, for all SES indicators we use the Low Disability pathway as the reference category. Relative risk ratios greater than one imply a raised probability of belonging to the High or Medium Disability pathways as opposed to the Low Disability one.

Two main observations can be made. First, contrary to our expectations, education and to some extent income do not appear to be associated with any particular pathway (for the association between pathways to death and education, we see that the coefficients are not significant in any of the models). Second, we find a wealth gradient in the pathways to death for almost all disability indicators. The relative risk ratios of the High Disability pathway for respondents in the third and fourth wealth quartiles are lower than one, indicating a lower probability of belonging to this pathway compared with the reference category (lowest wealth quartile). In other words, being in the highest wealth quartiles implies lower odds of belonging to the High Disability pathway than the Low Disability one. The decrease in the odds of the High Disability pathway ranges from 52 per cent (being in the third wealth quartile for ADL) to 80 per cent (being in the third wealth quartile for IADL). No wealth gradient is evident in the relative risk for Medium Disability with the exception of the fourth quartile of walking ability. Holding all other factors constant, the relative risk of Medium Disability does not exhibit any consistent gradient for any socioeconomic indicator.

Our results also point to the presence of non-linearity in the association between income and pathways to death in the WALK model. Surprisingly, the relative risk of experiencing the High Disability pathway, rather than the Low Disability one, is higher in the second and third income quartiles than in the lowest quartile. For example, for a person in the second income quartile, the relative risk of being in the High Disability pathway rather than the Low Disability one is 2.7 times higher for WALK (1.8 times for ADL) than that of a person in the first income quartile. Therefore, being in the second income quartile represents a disadvantage for health compared with being in the lowest one. However, this result is not evident in the IADL model. Some non-linearity is also evident in the wealth gradient. In general, the decrease in the odds of High Disability compared with Low Disability is larger in the third quartile compared with the fourth. Therefore, for all impairments, individuals in the third quartile of wealth seem to show lower odds of the most disabled pathway to death than individuals in the fourth quartile.

Our models consider the three SES indicators simultaneously. When examining these three indicators, we find that income, education, and wealth are somewhat distinct in their associations with disability at the end of life. Higher wealth is associated with a less disabled pathway to death, but higher income and educational attainment are not. Moreover, different disability indicators produce similar but not completely identical results. A very clear gradient for wealth is evident from the latent class analysis performed on IADL and WALK, but only partially for ADL. In particular, in the relative risk ratios for ADL, significant differences are only observed between the third and first quartiles of wealth. Being in the most advantageous wealth quartile does not seem to make any difference in terms of ADL, but it matters for IADL and WALK.

The control variables show the expected patterns. First, older age is consistently associated with worse disability pathways. Second, men are less likely than women to be in the High or Medium Disability pathways for all three health outcomes. This is in line with previous studies documenting a higher rate of functional limitations and disability among women during old age compared with men (Arber & Ginn, 1993; Case & Paxson, 2005). This finding demonstrates that, even in the terminal phase of life, men exhibit an advantage in terms of disability, as Lunney and colleagues (2003) have also pointed out. Unfortunately, further investigation of gender differences is not feasible because of the sample size. Moreover, the relative risks of high or medium impairment are generally smaller among married individuals than among those who are unmarried, although they are not statistically significant. This accords with the extensively documented positive association between marital status and low disability (Pienta, Hayward, & Jenkins, 2000; Liu & Zhang, 2013). However, there seems to be an exception in the IADL model, where being married four periods before death is positively associated with the High Disability pathway. Since the instrumental activities of daily living represent more practical abilities, we would expect married individuals to be able to rely on the other partner for these activities more than those remaining unmarried. Therefore, as a robustness check, we included an indicator of whether individuals experienced widowhood during the three periods before death. The experience of widowhood can have adverse health consequences for the surviving partner and even affect their mortality (Williams & Umberson, 2004; Elwert & Christakis, 2008). We found that experiencing widowhood was not associated with any pathway to death, and the results remained unchanged when we included this information in the model (results available on request). Similarly, we checked the robustness of the results by including in the analysis an indicator of whether the person was living with a child and the number of people in the household. Results also remained stable with the inclusion of these variables.

Table 5 Relative Risk Ratios for Belonging to the High or Medium Disability Pathways Relative to the Low Disability Pathway, for Walking Ability, ADL, And IADL, by SES

	WALK		ADL	IADL	
	High vs. low disability	Medium vs. low disability	High vs. low disability	High vs. low disability	Medium vs. low disability
Income **(ref. = 1, lowest quartile)**					
Income quartile 2	**2.723***	0.992	**1.781***	1.547	**3.887***
	(1.271–5.833)	(0.498–1.977)	**(1.039–3.051)**	(0.760–3.150)	**(1.358–11.13)**
Income quartile 3	**2.838***	1.162	1.459	1.045	1.863
	(1.147–7.023)	(0.524–2.577)	(0.771–2.758)	(0.489–2.232)	(0.583–5.950)
Income quartile 4	1.876	0.746	0.998	1.375	1.926
	(0.695–5.065)	(0.278–2.003)	(0.452–2.203)	(0.502–3.768)	(0.443–8.370)
Education **(ref. = 1, no qualifications)**					
Education 2	0.750	0.573	1.219	1.038	1.291
	(0.318–1.317)	(0.340–1.233)	(0.700–2.124)	(0.501–2.148)	(0.477–3.496)
Education 3	0.729	1.061	1.254	1.284	0.511
	(0.372–3.014)	(0.727–3.937)	(0.630–2.496)	(0.548–3.004)	(0.105–2.489)
Wealth **(ref. = 1, lowest quartile)**					
Wealth quartile 2	0.603	1.019	0.825	0.602	0.598
	(0.258–1.411)	(0.463–2.244)	(0.455–1.495)	(0.294–1.233)	(0.216–1.656)

Wealth quartile 3	**0.221****	0.587	**0.478***	**0.200****	0.288
	(0.085–0.574)	(0.264–1.304)	**(0.246–0.931)**	**(0.082–0.490)**	(0.065–1.275)
Wealth quartile 4	**0.324***	**0.323***	0.713	**0.370***	0.870
	(0.126–0.833)	**(0.121–0.866)**	(0.346–1.469)	**(0.146–0.938)**	(0.250–3.023)
Male	**0.336****	**0.486***	**0.600***	**0.506***	0.489
	(0.173–0.654)	**(0.272–0.868)**	**(0.378–0.951)**	**(0.278–0.922)**	(0.202–1.185)
Age	**1.069****	**1.058***	**1.030***	**1.086*****	1.011
	(1.028–1.113)	**(1.018–1.100)**	**(1.002–1.06)**	**(1.046–1.128)**	(0.96–1.064)
Married	0.678	1.030	0.902	**2.097***	**0.325***
	(0.328–1.401)	(0.572–1.857)	(0.550–1.479)	**(1.088–4.044)**	**(0.112–0.946)**
Intercept	**0.008****	**0.029***	0.083†	**0.001*****	0.142
	(0.000–0.225)	**(0.001–0.763)**	(0.007–0.946)	**(0.000–0.015)**	(0.002–10.403)
Number of observations	434		437	437	
AIC:	3,710,648		2,966.9	2,508.2	
BIC:	3,955,031		3,081.2	2,704.1	
χ^2:	3,955,572		131.1	95.1	

Notes: 95 per cent confidence interval in parentheses and significant relative risk ratios in bold. 'Ref.' denotes the reference category. WALK = walking ability indicator, ADL = activities of daily living indicator, IADL = instrumental activities of daily living indicator. AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. † p<0.10; *p<0.05; **p<0.01; ***p<0.001. *Source:* Authors' calculations based on ELSA data.

2.4 Conclusions

The patterns of disability prevalence and intensity in the last years of life have been successfully described in this work. The first aim of this study was to understand the most common pathways leading to death, without imposing a strict functional form when modelling the data. Using RMLCA, three patterns emerged from the analysis of WALK and IADL, but only two pathways in the case of ADL. First, the Low Disability pathway to death was characterized by general low disability over the eight years preceding death. The second pathway—High Disability—was characterized by increasing probability of high impairment over this period. Finally, the pathway of Medium Disability showed moderate impairment. The second aim of this study was to investigate how these pathways are shaped by socioeconomic factors. We found that the different SES indicators have uneven explanatory powers with respect to disability patterns at the end of life. The socioeconomic gradient in the pathways to death appears to be evident only for wealth. Higher wealth accumulation is associated with less disabled pathways to death. Therefore, those who reach death with low impairment tend to also be those with a high level of net wealth. On the contrary, there are no clear significant differences between income quartiles or educational groups in terms of pathways to death.

This analysis represents the first attempt to study the association between different socioeconomic indicators with disability patterns in the last years of life. Previous studies have limited their analysis to selected indicators with little attention to the comparison of different ones (Liao et al., 1999; Schoeni et al., 2003). By contrast, our analysis studied the association between various socioeconomic indicators and later-life functioning in the last years of life. The heterogeneous results of different socioeconomic indicators highlight the need for more detailed analyses of how socioeconomic variables relate to later-life functioning patterns. The disablement process might have early origins during the life course as well as being dependent on the length of exposure to unfavourable circumstances. In the context of older populations, income may be considered a reflection of economic circumstances over the life course to a lesser extent than wealth (Smith & Kington, 2004). Prolonged exposure to favourable economic conditions may lead to both accumulation of resources and a long-run health advantage, which—in turn—will affect disability patterns in the last years of life. This could explain why those with a high level of wealth are more likely to be in less disabled pathways to death, while high levels of income or education are not associated with a reduction in the time spent in disabled conditions in the last years of life.

The presence of a wealth gradient for disability patterns in the last years of life points in the direction of wealth as a relevant and independent dimension of social stratification (Killewald, Pfeffer, & Schachner, 2017). However, the emergence of a wealth gradient and the absence of a clear association with education and income could be related to the specificity of the English context. The important role of wealth in later life in England has also been found in other studies that have focused on specific health conditions and mortality (Banks, Muriel, & Smith, 2010; Demakakos, Marmot, & Steptoe,, 2012, Demakakos, Biddulph, Bobak, & Marmot, 2015; Tanaka, Gjona, & Gulliford, 2012). Moreover, most of the analyses on trajectories in the later years of life have been based on data from the United States (US). Unlike the US, England has universal health insurance coverage via the National Health Service (NHS), which provides a comprehensive range of free healthcare services over the entire life course. In particular, Banks and colleagues (2010) have found significant income and wealth gradients for mortality in the US but only a wealth-related one in England. According to them, one possible explanation for the absence of the income gradient in England is the welfare benefits received during retirement, as their distribution is largely flat below the median. However, this explanation would not completely be able to justify our result of a greater disadvantage for the second and third income quartile in terms of pathways to death.

Furthermore, by examining the association between disability and SES in the process of dying we overcame the limitations of conventional designs to learn about the socioeconomic gradient in old age. We found wealth-related differences in disability in the last years of life, which could in turn be reflected in the socioeconomic gradient over age. Since those with greater wealth experience a less disabled pathway and consequently a shorter period of impairment before death, the wealth gradient over age may mechanically widen as a consequence of differences in the years preceding death. Therefore, this study highlights the importance of considering the socioeconomic differences in pathways to death for a better understanding of health inequalities in old age.

Finally, our results are relevant to healthcare systems and policy, given the high costs typically incurred at the end of life. In the study of the healthcare expenditures across the life course, the *red herring* hypothesis emphasizes the differential roles of proximity to death and age for older people (Zweifel et al., 1999). While some studies have emphasized that age is a red herring for the large healthcare expenditures in later life and proximity to death is the real driver, other studies have suggested that proximity to death represents one of the important mediators together with morbidity and impairment (Seshamani & Gray, 2004; Howdon & Rice, 2015; Hazra, Rudisill, & Gulliford, 2017). Our results suggest that the healthcare expenditures in the last years of life may differ greatly according to the individuals' accumulated resources. Understanding better the mechanisms through which wealth is related to disability in the pathways to death could contribute to more efficient management of healthcare costs in last years of life.

Our analysis is not free from limitations. First, the time separating observed health outcomes is relatively large, since we observed individuals every two years. In addition, we were unable to differentiate those who were interviewed close to their death from those who were interviewed longer before their death. More frequent observations might allow us to distinguish more patterns, although this would obviously also require a lager sample size. Nevertheless, we were still able to examine major changes happening across individuals' last years of life. Second, we did not analyse causes of death, despite the information that different causes could provide regarding the rate of decline, as the reliability of stated cause of death is rather low in older populations because of high comorbidity in later life (Sington & Cottrell, 2002; Harteloh, Bruin, & Kardaun, 2010). Since diseases and health problems occurring in old age tend to be highly interrelated, it is often hard even for medical professionals to establish an individual's true cause of death. Third, sample size limited the extent of our analysis by forcing us to analyse the samples of women and men together. A small sample size could reduce the chances of detecting weaker but significant effects that might nonetheless be present. Consequently, we might be able to observe significant results only for relatively strong effects. Finally, the attrition of more socioeconomically disadvantaged individuals over time may have skewed the sample towards individuals with higher wealth and education, as presented in Table 2. Consequently, we may have underestimated the magnitude of wealth and educational disparities in the disability pathways to death.

Recommended next steps to further our understanding of social variations in pathways to death include investigating the role of different socioeconomic indicators in the disability pathways, by using data with more frequent observations and larger sample sizes. Moreover, the study of disability pathways at the end of life could be enhanced by models that consider the interactions between multiple socioeconomic indicators. The intersection of different socioeconomic conditions could be associated with distinct impairment patterns in the last years of life. We believe that further studies on how SES is related to the pathways to death could contribute to a better understanding of health inequalities in old age.

References

Arber, S., & Ginn, J. (1993). Gender and inequalities in health in later life. *Social Science & Medicine*, *36*(1), 33–46.

Banks, J., Muriel, A., & Smith, J. P. (2010). Disease prevalence, disease incidence, and mortality in the United States and in England. *Demography*, *47*(Supplement), S211–S231.

Bardsley, M., Georghiou, T., & Dixon, J. (2010). *Social Care and Hospital Use at the End of Life*. London. Retrieved from http://www.nuffieldtrust.org.uk/sites/files/nuffield/Social_care_and_hospital_use-full_report_Dec 2010.pdf

Beckett, M. (2000). Converging health inequalities in later life–an artifact of mortality selection? *Journal of Health and Social Behavior*, *41*(1), 106–119.

Biering, K., Hjollund, N. H., & Frydenberg, M. (2015). Using multiple imputation to deal with missing data and attrition in longitudinal studies with repeated measures of patient-reported outcomes. *Clinical Epidemiology*, *7*, 91–106.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: one-step versus three-step estimators. *Political Analysis*, *12*(1), 3–27.

Bradley, E. H., Fried, T. R., Kasl, S. V., & Idler, E. L. (2000). Quality-of-life trajectories of elders in the end of life. In P. M. Lawton (Ed.), *Annual Review of Gerontology and Geriatrics, Vol. 20. Focus on the end of life: scientific and social issues* (pp. 64–96). New York: Springer.

Cai, L., & Lubitz, J. (2007). Was there compression of disability for older Americans from 1992 to 2003? *Demography*, *44*(3), 479–495.

Case, A., & Paxson, C. H. (2005). Sex differences in morbidity and mortality.

Demography, *42*(2), 189–214.

Collins, L. M., & Lanza, S. T. (2010). *Latent Class and Latent Transition Analysis: with Applications in the Social, Behavioral, and Health Sciences*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Cutler, D. M., Lleras-Muney, A., & Vogl, T. (2011). Socioeconomic status and health: dimensions and mechanisms. In S. Glied & P. C. Smith (Eds.), *The Oxford Handbook of Health Economics* (pp. 1–26). Oxford: Oxford University Press.

Cutler, D. M., & Meara, E. (2001). The concentration of medical spending: an update. In D. A. Wise (Ed.), *Themes in the Economics of Aging* (pp. 217–240). Chicago: University of Chicago Press.

Demakakos, P., Biddulph, J. P., Bobak, M., & Marmot, M. G. (2015). Wealth and mortality at older ages: a prospective cohort study. *Journal of Epidemiology and Community Health*, *70*(4), 346–353.

Demakakos, P., Marmot, M. G., & Steptoe, A. (2012). Socioeconomic position and the incidence of type 2 diabetes: the ELSA study. *European Journal of Epidemiology*, *27*(5), 367–378.

Diehr, P., & Patrick, D. L. (2003). Trajectories of health for older adults over time: accounting fully for death. *Ann Intern Med*, *139*, 416–420.

Diggle, P. J., Sousa, I., & Chetwynd, A. G. (2008). Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture. *Statistics in Medicine*, *27*(16), 2981–2998.

Dodge, H. H., Du, Y., Saxton, J. A., & Ganguli, M. (2006). Cognitive domains and trajectories of functional independence in nondemented elderly persons. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *61*(12), 1330–1337.

Dupre, M. E. (2007). Educational differences in age-related patterns of disease: reconsidering the cumulative disadvantage and age-as-leveler hypotheses. *Journal of Health and Social Behavior*, *48*(1), 1–15.

Elwert, F., & Christakis, N. A. (2008). Wives and ex-wives: a new test for homogamy bias in the widowhood effect. *Demography*, *45*(4), 851–873.

Ferraro, K. F. (2011). Health and aging: early origins, persistent inequalities. In R. A. J. Settersten & J. L. Angel (Eds.), *Handbook of Sociology of Aging* (p. 689). New York: Springer.

Freedman, V. A., & Martin, L. G. (1998). Understanding trends in functional limitations among older Americans. *American Journal of Public Health*, *88*(10), 1457–1462.

Freedman, V. A., Martin, L. G., Schoeni, R. F., & Cornman, J. C. (2008). Declines in late-life disability: the role of early- and mid-life factors. *Social Science and Medicine*, *66*(7), 1588–1602.

Fried, L. P., Young, Y., Rubin, G., & Bandeen-Roche, K. (2001). Self-reported preclinical disability identifies older women with early declines in performance and early disease. *Journal of Clinical Epidemiology*, *54*(9), 889–901.

George, L. K. (2009). Conceptualizing and measuring trajectories. In G. H. Elder & J. Z. Giele (Eds.), *The Craft of Life Course Research* (pp. 163–186). New York: Guilford Press.

Gerstorf, D., & Ram, N. (2013). Inquiry into terminal decline: five objectives for future study. *Gerontologist*, *53*(5), 727–737.

Gerstorf, D., Ram, N., Lindenberger, U., & Smith, J. (2013). Age and time-to-death trajectories of change in indicators of cognitive, sensory, physical, health, social, and self-related functions. *Developmental Psychology*, *49*(10), 1805–1821.

Gill, T. M., Gahbauer, E. A., Han, L., & Allore, H. G. (2010). Trajectories of disability in the last year of life. *New England Journal of Medicine, 362*(13), 1173–1180.

Glaser, B. G., & Strauss, A. L. (1968). *Time for Dying*. Chicago: Aldine.

Goldman, D., & Lakdawalla, D. (2001). *Understanding health disparities across education groups* (NBER Working Paper 8328). Cambridge, MA: National Bureau of Economic Research.

Guralnik, J. M., Fried, L. P., & Salive, M. E. (1996). Disability as a public health outcome in the aging population. *Annual Review of Public Health*, *17*(1), 25–46.

Guralnik, J. M., LaCroix, A. Z., Branch, L. G., Kasl, S. V., & Wallace, R. B. (1991). Morbidity and disability in older persons in the years prior to death. *American Journal of Public Health, 81*, 443–447.

Harteloh, P., Bruin, K. D., & Kardaun, J. (2010). The reliability of cause-of-death coding in The Netherlands. *European Journal of Epidemiology*, *25*, 531–538.

Hazra, N. C., Rudisill, C., & Gulliford, M. C. (2017). Determinants of health care costs in the senior elderly: age, comorbidity, impairment, or proximity to death? *European Journal of Health Economics,* 1–12.

Herd, P., Goesling, B., & House, J. S. (2007). Socioeconomic position and health: the differential effects of education versus income on the onset versus progression of health problems. *Journal of Health and Social Behavior*, *48*, 223–238.

Hoffmann, R. (2008). *Socioeconomic Differences in Old Age Mortality*. Dordrecht: Springer Netherlands.

House, J. S., Lantz, P. M., & Herd, P. (2005). Continuity and change in the social stratification of aging and health over the life course: evidence from a nationally representative longitudinal study from 1986 to 2001/2002 (Americans' Changing

Lives Study). *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(Special Issue 2), S15–S26.

Howdon, D., & Rice, N. (2015). Health care expenditures, age, proximity to death and morbidity: implications for an ageing population. *IDEAS Working Paper Series from RePEc*.

Kelley-Moore, J. A., & Lin, J. (2012). Widening the view: capturing “unobserved” heterogeneity in study of age and the life course. In R. A. Settersten & J. L. Angel (Eds.), *Handbook of sociology of aging* (pp. 51–68). New York: Springer.

Killewald, A., Pfeffer, F. T., & Schachner, J. N. (2017). Wealth inequality and accumulation. *Annual Review of Sociology*, *43*(1), 379–404.

Klijs, B., Mackenbach, J. P., & Kunst, A. E. (2010). Disability occurrence and proximity to death. *Disability and Rehabilitation*, *32*(21), 1733–1741.

Liang, J., Wang, C. N., Xu, X., Hsu, H. C., Lin, H. S., & Lin, Y. H. (2010). Trajectory of functional status among older Taiwanese: gender and age variations. *Social Science and Medicine*, *71*(6), 1208–1217.

Liao, Y., McGee, D. L., Kaufman, J. S., Cao, G., & Cooper, R. S. (1999). Socioeconomic status and morbidity in the last years of life. *American Journal of Public Health*, *89*(4), 569–572.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: an R package for polytomous variable latent class analysis. *Journal of Statistical Software*, *42*(10), 1–29.

Liu, H., & Zhang, Z. (2013). Disability trends by marital status among older Americans, 1997-2010: an examination by gender and race. *Population Research and Policy Review*, *32*(1), 103–127.

Lleras-Muney, A., & Lichenberg, F. R. (2002). *The effect of education on medical technology adoption: Are the more educated more likely to use new drugs?*

(NBER Working Paper 9185). Cambridge, MA: National Bureau of Economic Research.

Lunney, J. R., Lynn, J., Foley, D. J., & Lipson, S. (2003). Patterns of functional decline at the end of life. *JAMA*, *289*(18), 2387–2392.

Lynch, S. M. (2003). Cohort and life-course patterns in the relationship between education and health: a hierarchical approach. *Demography*, *40*(2), 309–331.

Lynch, S. M. (2015). Commentary on "disability trajectories at the end of life: a countdown model" the problems with time-to-death as a predictor of disability. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *70*(5), 753–756.

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2011). Aging and the shape of cognitive change before death: terminal decline or terminal drop? *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *66 B*(3), 292–301.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 176–199). Thousand Oaks, Calif.: SAGE Publications, Inc.

Marmot, M., J. Nazroo, J. Banks, R. Blundell, B. Erens, C. Lessof, and F. A. Huppert. 2011. *English Longitudinal Study of Ageing: Waves 0–5, 1998–2011*. Colchester, Essex: UK Data Archive [distributor].

Martin, L. G., Schoeni, R. F., & Andreski, P. M. (2010). Trends in health of older adults in the United States: past, present, future. *Demography*, *47 S*, S17–S40.

Montez, J. K., & Hayward, M. D. (2014). Cumulative childhood adversity, educational attainment, and active life expectancy among U.S. adults. *Demography*, *51*(2), 413–435.

Morgan, L. A., & Kunkel, S. (Eds.). (2016). *Aging, Society, and the Life Course* (Fifth). New York: Springer.

Murphy, T. E., Han, L., Allore, H. G., Peduzzi, P. N., Gill, T. M., & Lin, H. (2011). Treatment of death in the analysis of longitudinal studies of gerontological outcomes. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *66 A*(1), 109–114.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569.

OECD. (2013). Framework for integrated analysis. In *OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth* (pp. 171–192). Paris.

Pienta, A. M., Hayward, M. D., & Jenkins, K. R. (2000). Health consequences of marriage for the retirement years. *Journal of Family Issues*, *21*(5), 559–586.

Poterba, J. M., Venti, S. F., & Wise, D. A. (2013). *Health, education, and the post-retirement evolution of household assets* (NBER Working Paper 18695). Cambridge, MA: National Bureau of Economic Research.

Poterba, J. M., Venti, S., & Wise, D. (2011). The composition and drawdown of wealth in retirement. *Journal of Economic Perspectives*, *25*(4), 95–118.

Power, C., Matthews, S., & Manor, O. (1998). Inequalities in self-rated health: Explanations from different stages of life. *The Lancet*, *351*(9108), 1009–1014.

Rajan, K. B., & Leurgans, S. E. (2010). Joint modeling of missing data due to non-participation and death in longitudinal aging studies. *Statistics in Medicine*, *29*(21), 2260–2268.

Robert, S. A., & House, J. S. (1994). Social structures, quality of life, and aging. In R. Abeles, H. Gift, & M. G. Ory (Eds.), *Aging and Quality of Life* (pp. 253–274). New York: Springer Publishing Company.

Robert, S. A., & House, J. S. (1996). SES differentials in health by age and alternative indicators of SES. *Journal of Aging and Health*, *8*(3), 359–388.

Rohlfsen, L. S., & Kronenfeld, J. J. (2014). Gender differences in trajectories of self-rated health in middle and old age: an examination of differential exposure and differential vulnerability. *Journal of Aging and Health*, *26*, 637–662.

Ross, C. E., & Wu, C.-L. (1996). Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior*, *37*(1), 104–120.

Schoeni, R. F., Freedman, V. A., & Wallace, R. B. (2003). Late-life disability trajectories and socioeconomic status. *Annual Review of Gerontology and Geriatrics*, *22*, 177–199.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Seshamani, M., & Gray, A. (2004). Ageing and health-care expenditure: The red herring argument revisited. *Health Economics*, *13*(4), 303–314.

Simm, A., Nass, N., Bartling, B., Hofmann, B., Silber, R. E., & Navarrete Santos, A. (2008). Potential biomarkers of ageing. *Biological Chemistry*, *389*(3), 257–265.

Sington, J. D., & Cottrell, B. J. (2002). Analysis of the sensitivity of death certificates in 440 hospital deaths: a comparison with necropsy findings. *Journal of Clinical Pathology*, *55*, 499–502.

Smith, J. P., & Kington, R. (2004). *Race, Socioeconomic Status, and Health in Late Life* (HEW). EconWPA. Retrieved from https://econpapers.repec.org/RePEc:wpa:wuwphe:0403001

Spillman, B. C. (2004). Changes in elderly disability rates and the implications for health care utilization and cost. *The Milbank Quarterly*, *82*(1), 157–194.

Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort profile: the English Longitudinal Study of Ageing. *International Journal of Epidemiology*, *42*(6), 1640–1648.

Stuck, A. E., Walthert, J. M., Nikolaus, T., Büla, C. J., Hohmann, C., & Beck, J. C. (1999). Risk factors for functional status decline in community-living elderly people: a systematic literature review. *Social Science & Medicine*, *48*(4), 445–469.

Tanaka, T., Gjona, E., & Gulliford, M. C. (2012). Income, wealth and risk of diabetes among older adults: Cohort study using the English longitudinal study of ageing. *European Journal of Public Health*, *22*(3), 310–317.

Taylor, M. G. (2010). Capturing transitions and trajectories: The role of socioeconomic status in later life disability. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *65 B*, 733–743.

Torssander, J., & Erikson, R. (2010). Stratification and mortality: a comparison of education, class, status, and income. *European Sociological Review*, *26*(4), 465–474.

Tsiatis, A. A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, *14*, 809–834.

Williams, K., & Umberson, D. J. (2004). Marital status, marital transitions, and health: a gendered life course perspective. *Journal of Health and Social Behavior*, *45*(1), 81–98.

Willson, A. E., Shuey, K. M., & Elder, G. H. (2007). Cumulative advantage processes as mechanisms of inequality in life course health. *American Journal of*

Sociology, *112*(6), 1886–1924.

Wilson, R. S., Beckett, L. A., Bienias, J. L., Evans, D. A., & Bennett, D. A. (2003). Terminal decline in cognitive function. *Neurology*, *60*(11), 1782–1787.

Wise, D. A. (1998). *Frontiers in Economics of Aging*. (D. Wise, Ed.). Chicago: The University of Chicago Press.

Wolf, D. A., Freedman, V. A., Ondrich, J. I., Seplaki, C. L., & Spillman, B. C. (2015). Disability trajectories at the end of life: a "countdown" model. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *70*(5), 745–752.

Yang, Y., & Lee, L. C. (2009). Sex and race disparities in health: cohort variations in life course patterns. *Social Forces*, *87*(4), 2093–2124.

Zimmer, Z., Liu, X., Hermalin, A., & Chuang, Y. L. (1998). Educational attainment and transitions in functional status among older Taiwanese. *Demography*, *35*(3), 361–375.

Zimmer, Z., Martin, L. G., Nagin, D. S., & Jones, B. L. (2012). Modeling disability trajectories and mortality of the oldest-old in China. *Demography*, *49*(1), 291–314.

Zweifel, P., Felder, S., & Meiers, M. (1999). Ageing of population and health care expenditure: a red herring? *Health Economics*, *8*(6), 485–496.

Chapter 3

Decomposing the Rising Educational Inequality in Health: The Role of Marriage

Abstract

According to recent studies, educational inequality in health is widening in the United States, but the reasons are still not clear. This study examines to what extent widening educational differences in self-reported health (SRH) in the United States can be attributed to changes over cohorts in the marital status composition of the least and highest educated groups. Therefore, how rising educational inequalities in SRH can be explained by changes in marital distribution by education is analysed using repeated cross-sectional data from the National Health Interview Survey (1976–2014). First, the changes in the relationship between marriage and health over cohorts within different educational groups is examined. Subsequently, the differences in SRH between high school or less educated and college-educated individuals is decomposed using a non-linear version of the Oaxaca-Blinder method, and the proportion attributable to marital status in different cohorts is calculated. The results suggest that married individuals over cohorts are better off in terms of SRH only within the college-educated group. Second, there is a sharp increase over cohorts in the proportion of educational differences in SRH which are explained by marriage. The lower number over time of married individuals in the least educated groups as compared to the highest educated group contributes more and more to the increase in inequalities of SRH; nonetheless, the contribution of marriage as an explanation of SRH differences remains modest. Therefore, married college-educated individuals appear over cohorts as an increasingly advantaged group in terms of SRH. This analysis highlights the need for greater attention to changing demographic characteristics in educational categories when looking at trends over time. The results indicate that while marital status composition across different educational groups is increasingly able to explain the educational SRH gap, marital status itself explains only a small part of the educational differences.

Keywords: health disparities; marriage; education; trends; U.S. population

3.1 Introduction

Educational inequality in health is widening in the United States (Elo & Preston, 1996; Montez, Hummer, Hayward, Woo, & Rogers, 2011; Olshansky et al., 2012). Differences in smoking and obesity trends, increasing income inequality and differential access to healthcare for the lowest strata of the population have been suggested as potential explanations (Ho & Fenelon, 2015; Meara, Richards, & Cutler, 2008; Montez & Zajacova, 2013). However, these widening educational differences might also be explained by changes in the composition of various social groups over time (Hayward, Hummer, & Sasson, 2015). After decades of educational expansion, the current batch of graduates from higher education institutions differ in many respects (e.g., parental socioeconomic background and other characteristics) from higher education graduates in birth cohorts of the 1920s and 1930s. The role of this changing composition of educational groups on the socioeconomic gradient and its implications on health has not been adequately analysed, although some studies have recognised its possible importance (Dowd & Hamoudi, 2014; Hendi, 2015; Ho & Fenelon, 2015; Montez & Friedman, 2015; Sasson, 2016).

In the past century, the *second demographic transition* and educational expansion have had an extensive impact on the living arrangements, partnerships and educational attainment of individuals. Education is increasingly connected to changing mechanisms of assortative mating, marriage and family stability (Schwartz, 2013; Sweeney, 2002). At the same time, marriage is associated with better health, as well as greater longevity (Martikainen, Martelin, Nihtilä, Majamaa, & Koskinen, 2005; Waite & Gallagher, 2000). Therefore, it is crucial to examine the role of marriage in the analysis of the changing relationship between education and health. The emergent educational gap in the prevalence and stability of marriage has been suggested as potentially responsible for increasing socioeconomic inequalities (Goldstein & Kenney, 2001; McLanahan, 2004; Meadows, Mclanahan, & Brooks-Gunn, 2008). Lower levels of education are increasingly associated with a decreased likelihood of marriage and as well as a higher probability of marital instability. Are these changes in marital status by education also associated with an increase in health disparities for different educational groups? This analysis examines to what extent widening educational differences in self-reported health (SRH) in the United States can be attributed to changes in the marital status composition of different educational groups over the 20th century. We measure health on the basis of SRH, which is has been shown to be a strong predictor of mortality (Benjamins, Hummer, Eberstein, & Nam, 2004; Idler & Benyamini, 1997; Idler, Russell, & Davis, 2000). Marital status is measured cross-sectionally, focusing on the association between current marital circumstances and SRH with the aim of offering a snapshot picture without considering marital history for which we do not have data available.

Previous literature has considered trends in the relationship between marital status, health and education separately, but has disregarded the intersection between marital status and education when studying trends in health (Goesling, 2007; Liu & Umberson, 2008). Many studies have found evidence for a *protective effect* of marriage on health, which does not seem to be exclusively determined by selection mechanisms (Hu & Goldman, 1990; Lillard & Panis, 1996; Mirowsky & Ross, 2003; Rendall, Weden, Favreault, & Waldron, 2011; Waite & Gallagher, 2000). Conversely, previous research has suggested that divorce negatively affects health and mortality (Kalmijn, 2017; Lorenz, Wickrama, Conger, & Elder Jr., 2006; Williams & Umberson, 2004; Wood, Goesling, & Avellar, 2007; Zhang & Hayward, 2006). Therefore, when analysing the determinants of the trend in which college graduates have increasingly better health as compared to less educated individuals, the role of marriage cannot be neglected.

The analysis of the role of marital status on SRH differences by education follows two steps. First, how the relationship between marriage and SRH changes across cohorts within different educational groups is analysed. Second, in order to determine which proportion of the difference in SRH can be explained by marriage in each birth cohort and for different age groups the observed differences in SRH between high school or less educated and college or more educated individuals is decomposed. Therefore, this analysis aims at decomposing the association between education and SRH over cohorts by examining the role of marital arrangements.

This study contributes to the existing literature in four ways. First, it highlights the increasingly important role of current marital status for trends in health disparities among educational groups. Miech and colleagues (2011) drew attention to a new area of research which studies the social determinants that change health disparities over time. This analysis advances research in this newly formed area by highlighting the importance of the demographic composition of the population to examine trends in educational inequality in health. Moreover, this is part of a broader social science issue, such that changing group characteristics need to be taken into account for a meaningful understanding of group differences.

Second, this study provides significant insights into trends of marital status on health by educational attainment. In particular, the s*econd demographic transition* and educational expansion happened at about the same time and resulted in profound societal changes. This analysis explicitly accounts for one possible changing factor—education—associated with different marital categories. Consequently, this analysis considers marital status and education simultaneously, which is important when looking at trends in health differentials.

Third, we take a cohort perspective in the analysis of these trends, unlike previous works that have focused on period changes. The cohort perspective has been underutilised for understanding trends in health in the United States, albeit with some exceptions (Lauderdale, 2001; Lynch, 2003; Masters, Hummer, & Powers, 2012). Thus far, the period perspective has been the mainstay of efforts to determine trends in the association between education and health, and between marriage and health over time (Goesling, 2007; Liu & Umberson, 2008). Therefore, is crucial to further develop the cohort analysis for socioeconomic differences in health in order to introduce the life-course perspective to this area of research. In fact, socio-historical conditions experienced by different cohorts shape the life-course events of individuals, these include education and marital arrangements (Davey Smith & Lynch, 2004; Riley, 1987). The cohort perspective helps to analyse the common factors that affect education, marriage and health simultaneously. Change can be the product of variations in the demographic composition of educational groups, possibly due to common experiences faced (Mannheim, 1952; Ryder, 1965). For example, decisions related to marriage and marital stability can be influenced by cohort-specific characteristics, as well as events (Pampel & Peters, 1995; Sweeney, 2002). Similarly, previous research has found that health decline associated with aging is different across cohorts (Willson, Shuey, & Elder, 2007; Yang & Lee, 2009).

Finally, this work aims to contribute to the larger literature on how family processes affect larger population outcomes. In particular, the focus for this analysis is on understanding the rising health disparities between high school or less educated (for simplicity identified as *low*) and college-educated (*high*) groups in the context of increasing differentiation of marital arrangements over cohorts. As family structures change, this can shape the reproduction of socioeconomic disparities (McLanahan & Percheski, 2008), and in particular, health disparities.

3.2 Background

First, this study analyses to what extent the returns of marriage on SRH are changing over cohorts within different educational groups; second, it investigates the extent to which the changing prevalence of married individuals at the top and bottom of the educational distribution contributes to rising health disparities. Therefore, this section provides an overview of partnership formation over time. It also reviews the hypothesised mechanisms, according to the literature, for the gap in marriage and marital stability in different educational groups. Finally, it discusses the theoretical frameworks used to explain the association between marriage and health over cohorts, while highlighting the role of education.

3.2.1 Marriage, marital stability and education

Marriage formation has undergone a number of important changes over the last century. While increasing economic prospects, as well as rising educational attainment for women, have not reduced the likelihood of marriage overall, they have nonetheless produced socioeconomic differences in marriage patterns (Goldstein & Kenney, 2001; Sweeney, 2002). In particular, an educational gradient in marriage has emerged: individuals with advanced education are more likely to live in a marital union than those with lower educational attainment, despite marrying at a later age (Torr, 2011). This reduction in the prevalence of individuals living in a marital union for the least educated group can be explained by their lower likelihood to marry, as well as higher prevalence of divorce (Goldstein & Kenney, 2001; White & Rogers, 2000). Several studies have highlighted the increase in marital instability in the United States by socioeconomic status and education (S. P. Martin, 2006; Raley & Bumpass, 2003). Previous studies have revealed the importance of marital biographies for health and mortality (Blomgren, Martikainen, Grundy, & Koskinen, 2012; Grundy & Tomassini, 2010; Hughes & Waite, 2009; Murphy, Glaser, & Grundy, 1997; Ploubidis, Silverwood, DeStavola, & Grundy, 2015). In particular, remarriage is associated with some health benefits, but to lower extent compared to first order marriages (Hughes & Waite, 2009; Williams & Umberson, 2004), despite being linked to health behaviours similar to those in the first marriage (Keenan, Ploubidis, Silverwood, & Grundy, 2017). However, given the cross-sectional nature of our data and the absence of retrospective information, we cannot observe exposure to different transitions and marital circumstances, but only to the current prevalence of marital status. Therefore, this analysis uses measures of prevalence of marriage instead of incidence with the aim of representing the association between health and the proportion of people in a certain marital status at a given moment, and not focusing on previous marital history or on the risk of a new marital event occurring within a certain time frame.

The changes that have taken place in marriage trends during the twentieth century have been extensive and can be summarised according to four changing elements: marriage timing, meaning of marital unions, patterns of assortative mating by education, and the emergence of cohabitation and divorce. First, the degree of variability in age at marriage first declined, then increased for birth cohorts of the 1950s before plateauing for birth cohorts from the 1960s and 1970s at around 25 years old (Sweeney, 2016). Moreover, Sweeney (2016) observes that different socioeconomic groups have followed similar trends over the past century despite dissimilarities in the magnitude and exact timing of the shifts in age at first marriage. Therefore, the changes seen in the timing of marriage seem to be driven by changes that are common to all social groups.

Second, Cherlin (2004) has suggested the idea that the meaning of marriage has changed over time and, in particular, that marriage has become increasingly deinstitutionalised. For example, this trend towards deinstitutionalisation can be seen the increasing rate and complexity of cohabiting unions and the emergence of same-sex marriages. In addition, Cherlin (2004) describes the changing meaning of marriage from the institutional meaning to companionship and then to individualised marriage, which is considered a means of self-development. According to his study, marriage has increasingly become a symbol of personal success and prestige. Other scholars have argued against the significance of marriage deinstitutionalisation, suggesting that other family arrangements, such as cohabitation and same-sex marriage, are simply alternative models of marriage and not necessarily changes in the institution of marriage (Lauer & Yodanis, 2010). These scholars emphasise how socioeconomic groups that are less likely to marry keep valuing marriage and viewing it as something to aspire towards.

Other important changes in marriage over time concern patterns of assortative mating, especially in regards to education and income. Educational expansion has created new opportunities for young people to meet potential partners with similar characteristics (Blossfeld, 2009). Educational homogamy has increased in the course of the twentieth century, principally due to a decrease in intermarriages between college and less than high school-educated individuals (Schwartz & Mare, 2005). Similarly, other studies have shown that the increase in educational homogamy in Canada and the United States can be explained by the increase in women with post-secondary education (Hou & Myles, 2008). Changes in the meaning of marriage from gender specialised to egalitarian partnerships are evident in the analysis of Schwartz and Han (2014). These differences in educational attainment within marital couples have implications for marital stability. A wife with an educational advantage was found to have a higher likelihood of marital dissolution, but this effect has since vanished over time. Moreover, the issue of wage/income disparity between men and women has gained considerable importance in recent years. As such, increases in what women can earn has implications for marital formation and marriage prospects over time (Sweeney, 2002; Sweeney & Cancian, 2004). In addition, homogamy is increasingly associated with a lower likelihood of divorce as compared to couples with traditional breadwinner arrangements.

In addition, Van Bavel et al. (2018) explore the consequences of changing patterns of assortative mating in the context of the reversing gender gap in education. Over time, wives with a higher level of education as compared to their husbands (i.e. educational hypogamy) has become more common than the reverse (Esteve, García-román, & Permanyer, 2012; Esteve et al., 2016). However, while union instability for educational hypogamous couples was higher in the past, this is not the case anymore, at least in the United States (Schwartz & Gonalons-Pons, 2016). These changes in the characteristics of partners might also affect individual health. For example, in addition to one's own education, the level of education of one's spouse can influence health status through the exchange of material and non-material resources (D. C. Brown, Hummer, & Hayward, 2014; Huijts, Monden, & Kraaykamp, 2010; Kravdal, 2017; Monden, Van Lenthe, De Graaf, & Kraaykamp, 2003). Brown and colleagues (2014) have shown that both higher own and partner education are associated with better SRH in a pooled sample of adults in the United States.

Finally, the emergence of cohabitation as an alternative marital arrangement in less socioeconomically advantaged and race groups represents a relevant change in the structure of the family in the last century (Bramlett & Mosher, 2002; Bumpass & Lu, 2000; Cherlin, 2010; Smock, 2000). Similarly, marital dissolution has become increasingly common in more disadvantaged socioeconomic groups (Phillips & Sweeney, 2005; Raley & Bumpass, 2003; Raley, Sweeney, & Wondra, 2015). The increasing association between socioeconomic deprivation, cohabitation and marital dissolution is cause for concern for different domains of inequality (Kennedy & Bumpass, 2008; Smock & Manning, 2004).

Ruggles (2015) attributes a significant part of this change in marital formation and stability to the worsening socioeconomic conditions faced by young adult men as compared with previous cohorts. The decline in relative income, waning importance of labour unions and outsourcing of manufacturing in the globalised economy are among the factors that have contributed to this shift in the relationship between socioeconomic status and marriage (Ruggles, 2015). Therefore, educational expansion, in combination with the aforementioned changes in the likelihood of marriage and divorce, has led to a massive transformation in the prevalence of married individuals by educational attainment.

3.2.2 Trends in the association between marriage and health: the role of education

This study looks at the intersection of marriage and education when analysing health trends. Resource substitution and multiplication have been proposed as theoretical frameworks for understanding how the health benefits of marriage vary across groups (Ross & Mirowsky, 2006, 2010; Tumin, 2017). Resource substitution theory hypothesises the existence of compensative effects: the absence of a resource is compensated by the presence of another resource. Resource multiplication supposes that additional resources create a multiplicative effect in the benefits derived from those resources. We use these theories to analyse how the health advantage linked to marriage changes over cohorts within different educational groups. If the health benefits of marriage for those with low levels of education are larger in less recent cohorts, there is support for resource multiplication theory because people with low levels of education are subject to less disadvantageous conditions in terms of employment and income distribution in earlier cohorts (Lemieux, 2008; Ruggles, 2015; Sum, Khatiwada, McLaughlin, & Palma, 2011). Similarly, for the highly educated, we expect greater benefit from marriage in the most recent cohorts. If the resource multiplication framework applies, then higher education should confer more advantages linked to marriage in recent cohorts. Alternatively, the resource substitution framework should predict more benefits for more recent cohorts within the less educated group and for earlier cohorts within the highly educated, because according to the model, resources replace each other as compensation.

Married individuals have better health and have a longer life expectancy than people who are unmarried or divorced (Rendall et al., 2011; Waite & Gallagher, 2000; Wood et al., 2007). Three theories have been proposed to explain the association between marital status and health: marital resource, stress and selection models. The positive association between marriage and health could be causal (according to marital resource and stress models) or due to selection mechanisms (Joung, Van De Mheen, Stronks, van Poppel, & Mackenbach, 1998). These three frameworks are used to examine trends over time in the marriage–health association considering education as an important dimension of heterogeneity. In fact, these theories may apply asymmetrically to different educational groups over time.

According to marital resource theory, marriage provides economic, social and psychological resources, as well as a sense of obligation, which reinforces social integration. First, following Becker (1981), the decline in the prevalence and stability of marital unions is attributable to a decline in the economic benefits associated with them. Although this hypothesis has found very weak empirical support, especially in micro-level studies (Oppenheimer, 1997), it implies that the association between marriage and health is expected to decrease over time. However, the economic benefits from marriage decrease to a different extent across the educational spectrum. On one side, the economic benefits from marriage for the less educated might be smaller because of the decreasing socioeconomic resources available over time. On the other side, in the presence of increasingly limited socioeconomic resources, marriage among the least educated might be more beneficial for health.

Second, the change in psychological benefits from marriage is not straightforward. In particular, marital quality might be a useful proxy for psychological support offered by the partner. A unified picture of how marital quality has changing over time has not emerge (Amato, Booth, Johnson, & Rogers, 2007). Amato and colleagues (2007) observe that while marital interactions have decreased over time, so too have marital conflicts and problems. Therefore, despite the previously found positive association between marital quality and stability, the increase in educational attainment has not affected marital quality over time.

Third, the social resources provided by marriage in terms of the number of close friends and connections have declined over time (Amato et al., 2007). Social networks seem to be especially important for a successful marriage (Bryant & Conger, 1999), and they might possibly have a positive influence on health (Uchino, 2004; Umberson & Montez, 2010). However, we have also observed that couples are, over time, becoming increasingly disconnected from society because they have increasingly smaller social networks. Therefore, because higher education is associated with more extensive social networks (McPherson, Smith-Lovin, & Brashears, 2006), the harmful health consequences of shrinking social networks might be more damaging among the least educated.

Stress theory (or crisis model) attributes the responsibility for the health differences between married and unmarried individuals to the stress connected to marital dissolution. The stress of marital dissolution might diminish over cohorts because of the increased social acceptability of divorce. On the one hand, the stress of marital dissolution is likely to be higher in the least educated groups where couples have less financial independence and smaller support networks from which to seek help (Aseltine & Kessler, 1993; Johnson & Wu, 2002). On the other hand, divorce has become more common among the least educated, which has correspondingly reduced the social stigma within the relative social group; the opposite is true for highly educated individuals going through marital dissolution, who might suffer from higher stress levels given the relative infrequency of divorce among their similarly educated peers.

Finally, the selection model assumes that the health differences between the married and unmarried are due to the better health of those selected into marriage. Thanks to educational expansion, many more individuals have been able to access education despite initial differences in health endowments; therefore, a weakening relationship between marriage and health among the highly educated should be expected. That said, those remaining at the bottom of the educational distribution have tended to be increasingly negatively selected as compared to the past, possibly in terms of health as well (Dowd & Hamoudi, 2014; Montez & Friedman, 2015; Sasson, 2016). This would suggest a strengthening positive association between marriage and health over cohorts for the less educated, because those who get married are those with a more advantageous health endowment.

Marital resource, crisis and selection theories, however, are not mutually exclusive explanations for the association between marriage and health. They do, however, represent useful frameworks to understand how marriage and health are related across cohorts. Similarly, resource multiplication and substitution frameworks are useful theories for examining marriage benefits for health within different educational groups. One of the hypothesis to have emerged from these frameworks is that college-educated people might possibly be a doubly advantaged group in more recent cohorts, since both education and marriage are increasingly associated with more favourable health outcomes. In particular, changes in the marital composition of educational groups could have important consequences for the relationship between education and health.

3.3 Research questions

While rising educational inequality in health has been documented in the literature (Montez et al., 2011), the reasons for this growing inequality remain unclear. Several hypotheses have been suggested and tested (Hendi, 2015; Ho & Fenelon, 2015; Sasson, 2016), but the changing demographic composition of different educational groups has not been adequately analysed. For example, Sasson (2016) suggests that less educated individuals face greater uncertainty about the length of their life over time, possibly because of the decrease in real wages, the increase in poverty and unemployment and decreasing prospects in the marriage market. This last aspect of decreasing possibilities in the marriage market and higher marital instability for the less educated has not been adequately explored in the health inequality research. Marital arrangements in different educational groups have changed considerably in the last century: marriage prevalence has unequally declined leaving the least educated with the lowest levels of ever-married (Cherlin 2010; Goldstein and Kenney 2001). First, this work sheds light on changes in this association over cohorts within different educational groups. Second, it links micro and macro perspectives by studying how the distribution of marital status within groups affects SRH inequalities over cohorts. Therefore, the question guiding this research asks: *How much of the rising educational health inequality can be ascribed to changing marital status composition?*

The first research question looks at whether the change in the relationship between marital status and health over cohorts is the same across the educational spectrum. The intersection between marital status and education is considered when studying increasing disparities in health. In fact, in order to correctly assess the association between education and health over cohorts, it is important to take into account the changing composition of marital status of educational groups. If the association between education and health over time is analysed without considering the simultaneous change in the relationship between marital status and education, it is possible to overestimate the association between education and health. Therefore, a goal of this work is to include education in the analysis of trends of marital status on health. So far, previous analyses have used marital status and education as static controls over time (Goesling, 2007; Liu & Umberson, 2008). However, it is fundamental to consider how the changing heterogeneity in marital status across various educational categories influences the observed relationship between marriage and health over cohorts. Therefore, it can be expected that if resource multiplication theory holds, then married college-educated individuals should have increasingly better health over cohorts.

The second research question analyses the extent to which marital status composition across the educational spectrum is able to explain educational differences in SRH. This analysis aims at explaining part of the rising educational inequality in SRH through the changing marital composition in different educational groups. Educational inequality in SRH is decomposed and the proportion of difference accounted for by marital status composition in each birth cohort is determined. The differential marriage formation and family stability in the least and highest educated groups has changed greatly over cohorts (Goldstein & Kenney, 2001; S. P. Martin, 2006). The Fairlie decomposition method is used to calculate how much of the educational gap in SRH is explained by marital status distribution in the two groups. This is used to quantify the extent to which the proportion of married individuals in the two educational groups is able to account for the observed educational difference in SRH.

3.3.1 Data

The data relies on a pooled repeated cross-section from the National Health Interview Survey (NHIS) 1976–2014. The NHIS is collected by the National Center for Health Statistics every year. Thanks to a multistage probability sampling strategy, the NHIS is representative of non-institutionalised civilian population in the United States. Unfortunately, it is not possible to include institutionalised populations in this analysis, despite the survey respondents usually being healthier as compared to the general population (Baeten, Van Ourti, & van Doorslaer, 2013; Keyes, Rutherford, Popham, Martins, & Gray, 2018). Moreover, institutionalisation rates are declining over time and are related to marital status (Crimmins & Saito, 2001; Liang, Brown, Krause, Ofstedal, & Bennett, 2005). For example, jail and prison inmates have worse health conditions even after controlling for sociodemographic characteristics (Binswanger, Krueger, & Steiner, 2009), potentially leading to the underestimation of educational differences in health. However, the exclusion of this part of the population is unlikely to produce large biases given the low percentage of the population institutionalised among non-Hispanic Whites, especially among people below 80 years of age (Crimmins, Saito, & Ingegneri, 1997). Excluding institutionalised populations can produce more significant biases among non-Hispanic Blacks, a demographic that has experienced increasing mass incarceration over time (Pettit & Western, 2004).

This survey provides the advantage of a long time-serie on health, as well as information on participant demographic and socioeconomic characteristics. This data is appropriate for studying changing health differentials among educational groups over time, given the availability of both socioeconomic and health indicators over a long period. Despite the clear advantage of the time series design, the data collected only reveals information about current marital status. As consequence, the study cannot make any distinction between first and second order marriages or the duration of a marital union, thus limiting the scope of this analysis to current marital circumstances and health. Trends in remarriage by education might influence the association between SRH and marital status. However, the increase in remarriage rates over time could lead us to underestimate the strength of current findings on marital status and health.

Synthetic cohort measures were derived because birth cohorts rather than periods were used to perform the analysis. This work included only non-Hispanic Whites and Blacks because of the relatively smaller number of other ethnic and racial groups. The study population was individuals aged between 30 and 70 (divided over four 10 year age groups) in order to avoid biased estimates due to the shift in the age at marriage. Moreover, the analysis was restricted to individuals in the birth cohorts 1937 to 1967, which represents the longest observable range for different age groups. Birth cohorts were organised in groups of 5 years. Therefore, this analysis concerned three generational groups which have different characteristics, according to Strauss and Howe's theory (1991): the Silent Generation (birth cohorts 1937–1942), the Baby Boomers (1947–1962), and Generation X (1967). Individuals belonging to the Silent Generation have shared experiences of the years following World War II and the Great Depression as their most distinctive trait (Pew Research Center, 2010). Baby boomers also possess a number of characteristic traits, such as a much higher probability of having a college education as compared to previous cohorts (Whitbourne & Willis, 2006). They are the first cohort to experience a decrease in marriage rates, increase in marital dissolution and pre-marital cohabitation (Cherlin, 1992). Moreover, pronounced differences by race and ethnicity emerged in the baby boom generation in socioeconomic resources and marital arrangements (Whitbourne & Willis, 2006). Finally, Generation X was observed only in the 1967 birth cohort at young ages (30–40 years old), which have similar demographic behaviours to the late baby boomers (Tamborini & Iams, 2011).

The health outcome in this analysis is SRH using two different categorisations. First, SRH was recoded as a binary variable assuming a value of 1 if the reported health score was fair or poor and 0 otherwise, a method suggested by other studies (Goesling, 2007; Liu & Hummer, 2008; L. G. Martin, Schoeni, Freedman, & Andreski, 2007). Second, SRH was recoded in four categories in order to be able to perform the linear decomposition. This approach was adopted because SRH is measured slightly differently across survey years. Before 1982, the variable was coded according to a 4-point scale (1 *excellent*, 2 *good*, 3 *fair*, and 4 *poor*); after 1982, however, a new 5-point scale was added (1 *excellent*, 2 *very good*, 3 *good*, 4 *fair*, and 5 *poor*). To ensure that the results would be comparable within periods, *very good* and *excellent* scores after 1982 are combined. Higher score values represent worse health status. Previous studies have demonstrated that SRH is a robust measure for trends over time, despite some variations depending on study design (Avendano, Huijts, & Subramanian, 2009; Benjamins et al., 2004; Farmer & Ferraro, 1997; Hays, Schoenfeld, Blazer, & Gold, 1996; Idler & Benyamini, 1997). However, this measure is not free from limitations: the reliability of SRH has been tested in some studies finding the presence of some bias in the reporting by educational groups (Burgard & Chen, 2014; Dowd & Todd, 2011; Dowd & Zajacova, 2007, 2010; Zajacova & Dowd, 2011). In particular, Zajacova and Dowd (2011) have found that repeated measurements of SRH, only one month apart, in North Americans samples, about 40 per cent of the respondents were unable to provide consistent answers. The reliability of SRH appears to be threaded especially among the most disadvantaged educational groups. Despite this limitation, the focus of this study is on the gap between the low and highly educated over time. Therefore, the presence of some bias in reporting SRH among different educational groups is not expected to change over time, and as a consequence, should not affect the conclusions of this study.

Education was measured according to the highest level of educational attained. In particular, educational attainment was divided over four groups: less than high school, high school, some college and college or more. Considering that there are no information about cohabitation for the whole period (data about cohabitation arrangements has only been collected since 1997), it was not possible to include cohabitation as a marital category for the entire analysis. However, a series of robustness tests using the data on cohabitation from 1997 are performed. Moreover, the likely implications for the absence of data on cohabitation for the whole period are discussed. Marital status was mainly measured with four categories: married, divorced or separated, never-married and widowed. Other control variables were geographic region (Northeast, Midwest, South and West). All the analyses were run separately by sex.

3.3.2 Methods

This work applied a variety of methods. First, logit models were employed to examine changes over cohorts in the association between SRH and marital status. Second, this analysis used Fairlie decomposition (Fairlie, 1999) to determine the role of marital status on the rising health differences between low and highly educated individuals. Fairlie decomposition is a non-linear version of Oaxaca-Blinder decomposition, which is widely used in the demographic literature (Barber, Yarger, & Gatny, 2013; Hayford, 2013; Maurer, 2011). However, the application of this approach to socioeconomic differences in health has been limited thus far (Sen, 2014). Mandel and Semyonov (2014, 2016) showed that Oaxaca-Blider decomposition can be an interesting method for decomposing socioeconomic differences over birth cohorts.

The analytic strategy was to decompose the difference between high school-educated or less and college-educated groups for different birth cohorts. Robustness tests were also conducted using only the two most extreme educational categories: less than high school and college. The four age groups were analysed separately. The decomposition included an indicator of the marital status of each individual, plus a set of controls that includes age and region. A simple married–unmarried distinction was used in the model following the recommendation by Rendall and colleagues (2011): the differentiation between unmarried groups may be more harmful than beneficial because the lower statistical power among the various unmarried categories may 'obscure more than illuminate' (Rendall et al. 2011 p. 502). However, as a robustness check, this study also explored the variability across the unmarried categories.

Fairlie decomposition for non-linear equations focuses only on the *endowment effect,* meaning the explained proportion of the gap. This is the part of the educational gap due to group differences in the distribution of marital status. Therefore, the endowment effect represents the difference in the proportion of married individuals between low and highly educated groups keeping fixed the effect of marital status on health. This term informs on how much the different demographic composition in terms of the marital status of the two educational groups account for the observed differences in SRH.

This non-linear decomposition technique is able to analyse only the endowment effect, but does not, unfortunately, capture the proportion of the educational gap due to unobserved characteristics. As a robustness check, this study employed the linear Oaxaca-Blinder decomposition, which breaks down the SRH differences between explained and unexplained parts. The unexplained part is due to the different effect of marriage on health between low and highly educated after fixing the proportion of married individuals at the least educated level. This term represents the contribution of a differing return for marriage on health between low and highly educated individuals. All of the analyses were performed in Stata 15. Survey weights were used in order to account for the complex sampling design using the 'svy' command in Stata.

3.3.3 Descriptive statistics

Table 1 presents the distribution of characteristics both in female and male samples. The analytic sample is composed of 1,029,671 observations. Missing data is present only for marital status and health; however, these missing observations represent less than 1 per cent of the sample. As Lynch (2003) pointed out, missing data in NHIS does not constitute a reason for concern. Greater access to university education and a lower percentage of people with less than high school education is observed over cohorts. During the educational expansion, the number of people obtaining a college degree increased drastically. Conversely, the number of people with less than a high school diploma decreased dramatically. Simultaneously, for non-Hispanic Whites, differences in SRH scores between high school and less and college graduates increased over cohorts, as is observable in Figure 1, with the exception of individuals aged 30–40, for which we observe a decline. The gap in SRH is much higher in the most recent birth cohorts, which suggests an increasing distance among educational groups over cohorts.

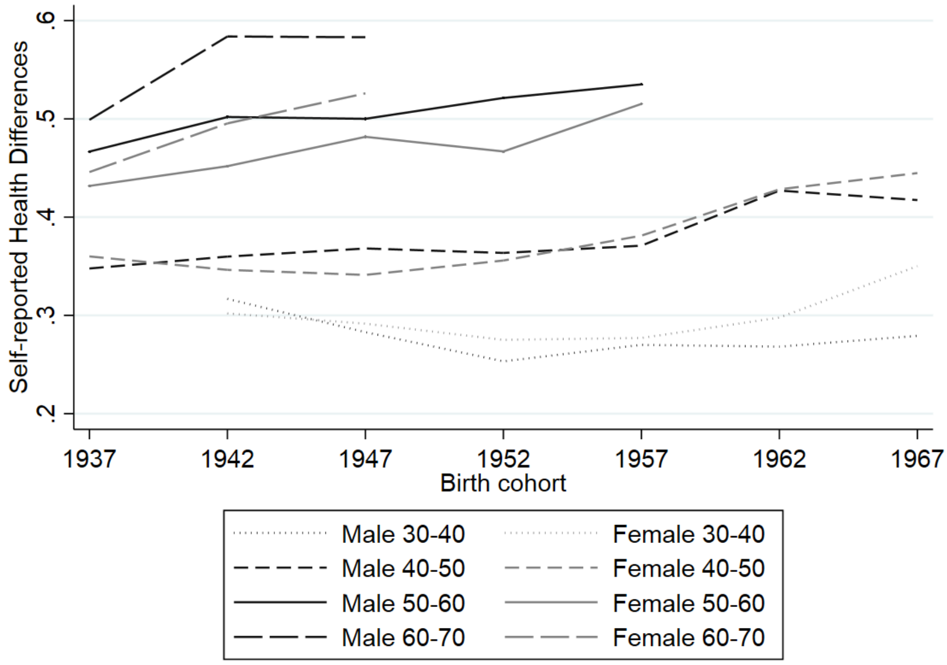

Figure 1 Difference in SRH between high school-educated or less and college or more (non-Hispanic White, cohorts 1937–1967).

Table 1 Characteristics of the Sample (Non-Hispanic White and Black, Percentages)

		White		*Black*	
	All	*Male*	*Female*	*Male*	*Female*
N. Observations (complete sample)	1,029,671	423,072	446,521	67,489	92,589
N. Observations (after 1997)	513,150	205,699	217,199	38,715	51,537
Marital Status					
Married	71.8	76.4	73.5	60.2	43.6
Divorced	15.5	12.4	16	18.7	27.8
Widow	2.4	0.9	3.4	1.7	5.9
Never-married	10.3	10.3	7.1	19.4	22.7
Cohabitation*	4.8	4.9	4.4	7.5	4.4
Education					
Less than high school	9.9	9.3	8.4	17.8	17.3

High school	35.3	33.3	36.3	39.8	37.6
Some college	26.2	24.9	27.3	25.3	28
College or more	28.6	32.5	28	17.1	17.1
Self-reported Health					
Excellent/very good	63.7	67.2	64.6	52.6	45.7
Good	25.4	23.5	25.4	30.5	33.7
Fair	8.1	6.8	7.4	12.7	15.7
Poor	2.8	2.5	2.6	4.2	4.9
Gender					
Male	48.7				
Female	51.3				

Notes: Weights were applied to represent the population distribution accurately. Note that for cohabitation the percentages are calculated only on the sample surveyed after 1997. *Source:* Author's calculation based on NHIS data.

Table 2 displays the characteristics of the sample across selected birth cohorts. The retreat from marriage is particularly evident. The percentage of married individuals declines both in the non-Hispanic White and Black sample. Concurrently, divorce and cohabitation become more common over cohorts. In addition, the number of never-married individuals also increases. Overall, we also observe improvements in SRH, since a larger number of individuals report *excellent*/*very good* health. However, these represent unconditional percentages, meaning that more recent cohorts are probably surveyed to a greater extent at younger ages. The description of how SRH is distributed across age groups is in Table 3. As expected, the prevalence of *fair* and *poor* health increases over age in both non-Hispanic White and Black samples. Conversely, the percentage of people reporting *excellent*/*very good* SRH declines with age. The rate of decline is faster in the non-Hispanic Black sample compared to the non-Hispanic White.

Table 2 Characteristics of the Sample Over Selected Cohorts (Non-Hispanic White and Black, Percentages)

	Birth Cohorts					
	White			Black		
	1937	1947	1957	1937	1947	1957
Marital Status						
Married	79.8	76.4	73.3	55.8	52.9	49.5
Divorce	12	14.5	14.7	24.8	26.8	22.7
Widow	5.2	2.2	1	10.3	5.1	1.8
Never-married	4	6.9	11	9.1	15.2	26
Cohabitation	1.9	3.2	5.3	2.7	3.8	6.4
Male	48.5	49.3	49.8	43.8	44.5	45.6
Education						
Less than high school	16.2	7.9	7.1	35.2	18.1	13.1
High school	41.4	33.6	35	37.7	39.3	39.2
Some college	19.9	26.2	27.8	15.9	25.3	30
College or more	22.5	32.2	30.1	11.2	17.3	17.7
Self-reported Health						
Excellent/very good	56.9	64.6	69.9	36.1	43.9	53.6
Good	29.7	25.2	22.1	35.8	34.7	30.6
Fair	9.6	7.5	6	20.4	16	12.4
Poor	3.8	2.7	2	7.7	5.4	3.4
N. observations	116,170	176,289	128,616	17,736	29,273	26,771

Notes: Weights were applied to represent the population distribution accurately. Note that for cohabitation the percentages are calculated only on the sample surveyed after 1997. *Source:* Author's calculation based on NHIS data.

Table 3 Characteristics of the Sample Over Different Ages (Non-Hispanic White and Black, percentages)

	Age Groups							
	Non-Hispanic White				Non-Hispanic Black			
	30-40	*40-50*	*50-60*	*60-70*	*30-40*	*40-50*	*50-60*	*60-70*
Excellent/ very good	72.4	67.8	59.7	52.6	57	50.8	39.4	30.5
Good	21.8	23.5	26.8	30.1	30.5	31.8	34.2	36.1
Fair	4.7	6.5	9.2	12.4	10.1	13.2	19	25
Poor	1.1	2.2	4.3	4.9	2.4	4.2	7.4	8.4
N. observations	330,968	292,814	171,065	74,746	57,219	55,537	33,531	13,791

Notes: Weights were applied to represent the population distribution accurately. *Source:* Author's calculation based on NHIS data.

Tables 4, 5 and 6 reflect the changes in the joint distribution of marital status and education over cohorts. Although these trends by education have already been partially explored in separate studies of marriage (Goldstein & Kenney, 2001) and divorce (S. P. Martin, 2006), table 4 provides some additional evidence using a different set of data (Cherlin, 2009; Cohen, 2015; Kennedy & Ruggles, 2014; Torr, 2011). First, educational differences in the prevalence of married and divorced people by education are present, offering some support in line with educational divergence in marital dissolution which has received some evidence both in favour (S. P. Martin, 2006; Raley & Bumpass, 2003) and against (South, 2001; Teachman, 2002). However, a more comprehensive test of educational divergence in marriage and marital stability would require data containing a complete marital history. Second, the presence of differences in the prevalence of marriage by education is observed (Goldstein & Kenney, 2001). Overall, the prevalence of marriage declines over cohorts. However, the decrease is larger for the least educated individuals. College graduates have a higher prevalence of marriage as compared to individuals with lower levels of educational attainment. The least educated individuals have both the fastest drop in marriage prevalence and the quickest increase in divorce. This result is consistent with previous studies reporting that the decline in marriage prevalence has been steeper in the most disadvantaged groups (Cherlin, 2004, 2010; Goldstein & Kenney, 2001; McLanahan, 2004). Likewise, the number of people divorced and never-married diverges between low and highly educated individuals. Cohabitation also increases over cohorts, as it is increasingly more prevalent for high school or less educated individuals. For example, Table 4 highlights differences in the age-standardised prevalence of marriage between less than high school and college-educated. For non-Hispanic Whites, it is around 5 percentage points in the 1937 birth cohort and 17 percentage points in the 1952 cohort. Similar patterns are observable for the prevalence of divorce over cohorts: 2.4 percentage point difference for the 1937 birth cohort and 13 percentage points for the 1952 cohort. Moreover, 5.3 per cent of non-Hispanic White men with less than high school education in the 1952 birth cohort were reported to cohabit as compared to less than 1 per cent in the 1937 cohort. Concerning non-Hispanic Black individuals, overall the prevalence of divorce across educational groups does not change substantially over cohorts; however, there is a sharp decrease in marriage prevalence, especially for non-Hispanic Black men, and a symmetrically marked increase in the never-married category.

Striking differences in the distribution of marital status are observable by race and ethnicity. In particular, the prevalence of widows among women is remarkably higher among non-Hispanic Blacks; across the educational distribution, non-Hispanic Blacks have ten percentage point higher prevalence rate of being windowed compared to non-

Hispanic Whites. Moreover, marriage prevalence exhibits pronounced differences between non-Hispanic Whites and Blacks. For example, the prevalence of less than high school-educated women in the 1937 birth cohort being married is 72.4 per cent among non-Hispanic Whites and 39.6 per cent among non-Hispanic Blacks. Therefore, there is more than a 30 percentage point difference between the two racial groups.

Focusing on the age-specific prevalence rates in Tables 5 and 6, we observe in the youngest age groups (30–40 and 40–50 years old) the presence of a pronounced educational gradient in marriage and divorce in 1967, but little or no difference among educational groups in the 1937 birth cohort. Similarly, the prevalence of never-married individuals has increased substantially in the 1967 birth cohort. For example, for less than high school-educated non-Hispanic Black men marriage prevalence in the 30–40 age group is less than half in the 1967 birth cohort (30 per cent) compared to the 1937 cohort (67 per cent). In conclusion, there is an increasing differential between the least and the highest educated individuals in terms of marriage, divorce and cohabitation prevalence over cohorts. Therefore, the emergence of an educational gradient in the prevalence of marriage, cohabitation and divorce over cohorts as educational expansion progresses is evident, but the availability of a full marital history would enable further confirmation of these findings.

Table 4 Prevalence of Different Marital Categories Over 1937 and 1952 Cohorts (Non-Hispanic White and Black, Percentages)

	Married		Divorced		Widow		Never-married		Cohabitation	
	1937	1952	1937	1952	1937	1952	1937	1952	1937	1952
Non-Hispanic White Men										
Less than high school	80.8	62.6	11.0	22.9	2.2	1.7	5.9	12.4	0.8	5.3
High school	84.6	72.5	9.5	17.3	1.6	1.2	4.1	8.6	0.6	4.4
Some college	84.3	73.3	10.3	16.2	1.6	1.0	3.8	9.1	0.6	3.7
College	85.3	79.7	8.6	9.9	1.0	0.7	4.9	9.4	0.5	2.4
Non-Hispanic White Women										
Less than high school	72.4	61.2	14.3	24.0	10.1	7.4	3.0	7.1	0.6	5.2
High school	78.3	72.7	12.1	18.5	6.7	3.7	2.7	4.6	0.4	3.4
Some college	74.1	69.4	15.8	21.2	6.7	3.3	3.1	5.8	0.5	3.4
College	73.6	72.8	14.4	14.5	4.4	2.7	7.4	9.5	0.5	2.3
Non-Hispanic Black Men										
Less than high school	63.4	43.5	19.8	23.3	4.5	3.5	12.1	29.3	1.1	5.6
High school	71.4	55.1	18.2	22.7	3.9	2.2	6.3	18.9	1.4	6.1
Some college	72.4	63.0	19.3	19.6	2.6	2.2	5.0	14.3	1.2	4.0
College	73.7	67.6	17.6	18.1	1.6	1.3	6.9	12.5	0.5	3.0
Non-Hispanic Black Women										
Less than high school	39.6	29.2	31.6	30.8	16.5	10.3	12.0	29.4	0.7	3.5
High school	49.9	40.7	27.1	29.4	12.9	6.5	9.1	22.3	0.3	3.6
Some college	46.5	41.8	32.8	32.4	12.8	6.5	7.5	18.1	0.6	3.5
College	54.2	52.3	25.3	26.6	10.4	3.7	9.4	16.1	0.3	1.6

Notes: All these rates are age-standardised based on the 1937 female non-Hispanic White age distribution. *Source:* Author's calculation based on NHIS data.

Table 5 Prevalence of Different Marital Categories Over 1937 and 1967 Cohorts by Age Groups (Non-Hispanic White, Percentages)

Age group	Education	Married		Divorced		Widow		Never-married		Cohabitation	
		1937	1967	1937	1967	1937	1967	1937	1967	1937	1967
Non-Hispanic White Men											
30-40	Less than high school	83.2	55.2	9.4	18.8	0.1	0.3	7.3	24.9	-	15.0
	High school	87.4	63.0	7.2	13.9	0.4	0.2	5.0	22.0	-	10.5
	Some college	86.4	67.5	7.8	11.6	0.3	0.1	5.5	20.3	-	8.0
	College	86.2	73.7	6.7	4.7	0.3	0.1	6.9	21.0	-	5.5
40-50	Less than high school	84.8	53.1	8.6	22.7	0.6	1.3	5.9	22.5	-	11.6
	High school	86.7	62.0	8.5	20.0	0.4	0.7	4.2	16.8	-	9.3
	Some college	86.0	68.6	9.9	16.1	0.4	0.7	3.5	14.3	-	7.2
	College	86.7	80.9	7.9	8.0	0.4	0.2	4.9	10.6	-	4.5
50-60	Less than high school	79.7	-	13.0	-	1.6	-	5.6	-	0.3	-
	High school	84.1	-	10.2	-	1.5	-	3.9	-	0.5	-
	Some college	84.3	-	10.3	-	1.4	-	3.8	-	0.5	-
	College	85.3	-	9.1	-	1.0	-	4.5	-	0.3	-
60-70	Less than high school	74.5	-	12.9	-	7.1	-	5.3	-	3.1	-
	High school	80.2	-	11.4	-	4.5	-	3.4	-	2.3	-
	Some college	80.2	-	12.2	-	4.3	-	3.2	-	2.3	-
	College	82.6	-	10.0	-	2.6	-	4.3	-	2.0	-
Non-Hispanic White Women											
30-40	Less than high school	81.3	59.4	13.8	25.7	1.8	0.9	3.1	13.7	-	13.5
	High school	86.2	68.7	9.3	17.0	1.1	0.7	3.4	12.8	-	8.4
	Some college	83.0	70.6	11.7	16.6	1.3	0.4	3.9	11.8	-	7.3
	College	80.7	75.4	9.8	7.8	0.4	0.3	9.2	15.9	-	4.8
40-50	Less than high school	79.2	50.8	13.8	32.7	3.4	1.1	3.5	15.4	-	14.3
	High school	82.3	64.0	12.0	23.4	2.7	1.6	2.9	10.6	-	10.0
	Some college	78.1	67.8	15.2	21.8	3.2	1.3	3.2	8.7	-	7.3
	College	76.9	78.0	13.3	12.1	1.7	0.7	8.0	9.0	-	4.9
50-60	Less than high school	72.0	-	15.1	-	9.8	-	2.9	-	0.3	-

	High school	78.0	-	13.0	-	6.3	-	2.4	-	0.3	-
	Some college	74.2	-	17.0	-	5.7	-	2.9	-	0.5	-
	College	72.1	-	16.5	-	4.6	-	6.5	-	0.4	-
60-70	Less than high school	56.8	-	14.5	-	26.3	-	2.2	-	2.1	-
	High school	67.5	-	12.4	-	17.3	-	2.3	-	1.6	-
	Some college	62.2	-	17.5	-	17.3	-	2.7	-	1.5	-
	College	66.4	-	15.7	-	10.7	-	6.6	-	1.6	-

Source: Author's calculation based on NHIS data.

Table 6 Prevalence of Different Marital Categories Over 1937 and 1967 Cohorts by Age Groups (Non-Hispanic Black, Percentages)

Age group	**Education**	**Married**		**Divorced**		**Widow**		**Never-married**		**Cohabitation**	
		1937	**1967**	**1937**	**1967**	**1937**	**1967**	**1967**	**1937**	**1967**	**1937**
Non-Hispanic Black Men											
30-40	Less than high school	67.0	30.5	17.5	13.8	1.8	0.7	13.7	53.9	-	18.2
	High school	74.7	44.9	16.5	15.2	2.4	0.1	6.4	38.8	-	13.3
	Some college	82.0	51.7	13.0	14.6	1.0	0.1	4.0	32.8	-	12.0
	College	67.1	58.7	15.8	8.4	-	0.5	17.1	30.9	-	6.9
40-50	Less than high school	67.9	31.1	17.9	23.5	1.8	0.5	12.2	44.8	-	13.1
	High school	74.0	48.3	17.1	18.4	1.8	0.9	6.9	32.0	-	10.5
	Some college	75.8	57.6	16.8	19.3	1.0	0.5	6.1	22.0	-	10.1
	College	74.9	66.3	17.4	14.0	1.2	0.7	6.5	18.1	-	4.9
50-60	Less than high school	63.8	-	21.0	-	4.2	-	11.0	-	1.1	-
	High school	71.2	-	18.4	-	3.9	-	6.3	-	1.1	-
	Some college	69.9	-	22.9	-	2.0	-	4.2	-	1.5	-
	College	74.5	-	19.4	-	1.8	-	4.3	-	0.7	-
60-70	Less than high school	53.4	-	22.3	-	11.0	-	12.5	-	3.8	-
	High school	65.3	-	20.8	-	8.1	-	5.1	-	5.1	-

	Some college	65.0	-	22.1	-	6.9	-	4.9	-	3.4	-
	College	74.5	-	16.3	-	2.9	-	5.4	-	1.3	-
Non-Hispanic Black Women											
30-40	Less than high school	48.1	24.1	34.4	19.2	4.6	1.1	12.8	54.1	-	8.3
	High school	56.1	34.3	28.1	18.3	6.0	0.8	9.8	44.7	-	8.9
	Some college	58.3	38.7	31.3	21.4	4.2	0.9	6.3	37.2	-	6.4
	College	61.2	47.8	17.6	14.3	4.7	0.7	16.5	36.1	-	4.5
40-50	Less than high school	44.4	27.4	35.0	23.7	9.3	2.2	11.4	45.6	-	9.6
	High school	54.4	33.1	28.1	22.8	8.6	2.1	8.5	40.6	-	6.3
	Some college	50.7	39.7	35.1	26.7	5.8	1.5	8.4	31.3	-	5.0
	College	58.7	50.0	27.5	21.5	4.6	0.9	9.2	26.0	-	4.9
50-60	Less than high school	38.3	-	30.8	-	18.4	-	12.2	-	0.8	-
	High school	49.8	-	28.9	-	11.6	-	8.9	-	0.2	-
	Some college	46.1	-	32.9	-	12.8	-	7.8	-	1.2	-
	College	53.6	-	27.6	-	10.6	-	7.5	-	0.5	-
60-70	Less than high school	28.5	-	25.7	-	32.5	-	12.3	-	1.9	-
	High school	39.1	-	22.2	-	25.8	-	10.0	-	1.1	-
	Some college	33.4	-	29.7	-	29.5	-	6.4	-	1.2	-
	College	43.6	-	22.5	-	22.8	-	8.7	-	0.7	-

Source: Author’s calculation based on NHIS data.

3.4 Results

3.4.1 Heterogeneity in marriage-health by education

Tables 7 and 8 and Figure 2 and 3 present the results from the logit regression analyses of the association between marital status and poor/fair SRH within different educational groups. Overall, divorced and never-married have worse SRH than married regardless of educational attainment. For never-married women, this disadvantage seems to be true only among those with high school or some college education.

Over cohorts, married men are less likely to report poor/fair SRH within the college graduates group, but they are also more likely to report poor/fair SRH within less than college educational groups. Similar results were obtained for women, but the improvement over cohorts in the highest educated group was not significant. Moreover, there was an increase in the probability of reporting poor/fair SRH for divorced, never-married and widowed in almost all educational groups for both genders. Therefore, the health advantage of married college-educated individuals increases over cohorts among men and remains stable for women. Remarkably, married individuals are better off over cohorts only within the college-educated group. Married women with high school or less education have even worse SRH over cohorts. For example, in the less than high school-educated group the predictive probability of poor/fair SRH for married men goes from around 20 per cent for the 1937 birth cohort to almost 30 per cent in the 1967 cohort.

To summarise, this analysis suggests that within the least educated groups, there is no improvement, and even deterioration, over cohorts in SRH for married people as compared to the other marital categories. But it also suggests that married college graduates (especially men) experience improved SRH over cohorts. Therefore, the association between marriage and health over cohorts seems to strengthen within the college-educated group. These results propose that over cohorts, education is a reinforcing mechanism for the association between marriage and health. Being married, however, is not associated with improvements in health over cohorts in non-college-educated groups: over cohorts, marriage does not confer an increasing health advantage to the least educated groups. Married college-educated individuals emerge over time as a double advantaged group, providing support to the resource multiplication framework.

Trends for divorced, widowed and never-married also differ depending on the educational group under consideration. In general, widowed and never-married are more likely to report poor/fair SRH in all educational categories over cohorts. Divorced women seem to have worse SRH in all educational groups, except within the college degree group. Divorced men, on the other hand, are more likely to report poor/fair SRH over cohorts only within the some college and college educational groups. These results suggest that being divorced has increasingly negative health consequences over cohorts within different educational groups. For women, this health disadvantage seems to be concentrated among the least educated, while for men, individuals with some or more college education are more likely to report a SRH disadvantage over cohorts.

As expected, age is associated with worse health. The odds of poor/fair SRH are between 6.4 (for men with some college) and 4.3 times (for women with college or more) higher for each additional year of age. Moreover, concerning regional differences in SRH, we observed a greater likelihood of a poor/fair SRH in the Southern, Midwestern and Western regions as compared to the Northeast one. In particular, the Southern region emerges as the most disadvantaged region in terms of SRH, followed by the Midwest and West. For example, less than high school-educated individuals living in the South are 88 times more likely to have poor/fair SRH than people in the Northeast. Differences by region persist across the educational spectrum. However, the relative difference among regions is lower within the college-educated group as compared to the less than high school-educated group.

The robustness of these results is tested in different ways. This analysis is replicated using an alternative dichotomous SRH indicator (i.e. *poor* SRH or otherwise). The results are qualitatively similar to the ones presented (see Table A1 and Table A2 in the Appendix). The analysis is also conducted including cohabitation using only data from 1997, the results of which are available in the Appendix in Tables A3 and A4. The results are approximately similar to the results reported when excluding cohabitation, as coefficients for cohabitation and its interaction with cohort are not significant for either men or women, which is also probably due to the small size of the cohabitating sample.

Table 7 Odds Ratios for Cohort Trends by Marital Status in Poor/Fair SRH from Logistic Regression Models within Various Educational Groups for Non-Hispanic White Men (Age 30–70, Cohort 1937–1967)

	Less than high school		High school		Some college		College or more	
	OR	SE	OR	SE	OR	SE	OR	SE
Marital status (ref. married)								
Divorced	1.584***	(0.144)	1.698***	(0.127)	1.354**	(0.128)	1.516**	(0.193)
Widow	0.719	(0.183)	0.570**	(0.115)	0.805	(0.230)	1.146	(0.575)
Never-married	1.678***	(0.188)	1.588***	(0.156)	1.507***	(0.187)	1.370*	(0.203)
Age	1.059***	(0.00218)	1.060***	(0.00139)	1.064***	(0.00185)	1.056***	(0.00227)
Cohort	1.038**	(0.0140)	1.022*	(0.00957)	1.029*	(0.0121)	0.948***	(0.0138)
Cohort*Divorced	0.999	(0.0194)	1.015	(0.0158)	1.070***	(0.0210)	1.079**	(0.0309)
Cohort*Never-married	1.006	(0.0226)	1.044*	(0.0184)	1.057*	(0.0241)	1.083**	(0.0322)
Cohort*Widow	1.189*	(0.0803)	1.291***	(0.0629)	1.157*	(0.0781)	1.167	(0.125)
Region (ref. Northeast)								
Midwest	1.310***	(0.0623)	1.133***	(0.0342)	1.154**	(0.0554)	1.045	(0.0620)
South	1.882***	(0.0878)	1.573***	(0.0457)	1.450***	(0.0629)	1.191**	(0.0645)
West	1.143*	(0.0657)	1.190***	(0.0431)	1.244***	(0.0564)	1.183**	(0.0655)
Recoded	1.275***	(0.0660)	1.317***	(0.0545)	1.556***	(0.102)	1.585***	(0.121)
N Observations	42,222		143,395		102,776		134,679	

Notes: All values are weighted. *p < .05. **p < .01. ***p < .001. *Source:* Author's calculation based on NHIS data.

Table 8 Regression Coefficients for Cohort Trends by Marital Status in Poor/Fair SRH from Logistic Regression Models within Various Educational Groups for Non-Hispanic Women (Age 30–70, Cohort 1937–1967)

	Less than high school		High school		Some college		College or more	
	OR	SE	OR	SE	OR	SE	OR	SE
Marital status (ref. married)								
Divorced	1.495***	(0.122)	1.536***	(0.0927)	1.421***	(0.112)	1.463**	(0.185)
Widow	0.786	(0.104)	0.813*	(0.0827)	0.820	(0.120)	1227	(0.320)
Never-married	1.102	(0.162)	1.386**	(0.151)	1.846***	(0.265)	1.201	(0.186)
Age	1.049***	(0.00197)	1.052***	(0.00134)	1.050***	(0.00168)	1.043***	(0.00226)
Cohort	1.057***	(0.0133)	1.075***	(0.00914)	1.071***	(0.0120)	0.983	(0.0151)
Cohort*Divorced	1.042*	(0.0187)	1.038**	(0.0138)	1.074***	(0.0167)	1.049	(0.0273)
Cohort*Never-married	1.065*	(0.0305)	1.069**	(0.0219)	0.999	(0.0256)	1.088**	(0.0324)
Cohort*Widow	1.227***	(0.0454)	1.196***	(0.0312)	1.212***	(0.0415)	1.135	(0.0733)
Region (ref. Northeast)								
Midwest	1.365***	(0.0617)	1.149***	(0.0344)	1.101*	(0.0479)	1.142*	(0.0616)
South	1.740***	(0.0745)	1.529***	(0.0429)	1.349***	(0.0514)	1.318***	(0.0683)
West	1.189***	(0.0620)	1.268***	(0.0431)	1.276***	(0.0532)	1.300***	(0.0672)
Recoded	1.345***	(0.0665)	1.592***	(0.0565)	1.543***	(0.0826)	1.503***	(0.108)
N Observations	40,932		168,663		117,794		119,132	

Notes: All values are weighted. *p < .05. **p < .01. ***p < .001. *Source:* Author’s calculation based on NHIS data.

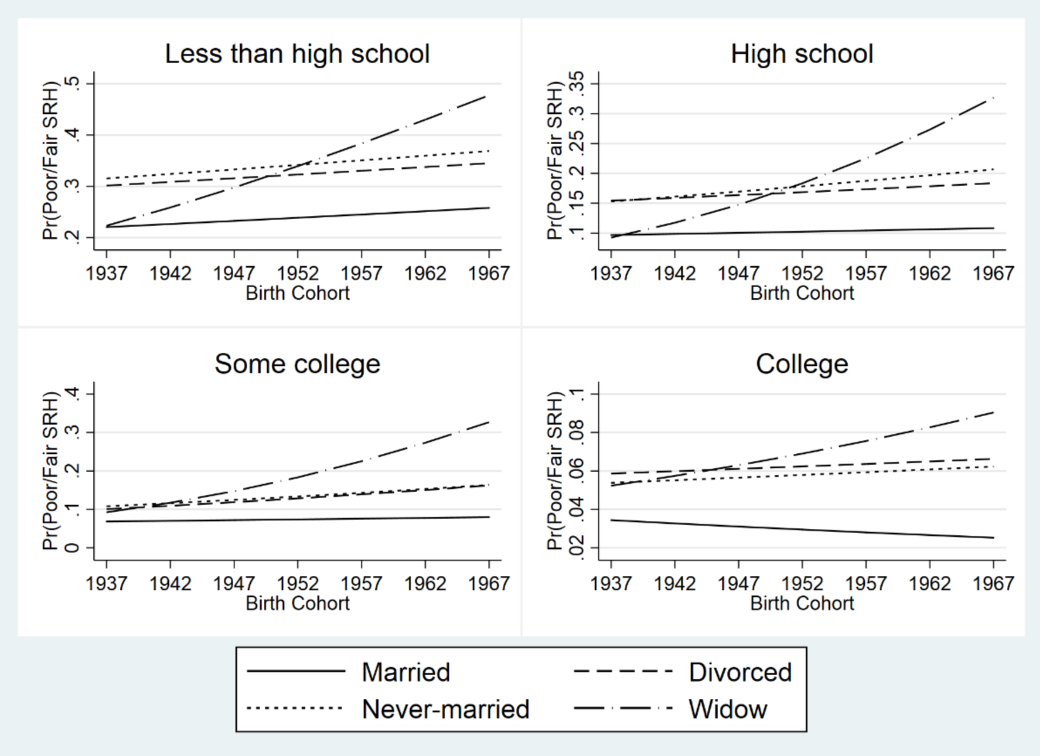

Figure 2 Predicted probability of poor/fair SRH by educational group (Men 30–70, cohorts 1937–1967, Non-Hispanic White).

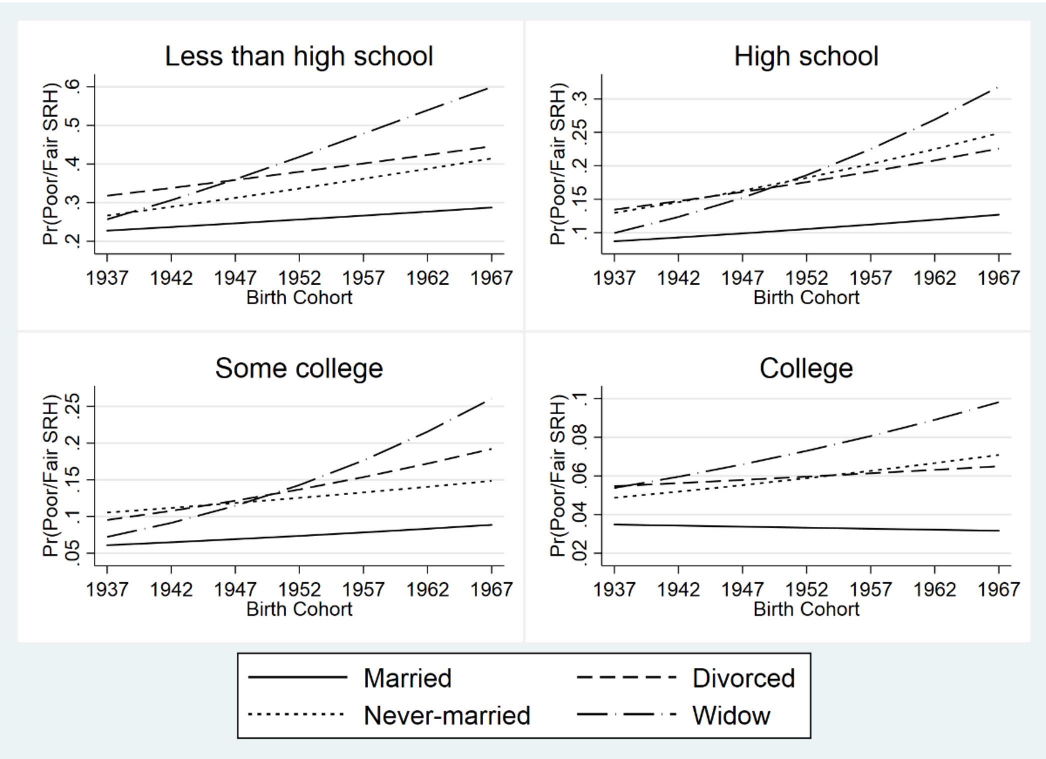

Figure 3 Predicted probability of poor/fair SRH by educational group (Women 30–70, cohorts 1937–1967, Non-Hispanic White).

3.4.2 Results using non-linear decomposition

The second objective of the paper is to relate the differences in SRH between educational groups to changes in the distribution of marital status within different education groups. The Fairlie non-linear decomposition technique is employed to identify the contribution of marital status distribution to the gap in SRH between low and highly educated respondents. In particular, we focus on the dichotomy between married and unmarried. The contribution of marriage to the SRH gap is calculated looking at the change in the average predicted probabilities if it is replaced the less educated distribution with the highly educated distribution, keeping all other variables constant. Fairlie decomposition focuses on the part of educational gap in SRH explained by group differences in the distribution of marriage. However, the part of the gap due to unexplained differences in unobservable endowments is not easily interpretable with this method; therefore, we perform this gap analysis using Oaxaca-Blinder linear decomposition.

The difference in the predicted probabilities of poor/fair SRH increases across age groups and birth cohorts (Table 9). Figure 4 plots the percentage of the SRH gap among high school or less and college-educated explained by marriage over cohorts. The distribution of marriage explains increasingly the differences between high school graduates or less and the college-educated in SRH. This seems to be especially true for individuals above the age of 40. For women in the early cohorts, marriage explains the gap negatively, which means that low educated women would have had a higher predicted probability of poor/fair SRH if they had the distribution of the highly educated for that variable. However, the sign reverses among more recent cohorts, such that less educated women would have a lower predicted probability of poor/fair SRH if they had the distribution of highly educated. High school or less educated women emerge as an increasingly disadvantaged group over cohorts since their marital distribution initially contributes to a narrowing of the SRH gap and subsequently widens the difference. The trend towards a reduction in the prevalence of marriage in the less educated group has a noticeable association with SRH differentials over cohorts. Despite the described steady increase in the importance of marital distribution for SRH, marriage is only able to explain a small part of the SRH differential (slightly more than 10% at maximum). Therefore, the magnitude of the contribution of marital status for the SRH differences remains small.

In conclusion, differences in the prevalence of marriage are increasingly able to explain the SRH gap by educational attainment over cohorts; nonetheless, they still represent only a small contribution to the overall SRH gap (up to a maximum of 10 per cent of the gap is explained). These results suggest a sharp increase in the relevance of marital composition for the SRH gap between the two educational groups. In the 40–50 age group, the largest increase in the proportion of the difference explained by marriage is observable. One possible interpretation for this result might be the greater importance of marital status for health in midlife. Similar results were obtained when using the dichotomous specification of SRH (1 if *poor* SRH, 0 otherwise). While the magnitude of the change in the explained proportion seems to be slightly higher, the results are qualitatively similar (Figure A3 in Appendix). The results replicate also using only the two most extreme educational groups: less than high school and college-educated. Despite slight changes in magnitude, the trend is in line with previous analyses (Figure A4 in Appendix).

Table 9 Difference in Average Predicted Probabilities between High School or Less and College-Educated for Poor/Fair Health by Cohort and Age Group (Non-Hispanic White)

	Age Groups			
	30–40	40–50	50–60	60–70
	Men			
1937		0.093	0.145	0.172
1942	0.078	0.100	0.169	0.205
1947	0.066	0.100	0.155	0.198
1952	0.053	0.097	0.166	0.205
1957	0.061	0.101	0.169	
1962	0.061	0.115		
1967	0.061			
	Women			
1937		0.095	0.125	0.1419
1942	0.075	0.093	0.142	0.1635
1947	0.067	0.088	0.148	0.1808
1952	0.062	0.095	0.153	0.2134
1957	0.060	0.104	0.167	
1962	0.062	0.129		
1967	0.087			

Source: Author's calculation based on NHIS data.

Figure 4 Percentage of SRH gap explained by marriage over cohorts in different age groups for men and women.

3.4.3 Other marital categories

The decomposition results indicate that the changing proportion of married people in the low and highly educated groups is strictly associated with increasing SRH inequality. This section explores the role of other marital categories to explain these increasing SRH differences. The analysis is replicated to include indicators for being divorced, widowed and have never-married. The full results from this additional analysis are presented in Figures 5 and 6. The SRH difference between high school or less and college-educated is increasingly explained over cohorts by the prevalence of divorce and never-married, particularly in the youngest age groups, whereas widowhood seems to play a role only in the oldest age group. The distribution of divorce by education seems to explain to a greater extent SRH differences over cohorts. The differential prevalence of never-married individuals in the two educational groups also explains part of the gap, but to a lesser degree as compared to divorce. Therefore, the role of family instability for the educational gap in SRH emerges as prominent.

Figure 5 Percentage of SRH gap explained by divorce, widowhood and never-married over cohorts in different age groups for men.

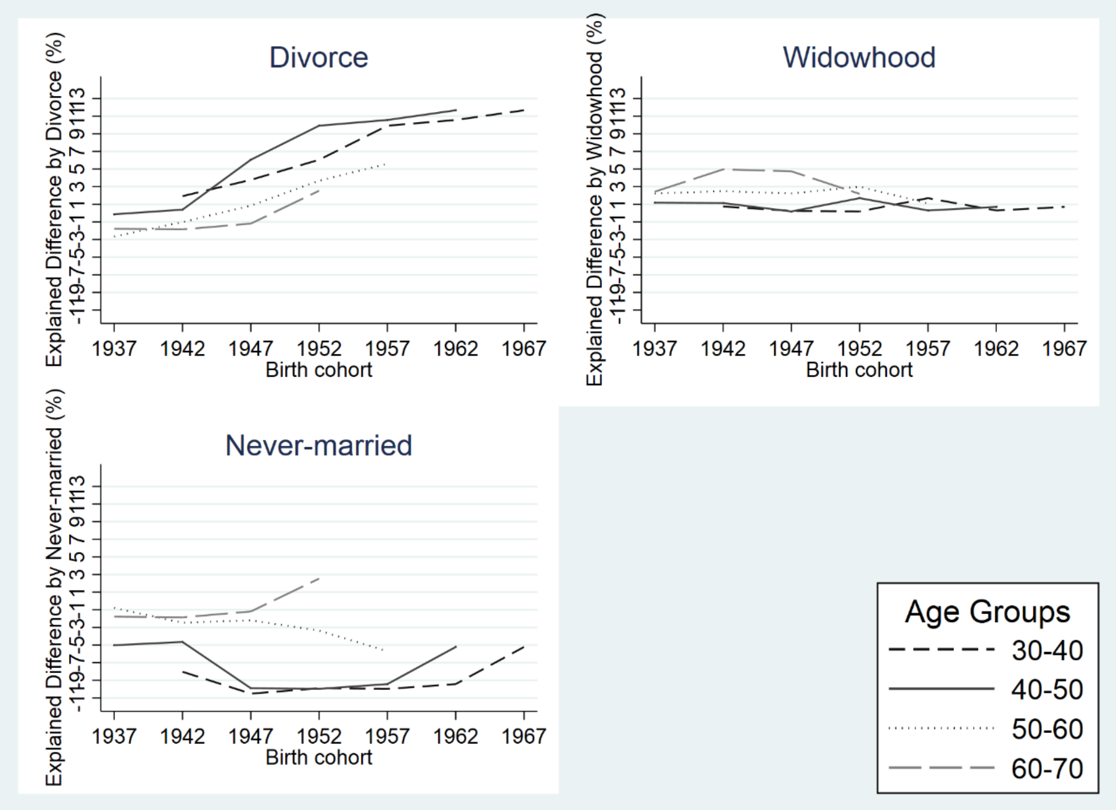

Figure 6 Percentage of health gap explained by divorce, widowhood and never-married over cohorts in different age groups for men.

3.4.4 Explained and unexplained component: linear decomposition

In order to explore the explained and unexplained decomposition components simultaneously, a 4-point SRH score is used and the differences among high school or less and college-educated is linearly decomposed. Tables 10 and 11 present the decomposition results over different cohorts, distinguishing between the explained and unexplained portions of the educational SRH differences. The decomposition procedure was run separately for men and women, and the role of different components specified. First, the two tables present the total gross difference in terms of SRH between high school or less and college-educated. Second, they show the contribution of different components to the explained and unexplained part. Finally, they present the unexplained part of the difference associated with the constant, which is the remaining unidentified quota after controlling for all covariates.

Table 10 Components of the Educational SRH Gap for Different Cohorts and Age Groups (Men)

		Explained				Unexplained				
	Difference	Married	Region	Age	Subtotal	Married	Region	Age	Constant	Subtotal
Age 30–40										
1937	0.321	0.000	0.007	0.001	0.008	-0.111	0.022	0.121	0.281	0.313
1942	0.317	-0.002	0.007	0.000	0.004	-0.095	0.039	0.515	-0.147	0.313
1947	0.283	-0.003	0.006	0.002	0.005	-0.097	0.033	0.256	0.087	0.278
1952	0.253	0.001	0.002	-0.004	-0.001	-0.101	0.024	0.280	0.051	0.254
1957	0.270	0.005	0.004	-0.003	0.005	-0.082	0.040	0.235	0.072	0.265
1962	0.268	0.011	0.006	-0.005	0.012	-0.056	0.052	0.445	-0.185	0.256
1967	0.279	0.021	0.004	-0.003	0.022	-0.057	0.025	0.400	-0.111	0.257
Age 40–50										
1937	0.348	0.001	0.009	0.001	0.012	-0.106	0.040	0.008	0.394	0.336
1942	0.360	0.001	0.006	-0.002	0.005	-0.070	0.095	0.090	0.239	0.355
1947	0.368	0.007	0.005	-0.003	0.009	-0.082	0.119	0.554	-0.232	0.359
1952	0.364	0.023	0.008	0.000	0.031	-0.097	0.031	0.205	0.194	0.333
1957	0.371	0.034	0.007	-0.003	0.038	-0.101	0.053	1.056	-0.676	0.333
1962	0.427	0.044	0.007	-0.001	0.050	-0.115	0.021	0.799	-0.328	0.377
Age 50–60										
1937	0.467	0.005	0.003	-0.005	0.004	-0.143	0.109	0.488	0.009	0.463

1942	0.502	0.013	0.010	-0.006	0.017	-0.087	0.091	0.729	-0.248	0.485
1947	0.500	0.024	0.012	-0.003	0.033	-0.161	0.124	0.920	-0.416	0.467
1952	0.521	0.036	0.007	-0.004	0.039	-0.135	0.078	1.102	-0.564	0.482
1957	0.535	0.045	0.010	-0.001	0.053	-0.156	0.125	0.240	0.274	0.482
Age 60–70										
1937	0.499	0.006	0.004	0.000	0.010	-0.019	0.079	-0.240	0.668	0.489
1942	0.584	0.013	0.012	0.000	0.025	-0.024	0.066	-0.389	0.904	0.559
1947	0.583	0.020	0.007	-0.001	0.027	-0.140	0.093	-0.100	0.703	0.556
1952	0.607	0.032	0.018	-0.004	0.046	-0.102	-0.120	-5.494	6.277	0.561

Source: Author’s calculation based on NHIS data.

Table 11 Components of the Educational SRH Gap for Different Cohorts and Age Groups (Women)

		Explained				Unexplained				
	Difference	Married	Region	Age	Subtotal	Married	Region	Age	Constant	Subtotal
Age 30–40										
1937	0.338	-0.004	0.003	0.000	-0.001	0.014	0.025	0.260	0.040	0.339
1942	0.302	-0.007	0.005	0.000	-0.003	-0.050	0.042	0.192	0.121	0.305
1947	0.292	-0.007	0.004	0.005	0.002	-0.054	0.034	0.024	0.286	0.290
1952	0.275	-0.004	0.005	-0.004	-0.004	-0.073	0.039	0.223	0.090	0.279
1957	0.277	0.000	0.003	-0.004	-0.001	-0.077	0.056	0.130	0.168	0.278
1962	0.298	0.005	0.008	-0.003	0.009	-0.061	0.054	0.093	0.202	0.289
1967	0.350	0.019	0.003	-0.004	0.018	-0.122	0.036	0.564	-0.146	0.332
Age 40–50										
1937	0.360	-0.006	0.011	0.002	0.007	-0.078	0.017	0.204	0.210	0.353
1942	0.346	-0.005	0.006	-0.003	-0.002	-0.085	0.070	0.291	0.073	0.348
1947	0.341	-0.005	0.008	-0.004	-0.001	-0.062	0.063	0.589	-0.247	0.342
1952	0.356	0.003	0.004	-0.005	0.002	-0.096	0.065	0.481	-0.096	0.354
1957	0.381	0.012	0.006	-0.003	0.015	-0.138	0.069	0.586	-0.151	0.366
1962	0.428	0.029	0.007	-0.003	0.034	-0.158	-0.004	0.930	-0.374	0.395
Age 50–60										

1937	0.432	-0.007	0.013	-0.002	0.005	-0.124	0.050	0.616	-0.115	0.427
1942	0.452	-0.006	0.011	-0.005	0.000	-0.087	0.033	0.644	-0.137	0.452
1947	0.482	-0.003	0.012	-0.002	0.008	-0.100	0.082	1.038	-0.546	0.474
1952	0.467	0.009	0.010	-0.003	0.015	-0.144	0.109	0.874	-0.388	0.451
1957	0.515	0.011	0.010	0.000	0.021	-0.161	0.042	0.482	0.131	0.495
Age 60–70										
1937	0.446	0.002	0.007	0.000	0.009	-0.024	0.041	-0.317	0.737	0.437
1942	0.495	0.002	0.001	0.002	0.006	-0.060	0.132	-0.443	0.861	0.490
1947	0.526	0.003	0.021	0.000	0.024	-0.129	0.101	-1.031	1.560	0.502
1952	0.618	0.013	0.006	-0.001	0.017	-0.047	0.069	-5.523	6.101	0.601

Source: Author’s calculation based on NHIS data.

The SRH difference between those at the bottom and at the top of the educational distribution increased over birth cohorts for males and females in all age groups (with the exception of men in the younger age group). This increase appears to have been gradual over cohorts and slightly steeper for men than for women. Looking at the subtotal for the explained and unexplained parts, it appears that most of the gap remained unexplained. Simultaneously, the explained and unexplained subtotals increased over cohorts. However, the increase appeared to be steeper in the explained portion. Similar results were evident in all age groups, except for the youngest group, where the total of the unexplained portion did not increase.

Figure 7 plots the SRH health gap explained by being married after controlling for age and region. Being married seems to play an increasingly important role in explaining health differences over cohorts in all age groups. In the youngest birth cohorts, marriage explains SRH disparities more than in the oldest cohorts. For example, educational differences in SRH for individuals aged 40–50 years born in 1962 are explained by marriage prevalence almost ten times more than for those born in 1937. Moreover, it is found that the effect was stronger for men than for women. For women, being married accounts less for the educational differences in SRH than for men. Therefore, using the linear decomposition, the importance of being married for educational inequality in SRH emerges as well.

Figure 7 Percentage of SRH gap explained by marriage over cohorts in different age groups for men and women using linear decomposition (non-Hispanic White).

The unexplained portion of the gap increases over cohorts and age groups, and is larger for women. The coefficient of the unexplained portion associated with being married serves as an indicator of the differential returns of marriage on health for the two educational groups while keeping the percentage of married individuals at the highly educated level fixed. In particular, the negative coefficient suggests that the benefits for health associated with being married are greater in the less educated group compared to the highly educated group. Therefore, the prevalence of negative coefficients on the unexplained portion associated with marriage suggests differential returns from marriage between the less and the highly educated, in which the less educated are favoured, contributing to the narrowing of the SRH gap.

To summarise, the Oaxaca-Blinder decomposition presented suggests that marriage has a role in explaining rising educational inequalities in SRH through two channels:

first, through the increasing difference in distribution of married individuals across educational group, and second, through the differentials in returns on health from marriage in the two educational groups. These two aspects go in two different directions: the diverging proportion of married individuals in the two groups over cohorts widens the SRH gap, while the differing returns from marriage favouring less educated married people contribute towards its reduction. Therefore, marriage appears to be beneficial for reducing the educational SRH gap because the returns from being married are higher in the less educated group, but the decreasing prevalence over time is associated with the widening gap over time. However, there is no consistent pattern of change over cohorts in the unexplained component (with the exception of women, who seem to benefit increasingly from marriage in the younger age groups), while there is a clear increase in the explained portion by marriage across age and sex. Similar results were obtained when analysing the gap for the two most extreme educational groups (results available upon request).

3.4.5 Cohabitation

Marriage has become less common, especially among less educated individuals, while cohabitation has become more widespread (Bumpass & Lu, 2000; Cherlin, 2010; Kennedy & Bumpass, 2008). However, the rise in cohabitation has not occurred evenly across the educational spectrum: the least educated are increasingly the most likely to ever cohabit (Bumpass & Lu, 2000). This alternative form of partnership is included in a separate analysis because data on cohabitation has only been available since 1997. Therefore, the consistency of previous results was checked in two different ways. First, the trends in the portion explained by marriage, including cohabitation, for the available survey years are recalculated (Figure 8). Secondly, short trends for the cohorts and age groups for which data was available are estimated (Figure 9). As expected, the role of marriage prevalence seems to attenuate following the inclusion of cohabitation. This is consistent with the literature showing that while cohabitation has a beneficial effect on health, this effect is less significant than the one of marriage (Wu, Penning, Pollard, & Hart, 2003). Similarly, from the analysis of short trends, it appears that cohabitation plays an attenuation role in explaining educational SRH inequality.

Figure 8 Percentage of SRH gap explained by marriage and cohabitation over cohorts in different age groups for women and men.

Figure 9 Percentage of SRH gap explained by cohabitation over cohorts in different age groups for women and men.

3.4.6 Non-Hispanic Blacks

This section analyses the role of marriage for the educational gap in SRH for the non-Hispanic Black group. Blacks have a higher degree of family instability as well as lower likelihood of marriage compared to Whites (Manning & Cohen, 2012; Raley et al., 2015). Many explanations have been proposed for the gap in marriage formation and stability between Blacks and Whites, such as the reduced availability of single Black men due to mass incarceration, decreasing economic prospects and unemployment. Numerous interrelated factors contribute to the emergence of these differences; therefore, it is difficult to isolate the cause of the higher prevalence of marital instability and lower proportion of ever-married individuals among Blacks. However, socioeconomic status represents an increasingly important factor for marriage formation and marital stability. Furthermore, race and socioeconomic disadvantage are closely related. Therefore, socioeconomic conditions may be among the driving forces behind the Black–White gap in marriage. Also within the Black group, marriage and marital stability rates vary depending on the position in the educational distribution (Raley et al., 2015). Despite the presence of differences between Blacks and Whites across the whole educational distribution, the largest differences have been observed within the least educated group. Over time, the educational gap in health and disability has widened among Blacks and Whites, but the difference between educational groups is larger among Whites (Liu & Hummer, 2008; L. G. Martin et al., 2007). Concerning SRH differences, previous studies have found a lower predicted probability of good health among Blacks as well as different trends by marital status for the two groups (Liu & Umberson, 2008). In particular, married Blacks experience a more dramatic improvement in SRH compared to Whites over time. In addition, divorced Blacks seem to have better SRH over time, unlike Whites.

The results from the logit models for different educational groups are presented in Tables 12 and 13. Divorced individuals have a higher probability of poor/fair SRH compared to married people across the educational spectrum. Never-married individuals have a higher likelihood of poor/fair SRH, but only in the least educated group. Moreover, married non-Hispanic Blacks are less likely to report poor/fair SRH at all levels of the educational distribution over cohorts. Small differences are present between males and females: the positive trend for married individuals over time holds significantly in all educational groups except for the least educated men and the highest educated women. No other significant trends over cohorts are evident across other marital categories.

The decomposition of the educational gap in SRH within non-Hispanic Blacks produces similar results as in the non-Hispanic White sample. The health gap is increasingly explained by the marital distribution of low and highly educated. Despite the clear differences in the meaning of marriage in the two race groups, the different distribution of marriage within the least and highest educated is increasingly important for explaining the educational gap in SRH. Marriage is still considered desirable in the Black community, but both marriage and marital stability have been declining at an increasing rate in the past century for the least educated (Raley et al., 2015). There are smaller gender differences in the explanatory role of marriage among non-Hispanic Blacks than among non-Hispanic Whites. Overall, the distribution of marriage seems to be equally important for men and women in explaining the educational gap in SRH among Blacks.

Table 12 Odds Ratios for Cohort Trends by Marital Status in Poor/Fair Health from Logistic Regression Models within Various Educational Groups for Men for Non-Hispanic Blacks (Age 30–70, Cohort 1937–1967)

	Less than high school		High school		Some college		College or more	
	OR	SE	OR	SE	OR	SE	OR	SE
Marital status (ref. married)								
Divorced	1.347*	0.182	1.383*	0.179	1.551*	0.278	1.391	0.379
Widow	1.348	0.471	1.220	0.362	0.617	0.302	0.065	0.103
Never-married	1.961***	0.281	2.100***	0.330	1.575*	0.365	1.093	0.444
Age	1.046***	0.003	1.054***	0.002	1.057***	0.003	1.055***	0.006
Cohort	0.977	0.019	0.964*	0.016	0.963	0.021	0.902**	0.034
Cohort*Divorced	1.027	0.033	1.045	0.027	1.000	0.037	1.010	0.060
Cohort*Never-married	0.973	0.030	0.987	0.028	1.023	0.043	1.082	0.083
Cohort*Widow	1.051	0.097	1.131	0.081	1.187	0.138	1.940	0.663
Region (ref. Northeast)								
Midwest	1.489***	0.125	1.347***	0.083	1.683***	0.143	1.237	0.188
South	1.518***	0.108	1.393***	0.073	1.511***	0.114	1.038	0.139
West	1.367*	0.183	1.197*	0.091	1.710***	0.151	1.210	0.201
N	12,948		27,127		16,683		10,731	

Notes: All values are weighted. *p < .05. **p < .01. ***p < .001. *Source:* Author's calculation based on NHIS data.

Table 13 Odds Ratios for Cohort Trends by Marital Status in Poor/Fair Health from Logistic Regression Models within Various Educational Groups for Women for Non-Hispanic Black (Age 30–70, Cohort 1937–1967)

	Less than high school		High school		Some college		College or more	
	OR	SE	OR	SE	OR	SE	OR	SE
Marital status (ref. married)								
Divorced	1.297*	0.138	1.191	0.119	1.046	0.148	1.620*	0.361
Widow	1.112	0.193	0.793	0.125	0.589*	0.126	1.589	0.615
Never-married	1.462**	0.199	1.215	0.149	1.175	0.207	1.307	0.371
Age	1.030***	0.002	1.041***	0.002	1.044***	0.003	1.042***	0.004
Cohort	0.947**	0.018	0.930***	0.014	0.954*	0.019	0.952	0.030
Cohort*Divorced	1.011	0.025	1.042	0.023	1.060*	0.029	0.973	0.044
Cohort*Never-married	0.981	0.027	1.049*	0.025	1.045	0.031	1.015	0.051
Cohort*Widow	1.046	0.050	1.146***	0.045	1.231***	0.059	1.030	0.097
Region (ref. Northeast)								
Midwest	1.383***	0.077	1.240***	0.062	1.295***	0.082	0.996	0.109
South	1.362***	0.058	1.202***	0.052	1.056	0.062	0.918	0.081
West	1.359***	0.113	1.349***	0.080	1.226**	0.089	1.078	0.130
N	17,141		35,766		24,991		14,691	

Notes: All values are weighted. *p < .05. **p < .01. ***p < .001. *Source:* Author's calculation based on NHIS data.

Figure 10 Percentage of health gap explained by marriage over cohorts in different age groups for men and women (non-Hispanic Black).

3.5 Summary and Discussion

Rising educational inequality in health is a largely debated phenomenon. Many factors have been proposed with which to explain the increasing differences in health among low and highly educated individuals. One element that has not been explored in the literature is the changing marital status composition among different educational groups. The distribution of different marital categories across educational groups has changed greatly over cohorts. This factor has had a large impact on the prevalence of marriage across the educational spectrum. Using repeated cross-sectional data from the NHIS (1976–2014), the increasing educational inequality in SRH over cohorts is decomposed. This work is able to explain part of the increasing health differences over cohorts between high school or less and college-educated individuals with the changing prevalence of married individuals within each group. Nonetheless, the results of this study explain only a small percentage of the overall gap.

First, this study examines how the observed marriage–health relationship within various educational groups changed over cohorts using a series of logit models. The main results of this investigation suggest that married individuals are better off in terms of SRH, but only within the college-educated group. On the contrary, divorced, never-married and widows are worse off over cohorts in most educational groups. Therefore, married individuals with a college education generally emerge over cohorts as being increasingly advantaged with regards to health. Dissimilarly, the returns from marriage within the least educated groups does not seem to rise over cohorts. Consequently, married, college-educated individuals are able to maintain and enhance their SRH advantage over cohorts. Regional differences in SRH also appear evident; in particular, people living in the Southern states are more likely to report poor/fair SRH. Other studies have documented these regional differences in the United States reflect true health difference and not only dissimilarities in reporting (Avendano & Kawachi, 2014; Berchick & Lynch, 2017).

Second, it is considered how the rising inequalities in health can be explained by the changing prevalence of marriage based on education. Despite this explanation having been proposed in the literature, it has not been tested empirically. In particular, SRH differences among high school or less and college-educated individuals are decomposed using the Fairlie method, calculating the proportion attributable to marriage in different cohorts. The results show a sharp increase over cohorts in the proportion of the SRH educational gap explained by marriage distribution, but the magnitude of the explained difference remains small despite the sharp rise. Therefore, the analysis suggests that the lower proportion of married individuals in the least educated group compared to the highest educated group make an increasingly greater contribution to SRH inequalities. In addition, the returns from marriage are higher in the least educated groups compared to the most educated group. Consequently, marriage returns appear beneficial for mitigating SRH inequalities by education. In this study, the significance of marriage for the SRH differentials seems to be especially evident in the younger age group (40–50 years old). The health advantage for married individuals is attenuated at an older age, which is consistent with previous research (Kallan, 1997; Rendall et al., 2011; Wood et al., 2007). However, it is important to note that these age differences might also be due to mortality selection issues that we were unable to investigate in the present study. Moreover, trends in the difference explained by marriage are comparable among men and women, but slightly stronger for men. Previous studies have found mixed evidence for the protective effects of marriage by gender (Manzoli, Villari, M Pirone, & Boccia, 2007; Zick & Smith, 1991), but more recent research has found a higher mortality advantage for men compared to women, which is consistent with the results from this analysis suggesting a greater explanatory ability of marriage for males (Rendall et al., 2011).

In conclusion, we found evidence for an increasing role of marriage in explaining educational disparities in SRH. The proportion of married individuals is increasingly greater in the college-educated group compared to the high school or less group. The changing proportion of married individuals in the two educational groups contributes to growing health inequalities. Given the results presented, it is important not to overlook the changing distribution of marital characteristics within various educational groups when the association between education and health is studied over time. These results also suggest that when studying cohort trends, it is important to fully consider the changing demographic composition of the groups of interest. This represents a broader issue to be applied to different puzzles in social sciences (Horowitz, 2015).

Thus, these results make at least two contributions to the literature. First, they suggest that the resource multiplication theory might be the most appropriate for describing the health benefits from marriage over cohorts for both the least and the highest educated. For the high school or less educated, the SRH advantages connected with marriage seem to be greatest in the oldest cohorts as lower levels of education for earlier cohorts were linked to socioeconomic disadvantage to a lesser extent than for more recent cohorts (Ruggles, 2015). Similarly, among the college-educated, the health benefits from marriage are more significant in among more recent cohorts where education is more strongly linked to several advantages in the labour market and income gains (Sum et al., 2011). Second, this analysis presents evidence for the importance of studying changes in demographic characteristics linked to education that are related to health trends over time. Similar to the cohort replacement framework, it is suggested that demographic aspects connected to different educational groups shape cohort profiles of health (Mannheim, 1952; Ryder, 1965). Therefore, differential cohort-specific demographic behaviours, such as partnership formation and stability, are essential elements to be studied in order to gain a better understanding of the increasing educational inequalities in health and mortality. Therefore, the highest educated are those able to maintain and enhance their health advantage over cohorts, particularly through demographic behaviours like marriage.

Despite this analysis being based only on United States data, these findings might be useful for understanding similar trends in other developed countries. A positive educational gradient in marriage and a negative association between education and divorce has been observed in some European countries (De Graaf & Kalmijn, 2006; Härkönen & Dronkers, 2006; Kalmijn, 2013). Many gender-egalitarian countries in Europe have reported a similar marriage gradient to that observed in the United States

(Kalmijn, 2013; Perelli-Harris & Lyons-Amos, 2016). However, Perelli-Harris and Amos (2016) highlight the importance of country of residence in explaining this relationship, finding weaker evidence for these associations in Europe compared to the United States. Therefore, despite United States does not appear very dissimilar to some of the countries in Europe in the association between education and partnership, large variation across countries in Europe is present. Therefore, it remains unclear to what extent current findings could be generalized for all European countries. Moreover, poor SRH among the less educated is much more prevalent in the United States compared to European countries (Präg & Subramanian, 2017); therefore, the results we found might, to a less extent, be valid in the European context.

One limitation for this study is the absence of data on cohabitation for the most of the study periods as this information was not collected in the NHIS until 1997. Some studies have analysed cohabitation in relation to SRH, psychological distress and mortality (S. L. Brown, 2000; Fuller, 2010; Wu et al., 2003). Most of these studies have established that cohabiters are in a better situation in terms of health and mortality compared to those who had never-married or to all other non-married individuals, but are worse off compared to married individuals (Liu & Reczek, 2012). Moreover, cohabitation appears to be a more short-lived phase, which is in accordance with the *trial marriage* view of cohabitation (Smock, 2000). However, additional analyses on the available sample have shown the robustness of our results to the inclusion of cohabitation.

A second limitation is that despite marriage having a protective effect on health (Hu & Goldman, 1990; Rendall et al., 2011), and education having been linked to favourable health outcomes (Mirowsky & Ross, 2003), these positive associations are not always informative with respect to causal effects. On the one hand, few studies have succeeded in disentangling the causal from the selection effects for the relationship between marriage and health (Lillard & Panis, 1996; Murray, 2000). On the other side, the causal effect of education on health has been analysed in several studies using such techniques as instrumental variables or twin fixed-effects designed to isolate causality (Baker, Leon, Smith Greenaway, Collins, & Movit, 2011; Behrman et al., 2011; Glied & Lleras-Muney, 2008; Lundborg, Lyttkens, & Nystedt, 2016). However, trying to isolate causal effects when analysing trends over time is especially complex. Finally, this analysis cannot take into account the changing patterns of assortative mating over time. For example, the rise in educational hypogamy might contribute to the increase in the health gap between the low and the highly educated. However, other studies have shown that changing patterns of educational assortative mating do not contribute significantly to the development of income inequality (Breen & Salazar, 2011).

Third, recent work by Kalmijn (2017) highlights the complexity of studying the health benefits from marriage. Kalmijn (2017) finds that the effect of divorce on health is much more damaging than the health benefits gained from entering marriage, which suggests a leading role for divorce in the marital status differences for health. Likewise, Williams and Umberson (2004) have suggested that marital advantage is likely to more closely reflect patterns of marital dissolution. Moreover, the role of marital biographies has been shown to contribute to health differences, since marital benefits accrue with time spent married (Dupre & Meadows, 2007; Mcfarland, Hayward, & Brown, 2013; Ploubidis et al., 2015). However, we are unable to take into account marital history, for example, by distinguishing first and second order marriages due to divorce and widowhood. Therefore, a third limitation of this study is the analysis of SRH differences focusing on marital status during a single point in time.

Moreover, the choice of using SRH to describe health differences by education over time is not free from limitations. In the current study, well-being in terms of health is measured using SRH, which refers to a more physical than psychological dimension of health. However, the effects of divorce on health are sensitive to the health outcomes being considered. As such, in this case, more negative effects are observed in relation to mental health as opposed to the physical dimensions of health. Therefore, we can hypothesise stronger patterns observed if using mental health as an outcome as compared to SRH. In addition, there might be differences in the reporting of SRH by socioeconomic status as shown by Dowd and Zajacova (2010). For example, SRH differences by education might be underestimated or overestimated. However, there is no reason to believe that the SRH gap might be biased differently over time; therefore, if some bias is present, it is likely to stay constant over time, thus not affecting the conclusions with respect to trends over time. Moreover, the NHIS excludes institutionalised civilians, which might lead to some bias, especially if there are changes in the demographic characteristics of institutionalised people over time. However, given that we are restricting the sample to those who are between 30 and 70 years old, no considerable bias should be present for non-Hispanic Whites. With respect to non-Hispanic Black men, however, incarceration trends and premature mortality might have affected those people who were interviewed (Pettit & Western, 2004).

Finally, we focus on the likelihood of marriage only by education, while Tumin and Zheng (2018) have highlighted the multidimensional socioeconomic aspects of the likelihood of marriage, which we were unable to capture in this analysis. However, the focus of this work on education is motivated by the desire to investigate an important piece of the puzzle with respect to rising educational inequalities in health over time (Goldman & Smith, 2011). Future research should aim to explore the health gap over time while taking into account the changing likelihood of marriage by socioeconomic status using multiple indicators of social disadvantage.

In conclusion, this analysis highlights the need for greater attention to the changing demographic characteristics within different educational categories when looking at trends over time. It also opens up some questions for future research. Has the educational expansion really produced a more selective and homogeneous group at the bottom? In what sense are those at the bottom of the educational distribution selected? Future research should focus on identifying the underlying processes that produce these changes in demographic characteristics and health outcomes in order to advance our knowledge of the consequences of family change for health inequality.

References

Amato, P. R., Booth, A., Johnson, D. R., & Rogers, S. J. (2007). *Alone Together: How Marriage in America is Changing*. Cambridge, Mass.: Harvard University Press.

Aseltine, R. H. ., & Kessler, R. C. . (1993). Marital disruption and depression in a community sample. *Journal of Health and Social Behavior*, *34*(3), 237–251.

Avendano, M., Huijts, T., & Subramanian, S. V. (2009). Re: "Are Americans feeling less healthy? the puzzle of trends in self-rated health." *American Journal of Epidemiology*, *170*(12), 1581–1582.

Avendano, M., & Kawachi, I. (2014). Why do Americans have shorter life expectancy and worse health than do people in other high-income countries? *Annual Review of Public Health*, *35*(1), 307–325.

Baeten, S., Van Ourti, T., & van Doorslaer, E. (2013). The socioeconomic health gradient across the life cycle: what role for selective mortality and institutionalization? *Social Science and Medicine*, *97*, 66–74.

Baker, D. P., Leon, J., Smith Greenaway, E. G., Collins, J., & Movit, M. (2011). The education effect on population health: a reassessment. *Population and Development Review*, *37*(2), 307–332.

Barber, J., Yarger, J., & Gatny, H. (2015). Black-white differences in attitudes related to pregnancy. *Demography*, *52*, 751–786.

Becker, G. (1981). *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Behrman, J. R., Kohler, H. P., Jensen, V. M., Pedersen, D., Petersen, I., Bingley, P., & Christensen, K. (2011). Does more schooling reduce hospitalization and delay mortality? New evidence based on Danish twins. *Demography*, *48*(4), 1347–1375.

Benjamins, M. R., Hummer, R. A., Eberstein, I. W., & Nam, C. B. (2004). Self-reported health and adult mortality risk: an analysis of cause-specific mortality. *Social Science & Medicine*, *59*(6), 1297–1306.

Berchick, E. R., & Lynch, S. M. (2017). Regional variation in the predictive validity of self-rated health for mortality. *SSM - Population Health*, *3*(January), 275–282.

Binswanger, I. A., Krueger, P. M., & Steiner, J. F. (2009). Prevalence of chronic medical conditions among jail and prison inmates in the USA compared with the general population. *Journal of Epidemiology and Community Health*, *63*(11), 912–919.

Blomgren, J., Martikainen, P., Grundy, E., & Koskinen, S. (2012). Marital history 1971-91 and mortality 1991-2004 in England & Wales and Finland. *Journal of Epidemiology and Community Health*, *66*(1), 30–36.

Blossfeld, H.-P. (2009). Educational assortative marriage in comparative perspective. *Annual Review of Sociology*, *35*(1), 513–530.

Bramlett, M. D., & Mosher, W. D. (2002). *Cohabitation, marriage, divorce, and remarriage in the United States* (Series 23 No. 22). *Vital Health Statistics*.

Breen, R., & Salazar, L. (2011). Educational assortative mating and earnings inequality in the United States. *American Journal of Sociology*, *117*(3), 808–843.

Brown, D. C., Hummer, R. A., & Hayward, M. D. (2014). The importance of spousal education for the self-rated health of married adults in the United States. *Population Research and Policy Review*, *33*(1), 127–151.

Brown, S. L. (2000). The effect of union type on psychological well-being: depression among cohabitors versus marrieds. *Journal of Health and Social Behavior*, *41*(3), 241–255.

Bryant, C. M., & Conger, R. D. (1999). Marital success and domains of social support in long-term relationships: does the influence of network members ever end? *Journal of Marriage and the Family*, *61*(2), 437–450.

Bumpass, L., & Lu, H. (2000). Trends in cohabitation and implications for children's family contexts in the United States. *Population Studies*, *54*(1), 29–41.

Burgard, S. A., & Chen, P. V. (2014). Challenges of health measurement in studies of health disparities. *Social Science and Medicine*, *106*, 143–150.

Cherlin, A. J. (1992). *Marriage, Divorce, Remarriage*. Cambridge, Mass.: Harvard University Press.

Cherlin, A. J. (2004). The deinstitutionalization of American marriage. *Journal of Marriage and Family*, *66*(4), 848–861.

Cherlin, A. J. (2009). *The Marriage-Go-Round: The State of Marriage and the Family in America Today* (1st ed.). New York: Vintage Books.

Cherlin, A. J. (2010). Demographic trends in the United States: a review of research in the 2000s. *Journal of Marriage and Family*, *72*(3), 403–419.

Cohen, P. N. (2015). *The Family: Diversity, Inequality, and Social Change* (First). New York: W.W. Norton & Company.

Crimmins, E. M., & Saito, Y. (2001). Trends in healthy life expectancy in the United States, 1970-1990: gender, racial, and educational differences. *Social Science and Medicine*, *52*(11), 1629–1641.

Crimmins, E. M., Saito, Y., & Ingegneri, D. (1997). Trends in disability-free life expectancy in the United States, 1970-90. *Population and Development Review*, *23*(3), 555–572.

Davey Smith, G., & Lynch, J. (2004). Life course approaches to socioeconomic differentials in health. In D. Kuh, Y. Ben Shlomo, & S. Ezra (Eds.), *A Life Course Approach to Chronic Disease Epidemiology* (Vol. 6, pp. 77–115). Oxford

University Press.

De Graaf, P. M., & Kalmijn, M. (2006). Change and stability in the social determinants of divorce: a comparison of marriage cohorts in the Netherlands. *European Sociological Review*, *22*(5), 561–572.

Dowd, J. B., & Hamoudi, A. (2014). Is life expectancy really falling for groups of low socio-economic status? Lagged selection bias and artefactual trends in mortality. *International Journal of Epidemiology*, *43*(4), 983–988.

Dowd, J. B., & Todd, M. (2011). Does self-reported health bias the measurement of health inequalities in U.S. adults? Evidence using anchoring vignettes from the Health and Retirement Study. *Journal of Gerontology*, *66*, 478–489.

Dowd, J. B., & Zajacova, A. (2007). Does the predictive power of self-rated health for subsequent mortality risk vary by socioeconomic status in the US? *International Journal of Epidemiology*, *36*(6), 1214–1221.

Dowd, J. B., & Zajacova, A. (2010). Does self-rated health mean the same thing across socioeconomic groups? Evidence from biomarker data. *Annals of Epidemiology*, *20*(10), 743–749.

Dupre, M. E., & Meadows, S. O. (2007). Disaggregating the effects of marital trajectories on health. *Journal of Family Issues*, *28*(5), 623–652.

Elo, I. T., & Preston, S. H. (1996). Educational differentials in mortality: United States, 1979-85. *Social Science Medicine*, *42*(1), 47–57.

Esteve, A., García-román, J., & Permanyer, I. (2012). The gender-gap reversal in education and its effect on union formation: the end of hypergamy? *Population and Development Review*, *38*(3), 535–546.

Esteve, A., Schwartz, C. R., van Bavel, J., Permanyer, I., Klesment, M., & García-Román, J. (2016). The end of hypergamy: global trends and implications.

Population and Development Review, *42*(4), 615–625.

Fairlie, R. W. (1999). The absence of the african-american owned business: an analysis of the dynamics of self-employment. *Journal of Labor Economics*, *17*(1), 80–108.

Farmer, M. M., & Ferraro, K. F. (1997). Distress and perceived health: mechanisms of health decline. *Journal of Health and Social Behavior*, *38*(3), 298–311.

Fuller, T. D. (2010). Relationship status, health, and health behavior: an examination of cohabiters and commuters. *Sociological Perspectives*, *53*(2), 221–246.

Glied, S., & Lleras-Muney, A. (2008). Technological innovation and inequality in health. *Demography*, *45*(3), 741–761.

Goesling, B. (2007). The rising significance of education for health? *Social Forces*, *85*(4), 1621–1644.

Goldman, D., & Smith, J. P. (2011). The increasing value of education to health. *Social Science & Medicine*, *72*(10), 1728–1737.

Goldstein, J. R., & Kenney, C. T. (2001). Marriage delayed or marriage forgone? New cohort forecasts of first marriage for U.S. women. *American Sociological Review*, *66*(4), 506–519.

Grundy, E., & Tomassini, C. (2010). Marital history, health and mortality among older men and women in England and Wales. *BMC Public Health*, *10*, 554.

Härkönen, J., & Dronkers, J. (2006). Stability and change in the educational gradient of divorce. A comparison of seventeen countries. *European Sociological Review*, *22*(5), 501–517.

Hayford, S. R. (2013). Marriage (still) matters: the contribution of demographic change to trends in childlessness in the United States. *Demography*, *50*, 1641–1661.

Hays, J. C., Schoenfeld, D., Blazer, D. G., & Gold, D. T. (1996). Global self-ratings of health and mortality: hazard in the North Carolina Piedmont. *Journal of Clinical Epidemiology*, *49*(9), 969–979.

Hayward, M. D., Hummer, R. A., & Sasson, I. (2015). Trends and group differences in the association between educational attainment and U.S. adult mortality: implications for understanding education's causal influence. *Social Science and Medicine*, *127*, 8–18.

Hendi, A. S. (2015). Trends in U.S. life expectancy gradients: the role of changing educational composition. *International Journal of Epidemiology*, *44*(3), 946–955.

Ho, J. Y., & Fenelon, A. (2015). The contribution of smoking to educational gradients in U.S. life expectancy. *Journal of Health and Social Behavior*, *56*(3), 307–322.

Horowitz, J. (2015). Doing less with more: cohorts, education, and civic participation in America. *Social Forces*, *94*(2), 747–774.

Hou, F., & Myles, J. (2008). The changing role of education in the marriage market: assortative marriage in Canada and the United States since the 1970s. *The Canadian Journal of Sociology / Cahiers Canadiens de Sociologie*, *33*(2), 337–366.

Hu, Y. R., & Goldman, N. (1990). Mortality differentials by marital status: an international comparison. *Demography*, *27*(2), 233–250.

Hughes, M. E., & Waite, L. J. (2009). Marital biography and health at mid-life. *Journal of Health and Social Behavior*, *50*(3), 344–358.

Huijts, T., Monden, C. W. S., & Kraaykamp, G. (2010). Education, educational heterogamy, and self-assessed health in Europe. A multilevel study of spousal effects in 29 European countries. *European Sociological Review*, *26*(3), 261–276.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior, 38*(1), 21–37.

Idler, E. L., Russell, L. B., & Davis, D. (2000). Survival, functional limitations, and self-rated health in the NHANES I. *American Journal of Epidemiology, 152*(9), 874–883.

Johnson, D. R., & Wu, J. (2002). An empirical test of crisis, social selection, and role explanations of the relationship between marital disruption and psychological distress: a pooled time-series analysis of four-wave panel data. *Journal of Marriage and Family, 64*(1), 211–224.

Joung, I. M. A., Van De Mheen, H. D., Stronks, K., van Poppel, F., & Mackenbach, J. P. (1998). A longitudinal study of health selection in marital transitions. *Social Science & Medicine, 46*(3), 425–435.

Kallan, J. (1997). Effects of sociodemographic variables on adult mortality in the United States: comparisons by sex, age, and cause of death. *Biodemography and Social Biology, 44*(1–2), 136–147.

Kalmijn, M. (2013). The educational gradient in marriage: a comparison of 25 European countries. *Demography, 50*(4), 1499–1520.

Kalmijn, M. (2017). The ambiguous link between marriage and health: a dynamic reanalysis of loss and gain effects. *Social Forces, 95*(4), 1607–1636.

Keenan, K., Ploubidis, G. B., Silverwood, R. J., & Grundy, E. (2017). Life-course partnership history and midlife health behaviours in a population-based birth cohort. *Journal of Epidemiology and Community Health, 71*(3), 232–238.

Kennedy, S., & Bumpass, L. (2008). Cohabitation and children's living arrangements. *Demographic Research, 19*, 1663–1692.

Kennedy, S., & Ruggles, S. (2014). Breaking up is hard to count: the rise of divorce in the United States, 1980-2010. *Demography*, *51*(2), 587–598.

Keyes, K. M., Rutherford, C., Popham, F., Martins, S. S., & Gray, L. (2018). How healthy are survey respondents compared with the general population? Using survey-linked death records to compare mortality outcomes. *Epidemiology*, *29*(2), 299–307.

Kravdal, Ø. (2017). Large and growing social inequality in mortality in Norway: The combined importance of marital status and own and spouse's education. *Population and Development Review*, *43*(4), 645–665.

Lauderdale, D. S. (2001). Education and survival: birth cohort, period, and age effects. *Demography*, *38*(4), 551–561.

Lauer, S., & Yodanis, C. (2010). The deinstitutionalization of marriage revisited: a new institutional approach to marriage. *Journal of Family Theory & Review*, *2*(1), 58–72.

Lemieux, T. (2008). The changing nature of wage inequality. *Journal of Population Economics*, *21*(1), 21–48.

Liang, J., Brown, J. W., Krause, N. M., Ofstedal, M. B., & Bennett, J. (2005). Health and living arrangements of older Americans: Does marriage matter? *Journal of Aging and Health*, *17*(3), 305–335.

Lillard, L. A., & Panis, C. W. A. (1996). Marital status and mortality: the role of health. *Demography*, *33*(3), 313–327.

Liu, H., & Hummer, R. A. (2008). Are educational differences in U.S. self-rated health increasing?: An examination by gender and race. *Social Science and Medicine*, *67*(11), 1898–1906.

Liu, H., & Reczek, C. (2012). Cohabitation and U.S. adult mortality: an examination by gender and race. *Journal of Marriage and Family*, *74*(4), 794–811.

Liu, H., & Umberson, D. J. (2008). The times they are a changin': marital status and health differentials from 1972 to 2003. *Journal of Health and Social Behavior*, *49*(3), 239–253.

Lorenz, F. O., Wickrama, K. A., Conger, R. D., & Elder Jr., G. H. (2006). The short-term and decade-long effects of divorce on women's midlife health. *Journal of Health and Social Behavior*, *47*(2), 111–125.

Lundborg, P., Lyttkens, C. H., & Nystedt, P. (2016). The effect of schooling on mortality: new evidence from 50,000 Swedish twins. *Demography*, *53*(4), 1135–1168.

Lynch, S. M. (2003). Cohort and life-course patterns in the relationship between education and health: a hierarchical approach. *Demography*, *40*(2), 309–331.

Mandel, H., & Semyonov, M. (2014). Gender pay gap and employment sector: sources of earnings disparities in the United States, 1970–2010. *Demography*, *51*(5), 1597–1618.

Mandel, H., & Semyonov, M. (2016). Gender differences in trends and sources of the racial pay gap, 1970-2010. *American Sociological Review*, *81*(5), 11039–106838.

Mannheim, K. (1952). *Essays on the Sociology of Knowledge*. London: Routledge & Kegan Paul.

Manning, W. D., & Cohen, J. A. (2012). Premarital cohabitation and marital dissolution: an examination of recent marriages. *Journal of Marriage and Family*, *74*(2), 377–387.

Manzoli, L., Villari, P., M Pirone, G., & Boccia, A. (2007). Marital status and mortality in the elderly: a systematic review and meta-analysis. *Social Science & Medicine*, *64*(1), 77–94.

Martikainen, P., Martelin, T., Nihtilä, E., Majamaa, K., & Koskinen, S. (2005).

Differences in mortality by marital status in Finland from 1976 to 2000: analyses of changes in marital-status distributions, socio-demographic and household composition, and cause of death. *Population Studies*, *59*(1), 99–115.

Martin, L. G., Schoeni, R. F., Freedman, V. A., & Andreski, P. (2007). Feeling better? Trends in general health status. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *62*(1), S11–S21.

Martin, S. P. (2006). Trends in marital dissolution by women's education in the United States. *Demographic Research*, *15*.

Masters, R. K., Hummer, R. A., & Powers, D. A. (2012). Educational differences in U.S. adult mortality: a cohort perspective. *American Sociological Review*, *77*(4), 548–572.

Maurer, J. (2011). Education and male-female differences in later-life cognition: international evidence from Latin America and the Caribbean. *Demography*, *48*(3), 915–930.

Mcfarland, M. J., Hayward, M. D., & Brown, D. C. (2013). I've got you under my skin: marital biography and biological risk. *Journal of Marriage and Family*, *75*(2), 363–380.

McLanahan, S. (2004). Diverging destinies: how children are faring under the second demographic transition. *Demography*, *41*(4), 607–627.

McLanahan, S., & Percheski, C. (2008). Family structure and the reproduction of inequalities. *Annual Review of Sociology*, *34*(1), 257–276.

McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social isolation in America: changes in core discussion networks over two decades. *American Sociological Review*, *73*(6), 353–375.

Meadows, S. O., Mclanahan, S., & Brooks-Gunn, J. (2008). Family structure changes and maternal health trajectories. *American Sociological Review*, *73*, 314:334.

Meara, E. R., Richards, S., & Cutler, D. M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981-2000. *Health Affairs*, *27*(2), 350–360.

Miech, R., Pampel, F., Kim, J., & Rogers, R. G. (2011). The enduring association between education and mortality: the role of widening and narrowing disparities. *American Sociological Review*, *76*(6), 913–934.

Mirowsky, J., & Ross, C. E. (2003). *Education, Social Status and Health*. New York, NY, USA: Aldine de Gruyter.

Monden, C. W. S., Van Lenthe, F., De Graaf, N. D., & Kraaykamp, G. (2003). Partner's and own education: does who you live with matter for self-assessed health, smoking and excessive alcohol consumption? *Social Science and Medicine, 57*(10), 1901–1912.

Montez, J. K., & Friedman, E. M. (2015). Educational attainment and adult health: under what conditions is the association causal? *Social Science and Medicine, 127*, 1–7.

Montez, J. K., Hummer, R. a., Hayward, M. D., Woo, H., & Rogers, R. G. (2011). Trends in the educational gradient of U.S. adult mortality from 1986 through 2006 by race, gender, and age group. *Research on Aging*, *33*(2), 145–171.

Montez, J. K., & Zajacova, A. (2013). Explaining the widening education gap in mortality among U.S. white women. *Journal of Health and Social Behavior*, *54*(2), 166–182.

Murphy, M., Glaser, K., & Grundy, E. (1997). Marital status and long-term illness in Great Britain. *Journal of Marriage and the Family*, *59*(1), 156–164.

Murray, J. E. (2000). Marital protection and marital selection: evidence from a historical-prospective sample of American men. *Demography*, *37*(4), 511–521.

Olshansky, S. J., Antonucci, T., Berkman, L., Binstock, R. H., Boersch-Supan, A., Cacioppo, J. T., … Rowe, J. (2012). Differences in life expectancy due to race and educational differences are widening, and many may not catch up. *Health Affairs*, *31*(8), 1803–1813.

Oppenheimer, V. K. (1997). Women's employment and the gains to marriage: the specialization and training model of marriage. *Annual Review of Sociology*, *23*, 431–453.

Pampel, F. C., & Peters, H. E. (1995). The Easterlin effect. *Annual Review of Sociology*, *21*, 163–194.

Perelli-Harris, B., & Lyons-Amos, M. (2016). Partnership patterns in the united states and across europe: The role of education and country context. *Social Forces*, *95*(1), 251–282.

Pettit, B., & Western, B. (2004). Mass imprisonment and the life course: race and class inequality in U.S. incarceration. *American Sociological Review*, *69*, 151–169.

Pew Research Center. (2010). Millenials: A Portrait of Generation Next. *Statistics*, (February), 140.

Phillips, J. A., & Sweeney, M. M. (2005). Premarital cohabitation and marital disruption among White, Black, and Mexican American women. *Journal of Marriage and Family*, *67*(2), 296–314.

Ploubidis, G. B., Silverwood, R. J., DeStavola, B., & Grundy, E. (2015). Life-course partnership status and biomarkers in midlife: evidence from the 1958 British birth cohort. *American Journal of Public Health*, *105*(8), 1596–1603.

Präg, P., & Subramanian, S. V. (2017). Educational inequalities in self-rated health across US states and European countries. *International Journal of Public Health*, *62*(6), 709–716.

Raley, K. R., & Bumpass, L. (2003). The topography of the divorce plateau: levels and trends in union stability in the United States after 1980. *Demographic Research*, *8*, 245–259.

Raley, K. R., Sweeney, M. M., & Wondra, D. (2015). The growing racial and ethnic divide in U.S. marriage patterns. *Future of Children*, *25*(2), 89–109.

Rendall, M. S., Weden, M. M., Favreault, M. M., & Waldron, H. (2011). The protective effect of marriage for survival: a review and update. *Demography*, *48*(2), 481–506.

Riley, M. W. (1987). On the significance of age in sociology. *American Sociological Review*, *52*(1), 1.

Ross, C. E., & Mirowsky, J. (2006). Sex differences in the effect of education on depression: resource multiplication or resource substitution? *Social Science and Medicine*, *63*(5), 1400–1413.

Ross, C. E., & Mirowsky, J. (2010). Gender and the health benefits. *The Sociological Quarterly*, *51*(1), 1–19.

Ruggles, S. (2015). Patriarchy, power, and pay: the transformation of American families, 1800–2015. *Demography*, *52*(6), 1797–1823.

Ryder, N. B. (1965). The cohort as a concept in the study of social change. *American Sociological Review*, *30*(6), 843–861.

Sasson, I. (2016). Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010. *Demography*, 1–25.

Schwartz, C. R. (2013). Trends and variation in assortative mating: causes and consequences. *Annual Review of Sociology*, *39*(1), 451–470.

Schwartz, C. R., & Gonalons-Pons, P. (2016). Trends in relative earnings and marital dissolution: are wives who outearn their husbands still more likely to divorce? *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(4), 218–236.

Schwartz, C. R., & Han, H. (2014). The reversal of the gender gap in education and trends in marital dissolution. *American Sociological Review*, *79*(4), 605–629.

Schwartz, C. R., & Mare, R. D. (2005). Trends in educational assortative marriage from 1940 to 2003. *Demography*, *42*(4), 621–646.

Sen, B. (2014). Using the oaxaca-blinder decomposition as an empirical tool to analyze racial disparities in obesity. *Obesity*, *22*(7), 1750–1755.

Smock, P. J. (2000). Cohabitation in the United States: an appraisal of research themes, findings, and implications. *Annual Review of Sociology*, *26*(1), 1–20.

Smock, P. J., & Manning, W. D. (2004). Living together unmarried in the United States: demographic perspectives and implications for family policy. *Law & Policy*, *26*(1), 87–117.

South, S. J. (2001). Time-dependent effects of wives' employment on marital dissolution. *American Sociological Review*, *66*(2), 226–245.

Strauss, W., & Howe, N. (1991). *Generations: The History of America's Future, 1584 to 2069* (1st ed.). New York: Morrow.

Sum, A., Khatiwada, I., McLaughlin, J., & Palma, S. (2011). No country for young men: Deteriorating labor market prospects for low-skilled men in the United States. *Annals of the American Academy of Political and Social Science*, *635*(1), 24–55.

Sweeney, M. M. (2002). Two decades of family change: the shifting economic foundations of marriage. *American Sociological Review*, *67*(1), 132–147.

Sweeney, M. M. (2016). Socioeconomic standing and variability in marriage timing in the twentieth century. *The ANNALS of the American Academy of Political and Social Science*, *663*(1), 270–291.

Sweeney, M. M., & Cancian, M. (2004). The changing importance of white women's economic prospects for assortative mating. *Journal of Marriage and Family*, *66*(November), 1015–1028.

Tamborini, C. R., & Iams, H. M. (2011). Are generation X'ers different than late boomers? Family and earnings trends among recent cohorts of women at young adulthood. *Population Research and Policy Review*, *30*(1), 59–79.

Teachman, J. D. (2002). Stability across cohorts in divorce risk factors. *Demography*, *39*(2), 331–351.

Torr, B. M. (2011). The changing relationship between education and marriage in the United States, 1940-2000. *Journal of Family History*, *36*(4), 483–503.

Tumin, D. (2017). Does marriage protect health? A birth cohort comparison. *Social Science Quarterly*.

Tumin, D., & Zheng, H. (2018). Do the health benefits of marriage depend on the likelihood of marriage? *Journal of Marriage and Family*.

Uchino, B. N. (2004). *Social Support and Physical Health: Understanding the Health Consequences of Relationships*. New Haven, Conn.: Yale University Press.

Umberson, D. J., & Montez, J. K. (2010). Social relationships and health: a flashpoint for health policy. *Journal of Health and Social Behavior*, *51 Suppl*(2010), S54-66.

Van Bavel, J., Schwartz, C. R., & Esteve, A. (2018). The reversal of the gender gap in education and its consequences for family life. *Annual Review of Sociology*, *44*(1), 341–360.

Waite, L., & Gallagher, M. (2000). *The Case for Marriage: Why Married People are Healthier, Happier, and Better-Off Financially. Westminster, MD: Broadway Books*. New York: Broadway Books.

Whitbourne, S. K., & Willis, S. L. (2006). *Baby Boomers Grow Up: Contemporary Perspectives on Midlife*. Mahwah, NJ: Lawrence Erlbaum.

White, L., & Rogers, S. J. (2000). Economic circumstances and family outcomes: a review of the 1990s. *Journal of Marriage and Family*, *62*(4), 1035–1051.

Williams, K., & Umberson, D. J. (2004). Marital status, marital transitions, and health: a gendered life course perspective. *Journal of Health and Social Behavior*, *45*(1), 81–98.

Willson, A. E., Shuey, K. M., & Elder, G. H. (2007). Cumulative advantage processes as mechanisms of inequality in life course health. *American Journal of Sociology*, *112*(6), 1886–1924.

Wood, R. G., Goesling, B., & Avellar, S. (2007). *The Effects of Marriage on Health: A Synthesis of Recent Research Evidence*. Retrieved from https://aspe.hhs.gov/report/effects-marriage-health-synthesis-recent-research-evidence-research-brief

Wu, Z., Penning, M. J., Pollard, M. S., & Hart, R. (2003). “In sickness and in health”: does cohabitation count? *Journal of Family Issues*, *24*(6), 811–838.

Yang, Y., & Lee, L. C. (2009). Sex and race disparities in health: cohort variations in life course patterns. *Social Forces*, *87*(4), 2093–2124.

Zajacova, A., & Dowd, J. B. (2011). Reliability of self-rated health in US Adults. *American Journal of Epidemiology*, *174*(8), 977–983.

Zhang, Z., & Hayward, M. D. (2006). Gender, the marital life course, and cardiovascular disease in late midlife. *Journal of Marriage and Family*, *68*(3), 639–657.

Zick, C. D., & Smith, K. R. (1991). Marital transitions, poverty, and gender differences in mortality. *Journal of Marriage and Family, 53*(2), 327–336.

Appendix

Table A1 Odds Ratios for Cohort Trends by Marital Status in Poor SRH from Logistic Regression Models within Various Educational Groups for Men (Age 30–70, Cohort 1937–1967)

	Less than high school		High school		Some college		College or more	
	OR	SE	OR	SE	OR	SE	OR	SE
Marital status (ref. married)								
Divorced	1.634***	(0.204)	1.825***	(0.225)	1.044	(0.189)	2.175**	(0.524)
Widow	0.310**	(0.114)	0.440*	(0.148)	0.418	(0.187)	0.638	(0.476)
Never-married	1.276	(0.221)	1.691**	(0.296)	2.009**	(0.426)	1.181	(0.384)
Age	1.055***	(0.00302)	1.062***	(0.00264)	1.067***	(0.00305)	1.063***	(0.00462)
Cohort	1.005	(0.0209)	0.984	(0.0170)	0.966	(0.0231)	0.896***	(0.0293)
Cohort*Divorced	0.982	(0.0278)	1.027	(0.0269)	1.172***	(0.0447)	1.040	(0.0613)
Cohort*Never-married	1.029	(0.0368)	1.031	(0.0344)	1.021	(0.0434)	1.142	(0.0777)
Cohort*Widow	1.425***	(0.132)	1.409***	(0.114)	1.445***	(0.141)	1.313	(0.248)
Region (ref. Northeast)								
Midwest	1.432***	(0.108)	1.160*	(0.0710)	1.145	(0.105)	1.035	(0.125)
South	2.299***	(0.158)	1.911***	(0.107)	1.661***	(0.135)	1.518***	(0.166)
West	1.169	(0.103)	1.318***	(0.0942)	1.329**	(0.115)	1.547***	(0.166)
Recoded	0.912	(0.0841)	0.984	(0.0884)	1.528**	(0.202)	1.449*	(0.273)
N	42,222		143,395		102,776		134,679	

Notes: All values are weighted. *p < .05. **p < .01. ***p < .001. *Source:* Author's calculation based on NHIS data.

Table A2 Regression Coefficients for Cohort Trends by Marital Status in Poor SRH from Logistic Regression Models within Various Educational Groups for Women (Age 30–70, Cohort 1937–1967)

	Less than high school		High school		Some college		College or more	
	OR	SE	OR	SE	OR	SE	OR	SE
Marital status (ref. married)								
Divorced	1.996***	(0.244)	2.270***	(0.249)	1.468**	(0.210)	1.254	(0.292)
Widow	0.914	(0.162)	0.995	(0.181)	0.562*	(0.150)	1.253	(0.509)
Never-married	1.307	(0.315)	1.776**	(0.351)	2.233**	(0.631)	1.151	(0.365)
Age	1.055***	(0.00283)	1.058***	(0.00224)	1.057***	(0.00306)	1.049***	(0.00468)
Cohort	1.100***	(0.0227)	1.106***	(0.0185)	1.048*	(0.0217)	0.928*	(0.0317)
Cohort*Divorced	0.992	(0.0279)	0.994	(0.0242)	1.101***	(0.0304)	1.125*	(0.0559)
Cohort*Never-married	1.019	(0.0491)	1.035	(0.0389)	0.947	(0.0511)	1.104	(0.0698)
Cohort*Widow	1.182***	(0.0558)	1.144**	(0.0537)	1.351***	(0.0813)	1.195	(0.112)
Region (ref. Northeast)								
Midwest	1.448***	(0.109)	1.192**	(0.0676)	1.225*	(0.106)	1.371**	(0.161)
South	2.026***	(0.138)	1.713***	(0.0885)	1.626***	(0.122)	1.674***	(0.179)
West	1.440***	(0.120)	1.418***	(0.0928)	1.487***	(0.118)	1.622***	(0.181)
Recoded	0.916	(0.0873)	1.317***	(0.0993)	1.323*	(0.171)	1.339	(0.228)
N	40,932		168,663		117,794		119,132	

Notes: All values are weighted. *p < .05. **p < .01. ***p < .001. *Source:* Author's calculation based on NHIS data.

Table A3 Odds Ratios for Cohort Trends by Marital Status Including Cohabitation in Poor SRH from Logistic Regression Models within Various Educational Groups for Men (Age 30-70, Cohort 1937-1967, Only Survey Years from 1997)

	Less than high school		High school		Some college		College or more	
	OR	SE	OR	SE	OR	SE	OR	SE
Marital status (ref. married)								
Divorced	1.639***	(0.243)	1.660***	(0.174)	1.435**	(0.170)	1.560*	(0.275)
Widow	0.555*	(0.156)	0.485**	(0.107)	0.855	(0.257)	1.272	(0.690)
Never-married	1.250	(0.237)	1.407*	(0.209)	1.385*	(0.220)	1.166	(0.240)
Cohabitation	0.986	(0.259)	1.132	(0.200)	0.946	(0.232)	0.787	(0.359)
Age	1.060***	(0.00443)	1.068***	(0.00270)	1.067***	(0.00319)	1.057***	(0.00419)
Cohort	1.050	(0.0266)	1.060***	(0.0167)	1.050**	(0.0185)	0.947*	(0.0216)
Cohort*Divorced	1.004	(0.0284)	1.031	(0.0205)	1.066**	(0.0251)	1.082*	(0.0389)
Cohort*Never-married	1.056	(0.0352)	1.074**	(0.0260)	1.075**	(0.0294)	1.121**	(0.0424)
Cohort*Widow	1.264**	(0.0904)	1.337***	(0.0690)	1.151*	(0.0802)	1147	(0.134)
Cohort*cohabitation	0.953	(0.0426)	0.955	(0.0296)	0.976	(0.0422)	0.996	(0.0783)
Region (ref. Northeast)								
Midwest	1.394***	(0.101)	1.113**	(0.0421)	1.155*	(0.0672)	1091	(0.0818)
South	1.957***	(0.139)	1.596***	(0.0585)	1.487***	(0.0772)	1.233**	(0.0827)
West	1164	(0.102)	1.140**	(0.0549)	1.259***	(0.0692)	1.176*	(0.0832)
N	15,526		64,842		56,952		68,379	

Notes: All values are weighted. *p < .05. **p < .01. ***p < .001. *Source:* Author’s calculation based on NHIS data.

Table A4 Odds Ratios for Cohort Trends by Marital Status Including Cohabitation in Poor SRH from Logistic Regression Models within Various Educational Groups for Women (Age 30–70, Cohort 1937–1967, Only Survey Years from 1997)

	Less than high school		High school		Some college		College or more	
	OR	SE	OR	SE	OR	SE	OR	SE
Marital status (ref. married)								
Divorced	1.750***	(0.241)	1.590***	(0.136)	1.415***	(0.139)	1.692**	(0.274)
Widow	0.687*	(0.111)	0.702**	(0.0839)	0.835	(0.132)	1.129	(0.333)
Never-married	0.934	(0.234)	1.348	(0.207)	1.656**	(0.314)	1.219	(0.257)
Cohabitation	0.727	(0.210)	1.047	(0.183)	1.245	(0.292)	1.089	(0.420)
Age	1.051***	(0.00432)	1.054***	(0.00267)	1.049***	(0.00295)	1.041***	(0.00409)
Cohort	1.103***	(0.0262)	1.090***	(0.0163)	1.070***	(0.0187)	0.991	(0.0241)
Cohort*Divorced	1.012	(0.0265)	1.043*	(0.0181)	1.083***	(0.0196)	1.030	(0.0320)
Cohort*Never-married	1.071	(0.0452)	1.087**	(0.0290)	1.022	(0.0327)	1.084*	(0.0412)
Cohort*Widow	1.265***	(0.0533)	1.229***	(0.0357)	1.219***	(0.0441)	1.149*	(0.0811)
Cohort*cohabitation	1.044	(0.0519)	0.962	(0.0290)	0.946	(0.0365)	0.950	(0.0638)
Region (ref. Northeast)								
Midwest	1.467***	(0.107)	1.141**	(0.0460)	1.112*	(0.0589)	1.169*	(0.0779)
South	1.858***	(0.127)	1.557***	(0.0582)	1.381***	(0.0640)	1.366***	(0.0872)
West	1.322**	(0.112)	1.225***	(0.0578)	1.278***	(0.0658)	1.297***	(0.0832)
N	13,999		69,692		67,548		65,960	

Notes: All values are weighted. *p < .05. **p < .01. ***p < .001. *Source:* Author's calculation based on NHIS data.

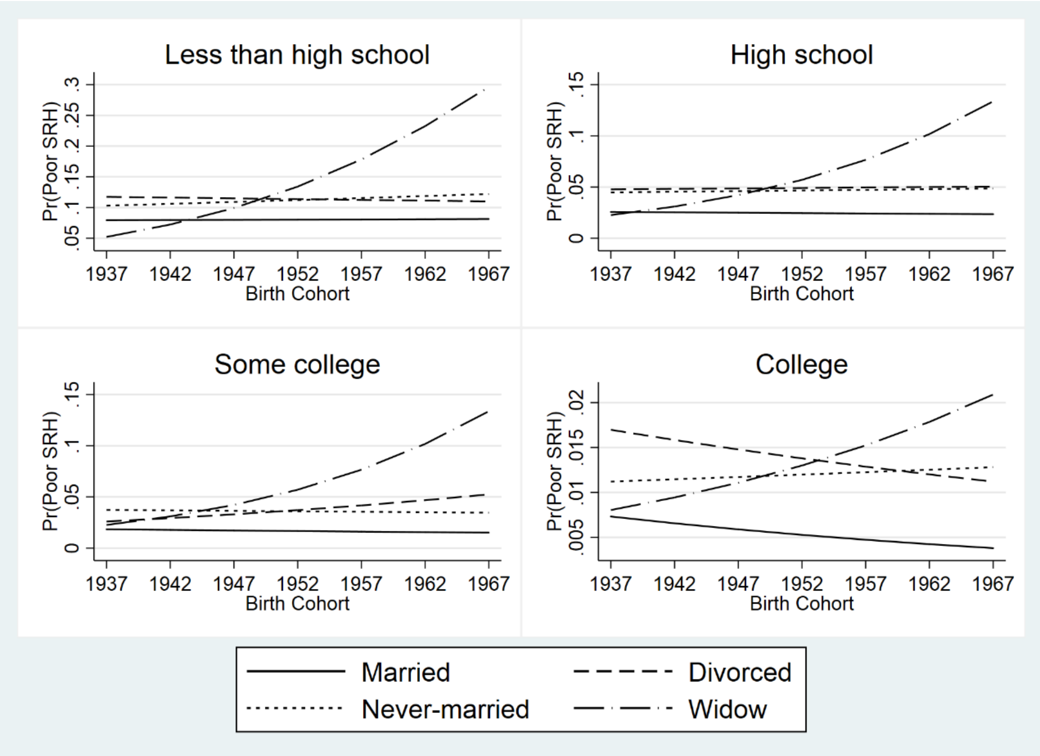

Figure A1 Predicted probability of poor SRH by educational group (men, age 30–70).

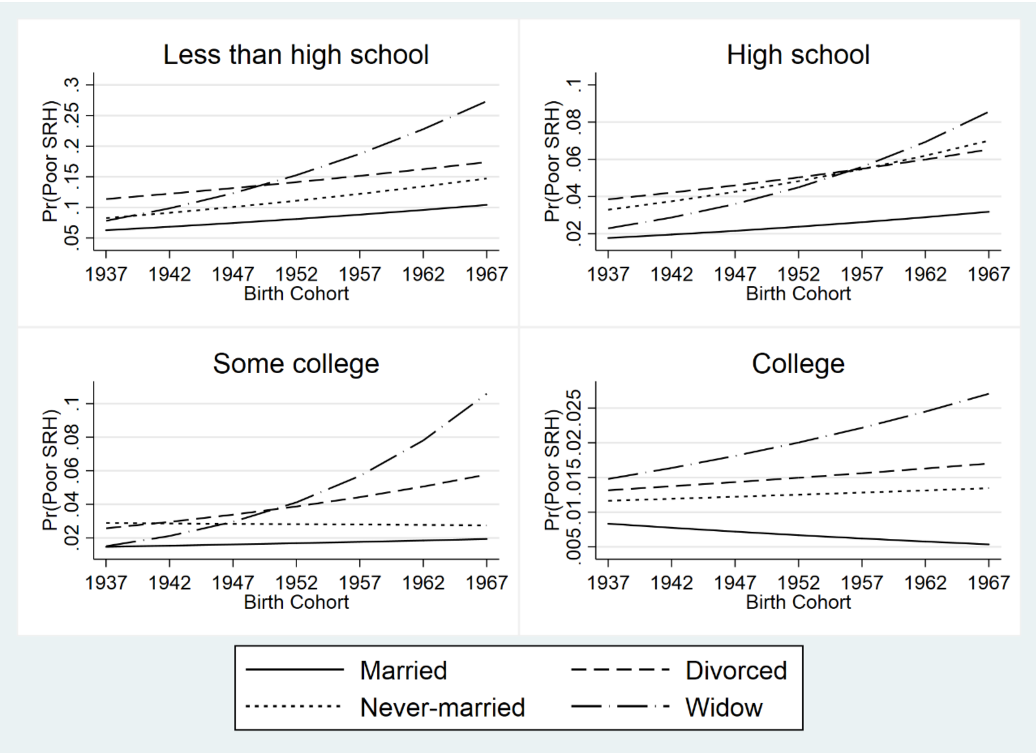

Figure A2 Predicted probability of poor SRH by educational group (women, age 30–70).

Figure A3 Percentage of health gap between high school and less and college-educated (only poor SRH) as explained by marriage over cohorts in different age groups for women and men (non-Hispanic White).

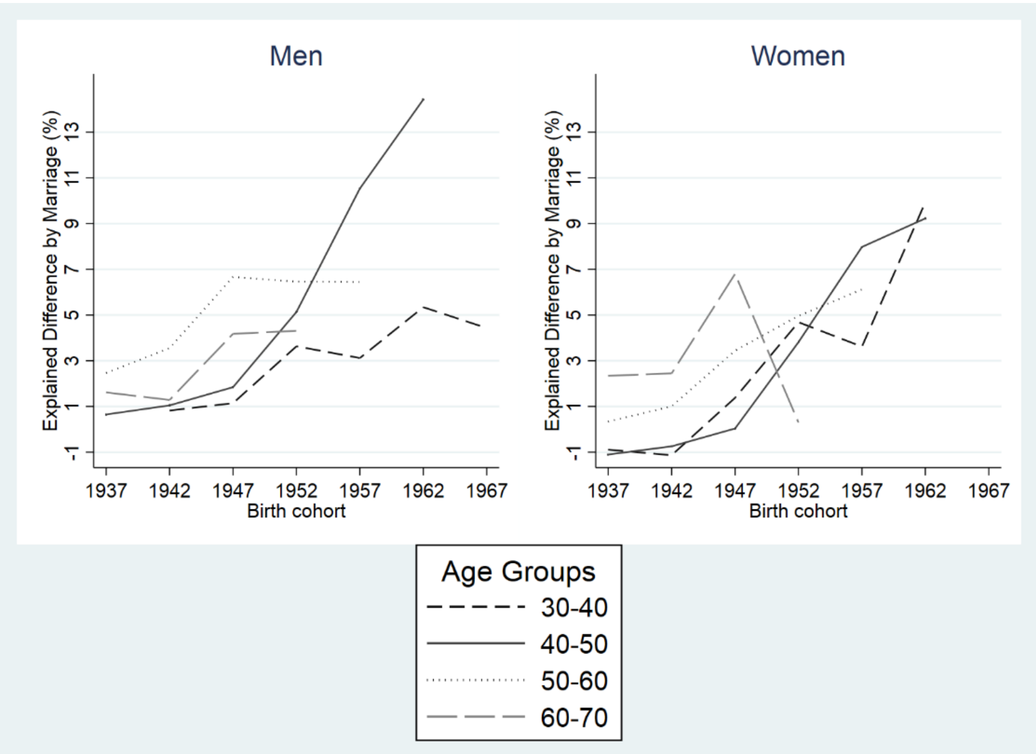

Figure A4 Percentage of health gap between less than high school and college-educated as explained by marriage over cohorts in different age groups for women and men (non-Hispanic White).

Chapter 4

The Causal Effect of Education on Cancer Risk and Survival in England and Wales

Abstract

A growing number of people are being diagnosed with cancer, so a better understanding of educational inequalities in cancer is becoming increasingly important. However, only a few studies have focused on the causal effect of education on cancer risk and mortality. Using data from the ONS Longitudinal Study (ONS LS) this quasi-experimental design exploits the exogenous variation in compulsory schooling due to policy changes in 1947 and 1972 in England and Wales, which affected a large proportion of the population. The identification of the causal effect of education on cancer risk and survival is achieved using a regression discontinuity framework with categorical outcomes. Therefore, this work aims to investigate the effect of education on the probability of developing, surviving and dying from cancer. The results show that the 1947 and 1972 educational reforms do not seem to be causally related to cancer incidence and mortality, suggesting an absence of a causal role of one additional year of compulsory schooling on cancer outcomes. Among the reasons for the absence of causal evidence

there are the heterogeneity by cancer site, as well as the local effect of the compulsory school reforms.

Keywords: causality; cancer; schooling reform; educational gradient in health; natural experiment

4.1 Introduction

Cancer is becoming an increasingly common disease, especially due to the ageing population and the reduction in infectious disease fatalities. Omran (1971) describes the epidemiological transition as the long-term shift in disease patterns, where deaths from degenerative and man-made diseases substitute those from infectious diseases as the leading causes of death. This change has increased the relevance of non-communicable diseases for population dynamics, and cancer has gradually become a leading cause of death. In 2016, the number of cancer deaths in the world was estimated at approximately 8.9 million and has been growing over time (Foreman et al., 2018). At the same time, socioeconomic factors are important drivers for the epidemiological transition. This research explores a particular pathway through which socioeconomic circumstances affect this increasingly common morbidity: education. An exogenous increase in compulsory schooling due to policy changes could reduce cancer incidence and fatalities if education provides additional resources to individuals exposed to this exogenous increase in years of schooling. In particular, scholars have called for more research on what are the circumstances enhancing or suppressing the causal effect of education on health (Montez & Friedman, 2015).

The existence of an association between educational attainment and cancer risk and survival is documented in the literature (Ø. Kravdal, 2000; Woods, Rachet, & Coleman, 2006). However, only a few studies have tried to explore directly the causal effect of education (Buckles, Hagemann, Malamud, Morrill, & Wozniak, 2013; Davies, Dickson, Davey Smith, van den Berg, & Windmeijer, 2018; Lager & Torssander, 2012; Leuven, Plug, & Rønning, 2016; Palme & Simeonova, 2015). This work aims to analyse the causal effect of education on cancer incidence, mortality and survival in England and Wales, taking advantage of two reforms, in particular, the compulsory schooling law changes in 1947 and 1972 in England and Wales. The English context represents an ideal setting for exploring the causal relationship between education and cancer as England has universal health insurance coverage, the National Health Service (NHS), which provides a comprehensive range of healthcare services for free over the entire life-course. Despite this, significant socioeconomic differences in health and mortality are present in the country (Mackenbach et al., 2015, 2008; Marmot, 2010). The specificity of the English context, as well as the two different time periods in which the reforms took place, provide an ideal setting for exploring the conditions under which the causal effect of education for cancer outcomes exists.

This work contributes to the literature in three important ways. First, it was determined whether a causal nexus between education and cancer survival exists. Previous studies have shown that survival after cancer diagnosis differs by socioeconomic status to a greater extent than incidence (Ø. Kravdal, 2000). It is important to study how education causally affects survival after cancer diagnosis given the presence of a gap in survival rates by socioeconomic status in observational studies. Simultaneously, there is not much known about the reasons for the existence of those inequalities in survival (Kogevinas & Porta, 1997). The present study significantly contributes to the literature by finding no effect of one additional year of compulsory schooling on cancer survival following the 1947 and 1972 reforms in England and Wales.

Second, whether education has a causal role in cancer incidence and survival in two time periods and at two different points of the educational distribution was explored. Compared to other studies, the present analysis provides evidence on the causal effect of education on cancer outcomes for the same country in two distinct historical periods. In particular, while the first reform took place in a period when risk factors associated with cancer were still not well known, more information was increasingly available for individuals affected by the second reform, in addition to free screening provided by the NHS for some cancer sites. For example, the NHS introduced cancer screening for breast and cervical cancer in 1988 (Albrow, Kitchener, Gupta, & Desai, 2012), therefore, individuals affected by the 1972 reform could benefit from free cancer screening for a large part of their adult life. Moreover, the increase in years of compulsory education happened at slightly different points in the educational distribution: the first reform increased the school leaving age from 14 to 15, the second increase from 15 to 16. In conclusion, the results from the two reforms give a comprehensive view of the effect of an additional year of compulsory schooling for cancer outcomes in England and Wales.

Third, by using a regression discontinuity (RD) design with categorical outcomes, this work aims to model different outcomes simultaneously, which is crucial for understanding cancer risk and mortality from other causes. Moreover, short-term and long-term mortality after cancer diagnosis is modelled. Therefore, this method gives an overall picture about the timing of mortality after diagnosis, as well as solving some methodological problems arising from the use of binary variables as outcomes. Given the novelty of the method, this work represents the first empirical application.

The presence of a causal effect of education on cancer is plausible for at least two reasons. First, education might influence cancer incidence and survival through greater access to resources in terms of risk behaviours and prevention. Many advocate the central role of prevention in reducing cancer incidence through lifestyle modifications (Ashford et al., 2015; Song & Giovannucci, 2016). Healthy behaviours, such as no smoking, physical activity and balanced diets, have been found to be associated with cancer incidence, in addition to survival after diagnosis (Donaldson, 2004; Greenwald, Clifford, & Milner, 2001; McTiernan, 2008; Naik et al., 2016; Sasco, Secretan, & Straif, 2004). Previous research has shown that an educational gradient in healthy behaviours exists (Cutler & Lleras-Muney, 2010), therefore it can be expected that cancer incidence and survival might be affected by education through this channel. However, the presence of some knowledge about cancer aetiology is essential for giving more educated groups the ability to benefit from their advantageous position. Information about preventative measures for many cancers has not been always been unambiguous and widespread for the general public. Hence, it is unclear the extent to which resources given by education might have influenced cancer prevention. In addition, screening might be related to cancer outcomes. Some studies have shown that stage at diagnosis could explain survival differences (Sant et al., 2003) and differences in the stage of diseases are observed across social groups (see for review Woods, 2006). At the same time, other studies have found that socioeconomic differences in survival persist even after accounting for the stage of cancer (Ø. Kravdal, 2000).

Second, the adoption of new therapies and the receipt of adequate treatments are other possible mechanisms that may influence the relationship between education and cancer outcomes. Previous studies have found that higher educational attainment leads to an increasing adoption of new medical technologies, which, in turn, affects health and mortality differentials (Glied & Lleras-Muney, 2008). Similarly, receiving adequate medical treatments has been found to depend on socioeconomic characteristics in different contexts (Cookson, Propper, Asaria, & Raine, 2016; Henley et al., 2005; Pollock & Vickers, 1998; Vaneenwyk, Campo, & Ossiander, 2002).

4.2 Background

4.2.1 Theoretical frameworks

Education is associated with different aspects of cancer epidemiology from incidence to detection to treatment and to survival (Kawachi & Kroenke, 2006). However, these results are greatly dependent on the cancer type as well as the growth site. Moreover, the mechanisms through which education affects cancer are often unclear. For example, more education is associated with a higher incidence of prostate cancer, breast cancer and melanoma, but the reasons for this have yet to be unequivocally determined. Therefore, it appears difficult to suggest a unique mechanism that explains all the associations. Moreover, knowledge about cancer prevention and treatment has increased slowly over time, offering few opportunities for more advantaged socioeconomic groups to make use of this differential knowledge.

According to the *fundamental cause* perspective, socioeconomic inequalities emerge when more advantaged individuals are able to make use of information and resources to disproportionately benefit their health (Link & Phelan, 1995; Phelan, Link, Diez-Roux, Kawachi, & Levin, 2004). Therefore, according to this framework knowledge plays an essential role for understanding the emergence of educational disparities in cancer incidence and mortality. Causality appears to be extremely dependent on context, since information and resources are contingent to time and place (Montez & Friedman, 2015). Specific cancer sites have been used to test the *fundamental cause* theory by contrasting specific types of preventable and unpreventable cancer (Rubin, Clouston, & Link, 2014; Wang, Clouston, Rubin, Colen, & Link, 2012). However, few studies have been able to complement it with causal evidence.

The *fundamental cause* framework is considered as a meta-mechanism directing the attention on the resources which help to avoid diseases and mitigate the consequences in case they occur. Therefore, according to it, socioeconomic status provides a flexible set of resources which allows to produce health benefits. When the mechanism leading to socioeconomic differences is replaced, those with greater resources keep having advantageous conditions by using different resources or mechanisms. However, this resource flexibility is particularly difficult to measure since it can take different manifestations, especially when looking at health outcomes over time. The *fundamental cause* framework puts less emphasis on proximal causes, such as health behaviours and occupational risk, and more on the resource flexibility that helps to maintain benefits for more advantaged individuals.

The *social determinants of health* perspective shifts the attention on the mechanisms through which health inequalities arise, as well as on the role of the institutions in shaping inequalities in health (Marmot, 2010). Therefore, there is more emphasis on the role of public policies to mitigate and reduce health inequality by shaping the environment (Marmot, 2018). Moreover, the *life-course* perspective, which especially emphasises the role of early life experiences in determining adult health, is another general framework often used to explain the presence of health inequality in society (dos Santos Silva, 2004; Mackenbach, 2012; Potischman, Troisi, & Vatten, 2004). These theories are not mutually exclusive, but they can help understanding the educational differences in cancer incidence and survival. Therefore, the *fundamental* *cause* and the *social determinants of health* are used as interpretative lenses in this study of the causal effect of education on cancer outcomes.

4.2.2 Socioeconomic inequalities in cancer incidence and survival

Investigations of cancer incidence and survival have been conducted in England and Wales, documenting an important role of socioeconomic status in understanding these inequalities but not specifically education (Kogevinas, 1990; Leon, 1988; Shack, Jordan, Thomson, Mak, & Møller, 2008). Associations between educational attainment and cancer incidence are observable in different contexts, but are sometimes weak and inconsistent (Clegg et al., 2009; Ma, Xu, Anderson, & Jemal, 2012). In particular, the association between cancer risk and education is vastly dependent on the type and site of cancer, while leukaemia and lymphoma incidences are not associated with education (Hermann et al., 2010), other cancer sites are positively or negatively linked.

On one side, higher educational attainment reduces the risk of development in some cancer sites, such as lung, colorectal, and cervical. First, lung cancer has a strong educational gradient, mostly due to a higher prevalence of smoking among the low educated (Mackenbach et al., 2004), for example, more than 50 per cent of the differences in lung cancer incidence by education is due to smoking (Menvielle et al., 2009). However, considerable differences in smoking prevalence and lung cancer rates are present across Europe (Janssen-Heijnen & Coebergh, 2003; Mackenbach et al., 2004). As Janssen-Heijen and Coebergh (2003) have pointed out, trends in lung cancer incidence over time are highly correlated with the smoking prevalence in the population and they largely differ by country. Similarly, cervix cancer is associated with low education, and smoking has also an important role, with a relative contribution of 30 per cent for reducing educational differences (Braaten, Weiderpass, Kumle, & Lund, 2005). Finally, the educational gradient in colorectal cancer incidence is dependent on the country context to a very large extent (Manser & Bauerfeind, 2014), colorectal cancer incidence is higher in the United States among the low educated, while in European countries, this pattern is not always consistent (Aarts, Lemmens, Louwman, Kunst, & Coebergh, 2010).

In contrast, the opposite educational gradient is observable for other cancers, such as skin, testicular, prostate, and breast cancer (Braaten et al., 2005; Hemminki & Li, 2003; Menvielle & Kunst, 2008). The higher incidence of breast cancer has been linked to some reproductive and behavioural factors among the highest educated, such as age at first birth, number of children, alcohol consumption, and use of hormone replacement therapies. Similarly, educational differences in melanoma incidence can be partially explained by lifestyle factors, such as excessive exposure to sunlight. Braaten et al. (2005) quantified the contribution of these different factors for the incidence of breast cancer, finding a large contribution of reproductive factors and alcohol consumption, explaining in total more than 50 per cent of the difference in the incidence by education. However, this educational gradient can vary by country as well as over time (Klassen & Smith, 2011; Lundqvist, Andersson, Ahlberg, Nilbert, & Gerdtham, 2016). Moreover, prostate cancer incidence is more common among men with more education, also due to a higher take up of screening and more precise detection (Steenland, Rodriguez, Mondul, & Calle, 2004). Therefore, education is also positively associated with overdiagnosis of low-risk prostate cancers (Kilpelaïnen et al., 2016; Loeb et al., 2014). The presence of more accessible screening tools, such as prostate-specific antigen blood test, may have enlarged this gradient over time (Liu, Cozen, Bernstein, Ross, & Deapen, 2001).

Overall, socioeconomic differences in cancer survival are more pronounced than those in cancer incidence (Braaten, Weiderpass, & Lund, 2009; H. Kravdal, 2014; Sloggett, Young, & Grundy, 2007). Previous research on cancer survival and its relationship with different socioeconomic indicators have identified a general positive association (Coleman et al., 2011, 2004; Du, Lin, Johnson, & Altekruse, 2011; Ø. Kravdal, 2000; Lyratzopoulos et al., 2011; Rachet et al., 2009; Steenland, Henley, & Thun, 2002; Woods et al., 2006). Few studies have looked only at education when studying the association between socioeconomic status and cancer survival (Braaten et al., 2009; Kogevinas & Porta, 1997; Ø. Kravdal, 2000). In particular, Ø. Kravdal (2000) found strong associations between socioeconomic status and cancer survival in Norway using three different measurements for socioeconomic status: income, education and occupation. Consistent associations were found across income, occupation and education, in particular, survival was 15 per cent higher for those with educational attainment more than compulsory schooling. In England and Wales, first Leon (1988) and Kogevinas (1990) explored socioeconomic gradients in cancer in incidence and survival, then, Sloggett and colleagues (2007) examined the socioeconomic gradient in overall cancer survival using three individual-level measures of socioeconomic status (social class, housing tenure, and household access to a car) and one area-level indicator (carstairs index). These studies found a higher incidence and lower survival for individuals in more disadvantageous socioeconomic conditions.

Moreover, the association between education and cancer survival changes over time and by country, for example, differences in breast cancer mortality among educational groups are narrowing over time in some countries (Elstad, Torstensrud, Lyngstad, & Kravdal, 2012; Martikainen & Valkonen, 2000; Menvielle, Leclerc, Chastang, & Luce, 2006; Strand et al., 2007; Wagener & Schatzkin, 1994). Another example is the large variation across Europe in survival from lung cancer. The health care structural factors may play a role in explaining these differences as the proportion of patients receiving surgery and other in-depth treatment for lung cancer in the United Kingdom appears to be lower compared to other European countries (Janssen-Heijnen & Coebergh, 2003).

Primarily, three different mechanisms have been hypothesised in the literature to explain the higher survival after the onset of cancer for more advantaged socioeconomic groups (Kawachi & Kroenke, 2006; Ø. Kravdal, 2000). First, the characteristics of the cancer, such as site and stage, can present an educational gradient, particularly, diagnosis time and type of cancer might differ across educational groups, leading to higher survival for more advantaged groups. Second, treatment might differ across educational groups. In England and Wales, the NHS provides free universal access to care, but type and quality of treatment might be different by education, as well as lower compliance to therapies and follow-up. For example, Cookson et al. (2016) reviewed inequalities in health care utilisation in England, finding meaningful differences in cancer treatment by socioeconomic status. Third, individual characteristics, such as comorbidities and immune functions, might be important mediators for the association between education and survival. For example, better health conditions could affect progression rate as well as treatment tolerance. Moreover, lower education is associated with some of the behavioural risk factors related to faster cancer progression and smaller efficacy of treatments.

4.2.3 Beyond the associations

Despite the strong correlational evidence, the association between education and cancer might not reflective a causal relationship. Indeed, there might be omitted "third variables" in associational studies which affect both schooling and cancer outcomes.

First, childhood health and adversities might influence both schooling and cancer. For example, the association found between education and cancer might not reflect a causal nexus if childhood circumstances have not been taken into account (Ferraro, Schafer, & Wilkinson, 2016; Hayward & Gorman, 2004). Second, time and risk preferences might affect health behaviours, in turn, cancer outcomes and education (Fuchs, 1982; Lawless, Drichoutis, & Jr, 2013). Third, cognitive ability could be linked to schooling and cancer incidence and mortality by decreasing risky behaviours (Cutler & Lleras-Muney, 2010). Finally, there might be other unobservable characteristics that are responsible for confounding the causal relationship between education and cancer outcomes.

In order to overcome these limitations in understanding the causal nexus, several methods have been proposed in the literature. Randomised control trials (RCT) represent an extremely reliable method for examining causality (Imbens, 2010), but their implementation poses many challenges that often lead to samples that are not representative. Twin studies aim to look at the variations in education within twins, given that they share similar conditions early in life and genetic endowment, but twin samples are difficult to obtain and might not be representative of the overall population. Finally, quasi-experimental designs or natural experiments can provide local causal estimates by making use of local exogenous variations in education. The presence of random variation in education provides a good setting for exploring the causal relationship between education and different outcomes. Instrumental variables (IV), difference-in-difference estimators and RD designs can be used to estimate causality in the presence of a natural experiment.

This work uses a RD design that compares individuals right before and after two compulsory school reforms in England and Wales. The identification starts from the assumption that individuals right below and above the cut-off for the policy change have similar characteristics but different assignment to the reform treatment. Therefore, the average cancer outcomes of those right before the reform implementations could be used as counterfactuals for those affected by the reform (Lee & Lemieux, 2013). RD designs exploit the local randomisation around the cut-off, representing a reliable method for the estimation of treatment effects. They are also closely linked to RCTs, where the randomisation holds globally. A key advantage of using RD methods for studying policy changes over time is the lower sensitivity to cohort trends compared to IV regressions.

4.2.4 The 1947 and 1972 reforms

Compulsory school laws in Britain define the maximum age at which children should start education and the minimum age at which they are allowed to leave education. Consequently, these laws determine how many years children spend in compulsory schooling. During the 20th century, two legal changes have increased the minimum age at which children were allowed to leave school in England and Wales. First, the 1944 Educational Act established an increase in leaving school age from 14 to 15, which was implemented from the 1st April 1947. Second, the same Education Act authorised a further increase in school leaving age from 15 to 16, which was implemented from 1st September 1972. The economic and social contexts surrounding these two changes were quite different, which offer a more comprehensive view of the causal effect of education on cancer risk and survival after two reforms occurring in different time periods.

The two aforementioned reforms in England and Wales have been used for identifying the causal effect of education on different outcomes, such as earnings, fertility, assortative mating, health and mortality (Clark & Royer, 2013; Geruso & Royer, 2018; Oreopoulos, 2006). The exogenous increase in education has been associated with higher earnings, especially for males (Devereux & Hart, 2010; Grenet, 2013; Oreopoulos, 2006). Geruso and Royer (2018) found a reduction in teen pregnancies, but not in completed fertility for women affected by the 1972 reform. In addition, they found that the reform led individuals to find more educated and younger partners. Concerning health and mortality, some studies have found no significant effects of the reform (Braakmann, 2011; Clark & Royer, 2013), whereas Silles (2009) identified a protective effect of education for health using the same reforms. Likewise, Davies and colleagues (2018) found that education is causally related to lower mortality and diabetes diagnosis after the 1972 reform. This protective effect is in line with previous research that revealed the presence of a causal effect of education on health in other contexts (Fletcher, 2015; Lleras-muney, 2005). Overall, the causal effect of education on health and mortality does not seem to be univocally supported across different institutional contexts and identification strategies (Mazumder, 2011). These mixed results leave open the debate on the causal effect of education on health and mortality. Therefore, this work aims to enter the debate by providing causal evidence for the effect of education on health, specifically on cancer risk and survival in England and Wales using the 1947 and 1972 compulsory school reforms.

4.2.5 Causal role of education on cancer

Previous research has used compulsory school law changes to identify the effect of education on cancer across different institutional settings (Davies et al., 2018; Lager & Torssander, 2012; Leuven et al., 2016; Palme & Simeonova, 2015). However, the causal effect of education on cancer survival has not been investigated in the literature so far.

Therefore, this work represents the first investigation of the causal role of education on overall cancer survival after diagnosis.

It is still an open question whether and why education affects health dissimilarly across different institutional settings. Research in Nordic countries found a protective effect of education for cancer incidence mortality, with the exception of breast cancer where the reverse was observed (Lager & Torssander, 2012; Leuven et al., 2016; Palme & Simeonova, 2015). Leuven and colleagues (2016) studied the effect of compulsory school reforms for cancer risk in Norway. Overall, they did not find a causal relationship of education on cancer risk, but education seemed to be related to a diminishing incidence of lung and prostate cancer in men. Lager and Torssander (2012) utilised the one-year increase in compulsory schooling in Sweden to study whether education causally affects mortality. Sweden had an extensive school reform between 1949 and 1962, with educational changes adopted gradually across municipalities and the study exploited this heterogeneity in adoption. They also explored cause-specific mortality and found that those affected by the reform had reduced overall cancer mortality, particularly, lung cancer mortality. Palme and Simeonova (2015) analysed the causal effect of education for breast cancer in Sweden using the same exogenous variation in compulsory schooling laws across municipalities. Breast cancer is usually more common among higher educated women, and they found support for a causal positive effect of education on breast cancer incidence and mortality.

In the United States (US), Glied and Lleras-Muney (2008) explored the education gradient in overall and cancer mortality, finding that the educational differences in mortality are more pronounced for diseases in which more innovations are available. They exploited as identification strategy changes in the compulsory school laws across states, finding that regarding cancer mortality, education has a negative and significant effect, especially for men. Similarly, Buckles (2013) identified a strong causal effect of college completion in the US for cancer mortality (especially lung cancer). Using entrance to college to avoid conscription during the Vietnam War as an instrumental variable, they uncovered a large causal effect of college education on health. Overall, evidence from the US suggests that education is causally related to a lower likelihood of cancer mortality at different points of the educational distribution.

In England and Wales, Davies and colleagues (2018) used Biobank data and looked at the effect of the 1972 reform on several health outcomes, including cancer. They found that being exposed to the educational reform lowered the risk of diabetes and mortality, but there was a lack of a causal effect of education on cancer incidence. This study differs from their work by overcoming some of their limitations. First, their work did not analyse survival after cancer diagnosis and mortality from cancer. Second, their analysis was based on Biobank data, which has been criticised for not being representative of the UK population (Fry et al., 2017; Swanson, 2012). Indeed, Fry and colleagues (2017) suggested that cancer incidence in the UK Biobank sample is 10-20 per cent lower compared to the general population. Second, individuals participating in the Biobank are from more socioeconomically advantage areas and have a lower prevalence of risky behaviours. This non-representability does not disprove their results, but it might introduce some bias, especially given the differences in cancer incidence rates compared to the general population. On the contrary, our analysis is based on a high-quality and representative sample of English and Welsh population derived from the census that takes into account both educational reforms, thereby providing a more holistic view of the causal effect of education for cancer outcomes, since both the 1947 and 1972 reforms are considered. Third, they were unable to consider the heterogeneity in preventable and unpreventable cancer sites, unlike the present study.

4.2.6 Some mechanisms linking causally education and cancer

The mechanisms through which education affects cancer incidence are still unclear in the literature and are highly dependent on cancer sites. However, some risk factors are known to be consistently associated with cancer outcomes. Some of the main risk factors will be briefly reviewed, which are also found to be linked to educational attainment (sometimes also causally), such as smoking, obesity, absence of physical activity, diet, and excessive alcohol consumption (Kawachi & Kroenke, 2006).

First, smoking is an important determinant in the association between socioeconomic status and cancer. It has been established that smoking increases cancer incidence, especially lung cancer (Alberg, Shopland, & Cummings, 2014; Ho & Fenelon, 2015; McCormack & Boffetta, 2011). A strong educational gradient exists in England and Wales for smoking prevalence (Jarvis & Wardle, 2005), but whether a causal effect for education on smoking behaviour exists is very much debated. An additional year of compulsory schooling does not seem to be related to lower smoking prevalence in England (Braakmann, 2011; Clark & Royer, 2013; James, 2015), with the exception of Davies et al. (2018), who found an effect in the probability of current and ever smoking. However, other aspects of education in different countries, such type of school and tracking, seem to be related to a reduction in smoking (see for review Galama et al., 2018).

Secondly, being obese is associated with a higher incidence of different types of cancers, like breast, colorectal, prostate, thyroid and ovarian cancer (Wolin, Carson, & Colditz, 2010). Moreover, absence of physical activity might be one of the mechanisms through which obesity is positively associated with cancer incidence and mortality. Similar to smoking, socioeconomic inequalities in obesity are present in England and Wales (Rennie & Jebb, 2005; Zaninotto, Head, Stamatakis, Wardle, & Mindell, 2009). The evidence about the causal role of education for obesity is overall weak, with many studies showing insignificant effects but some significant effects in reducing only body mass index (Arendt, 2005; Brunello, Fabbri, & Fort, 2013; Davies et al., 2018; Galama et al., 2018; James, 2015).

Third, reproductive behaviour is one of the explanations for the association between educational attainment and breast cancer. The biological pathways through which pregnancy characteristics are related to breast cancer are still unclear and not well understood (Nechuta, Paneth, & Velie, 2010), but Palme and Simeonova (2015) tested some of the fertility behaviours previously found to be associated with education and breast cancer, such number of children and age at first child, finding no evidence of a causal role of education for delayed childbearing in Sweden. Other studies in England and Wales have looked at effect of the educational reform in 1947 on fertility, but it was unclear whether the total number of children for women was affected by the reform (Fort, Schneeweis, & Winter-Ebmer, 2016; Geruso & Royer, 2018). Moreover, Geruso and Royer (2018) found a significant effect of the 1972 reform in lowering teen fertility leading to delayed childbearing. Therefore, it seems that fertility behaviours might represent a plausible mechanism for the causal effect of education on breast cancer in England and Wales.

Finally, an unhealthy diet including excessive consumption of fats, red meat and alcohol represents a risk factor for the incidence of some cancers, such as colorectal cancer. Heavy alcohol consumption is also associated with an increased incidence of mouth, pharynx, larynx, oesophagus, liver, and breast cancer (Boffetta, Hashibe, La

Vecchia, Zatonski, & Rehm, 2006). However, the relationship between education and alcohol consumption is complex, with some studies showing that college education increases alcohol consumption among young adults (Crosnoe & Riegle-Crumb, 2007). Moreover, Braakman (2011) have found in England and Wales no causal effect of education on various health behaviours related to diet and alcohol consumption.

4.3 Data and method

4.3.1 Data

The data source is the ONS Longitudinal Study (ONS LS) for England and Wales. ONS LS contains census-linked information concerning vital events for 1 per cent of the population of England and Wales from 1971 to 2011. Events registered are births, deaths, immigrations, emigrations, and cancer. Data are also linked to death records that allow for the examination of cancer survival. Approximately 500,000 individuals were followed longitudinally across different censuses (1971–2011), of which, 135,000 experienced a cancer event and there was information concerning their diagnosis and disease survival. The National Cancer Registration System (NCRS) is the source for cancer diagnosis. However, it is important to note that before 1993, cancer registration in England and Wales was not mandatory, which led to a large regional variation in registration methods and completeness. Overall, previous studies have shown the general reliability of Cancer Registry records (Gulliford, 1996; Gulliford, Bell, Bourne, & Petruckevitch, 1993), despite the presence of small inconsistencies (Pollock & Vickers, 1995). Deaths are linked to the ONS LS through two different methods: routine notification of events to the National Health Service Central Register (NHSCR) and annual searching of ONS computer files looking for events to persons with a LS date of birth. It is estimated that currently around 1 per cent of the deaths are missing.

Cancer incidence

A dichotomous variable for cancer incidence at any age was created, only including the first cancer diagnosis. Cancer registrations are up-to-date to 31st December 2015.

Cancer mortality and survival

Cancer mortality was measured using two different indicators: a dummy variable for mortality with cause of death cancer for all sample and a series of indicators were created for dying after cancer diagnosis (dying from any cause within 5 years after first cancer diagnosis; and dying within 5–10 years from cancer diagnosis; being alive 10 years after cancer diagnosis) only for those diagnosed with cancer. The mortality follow-up was until 31st December 2015.

Cancer sites

The most common cancer sites, such as the lung, colon-rectum, breast, and prostate, were considered. Moreover, a dichotomous indicator was created for marking whether that particular site was among the most preventable cancer. The ten cancers which account for the most preventable cases in the UK in 2015 were derived from Brown and colleagues (2018) were in the following sites: lung, colon-rectum, melanoma, breast, oesophagus, bladder, kidney, stomach, pancreas and cervix. The incidence from any of these cancers was aggregated to explore the presence of heterogeneity by preventability.

Education

Across censuses, different educational qualifications were reported. The educational attainment was divided into three groups reporting the information from the census with a higher level of detail (in order of wealth of information 1971, 2001, 2011, 1981 and 1991). The lowest level of education was no qualifications, then secondary level, which included A level or more but not university degree [this includes Certificate of Secondary Educations (CSE) (grades 2–5), General Certificate of Secondary

Educations (GCSE) (grades D–G), 1–4 CSEs (grade 1), 1–4 GCSEs (grades A–C), 1–4 O levels, NVQ level 1, Foundation GNV, 5+O levels, 5+CSEs (grade 1), 5+GCSEs (grades A–C) etc., 1 A level, 1–3 AS levels, NVQ level 2, Intermediate GNVQ), then tertiary level, which includes degrees of various types, such as degree (BA, BSc), higher degree (MA, PhD, PGCE), NVQ Level 4–5, HNC, HND, RSA Higher Diploma, BTEC Higher level, Foundation degree (NI), professional, or qualifications (teaching, nursing, accountancy)]. However, each census had different levels of detail available, so an education variable was created including the qualifications from the census available with more information.

4.3.2 Empirical strategy

First, linear probability models were used to examine the association between educational attainment, cancer incidence, and mortality after cancer diagnosis. According to the literature, the probability of getting cancer is expected to be lower for the most educated groups. However, the probability of experiencing some cancers is higher in the most educated groups, such as breast and prostate cancer. The linear probability was estimated instead of survival models to obtain some baseline estimates for comparison to the RD estimates. Linear probability models are used to model dichotomous dependent variables with a straightforward estimation and interpretation. Overall, it is expected that education will be negatively related to cancer diagnosis and mortality.

Next, to identify the causal effect of education on cancer risk and survival in England and Wales, a RD approach (Imbens & Lemieux, 2008; Lee & Lemieux, 2010) was used with categorical outcomes (Xu, 2017). RD with categorical outcomes was chosen because it allows the inclusion of competing risks in the analysis, as well as solving some of the problems arising from the use of a simple binary variable as an outcome; duration outcomes can be discretised and used as outcomes in a RD framework (Caliendo, Tatsiramos, & Uhlendorff, 2013; Landais, 2013; Sueyoshi, 1995). Given the novelty of the method, this paper represents its first empirical application.

RD designs have four main features: the running variable, the cut-off, the treatment, and the bandwidth. The running variable is the score that leads to the assignment of the treatment. In this work, the assignment is determined by the date of birth, which is measured in quarter of births due to confidentiality reasons by ONS. Second, the cut-off represents the score after the treatment is given. Individuals born after 1st April 1933 for the first reform and 1st September 1957 for the second reform had an additional year in compulsory education. Third, the treatment is affected by this increase in years of compulsory education, so the probability of treatment assignment based on the running variable has discontinuous changes after the reform implementation. Finally, the accuracy of local polynomial estimation depends on the bandwidth around the cut-off, therefore it is crucial that the bandwidth estimation procedure is conducted in a data-driven way, such that the trade-off between bias (when the bandwidth is large) and variance (when the bandwidth is small) is minimised.

The estimated treatment effect represents intent-to-treat (ITT), which is the effect of reform eligibility defined by the threshold rule (born after 1st April 1933 or 1st September 1957) on cancer outcomes. ITT uses the random assignment rather than the treatment actually received, in this case having one more year of compulsory schooling. Therefore, ITT is the average impact of the policy reform that increased age at school leaving: ITT averages the causal effect for those who actually stay one additional year in compulsory education and those who do not. It would be interesting to look at the causal effect on those who comply to the policy reform. However, the ITTs can only be estimated, since the actual age at which they left school is not available from census data. ITTs represent conservative estimates because they include everyone who was born after the threshold defined by the compulsory school law, regardless of whether they remained in school according to the threshold rule. Therefore, ITT does not take into account non-compliance, which can be expected to be low, since the compulsory schooling reforms have had very large effects on education (Clark & Royer, 2013; Davies et al., 2018). In conclusion, the ITT measures the effect of policy assignment rather than the effect of the policy itself, but it allows the inclusion of all individuals in the sample. By reporting the ITT it is possible to estimate the effect of creating a policy about increasing compulsory schooling on cancer outcomes. The analysis is divided in two parts, first analysis of the effect of education for cancer incidence, and second, for mortality and survival after cancer diagnosis.

In order to study the effect of education on cancer incidence, the outcomes were divided into three categories: dead without any cancer diagnosis, diagnosed with cancer and alive at the end of the observation period without experiencing cancer (base category). Moreover, the same analysis was replicated by dividing the cancer incidence into preventable and unpreventable cancers, so creating four categories: dead without any cancer diagnosis, diagnosed with a preventable cancer, diagnosed with unpreventable cancer and alive at the end of the observation period without experiencing cancer (base category). In addition, the effect of education on mortality from cancer was analysed in two different ways. First, at mortality after cancer diagnosis by breaking down the outcome in three categories for each individual: dead within 5 years from the cancer diagnosis, dead between 5 and 10 years from the cancer diagnosis and being alive 10 years after cancer diagnosis (base category). This classification takes into account the fact that cancer patients have higher overall mortality, given the aggressiveness of treatment and psychosocial factors related to cancer diagnosis. Second, three other mutual exclusive categories were considered to estimate the effect on cancer mortality: dead with cause of death cancer, dead with any other cause of death, and alive (base category).

The decision not to estimate survival models was motived by clearer methodological features and the development in the RD with categorical measures. Despite the increasing interest in applying the methods, few studies have used RD designs in combination with survival models (Bor, Moscoe, Mutevedzi, Newell, & Bärnighausen, 2014; Moscoe, Bor, & Bärnighausen, 2015; Venkataramani, Bor, & Jena, 2016). Indeed, regression discontinuity within the survival framework poses some challenges for the local polynomial methods used for approximating the regression functions near the cut-off and it is unclear how to find the optimal bandwidth, which minimises the bias-variance trade-off following this method.

As suggested by Xu (2017), the RD with categorical variables approach is considered superior to the simple binary outcome in a RD framework for several reasons. First, when a binary outcome is used, the selection procedure for the bandwidth is only suboptimal, and the bias corrector and robust standard error can no longer be used. Second, many studies have tried to overcome these issues by aggregating over different bins of the running variable, so that fractions become the outcome; among others, Clark and Royer (2013) have also applied this methodology. However, this procedure could influence window size and RD estimates in an unknown way. Finally, by using binary outcomes, it is not possible to generate simultaneous inference across categories. This methodology allows to jointly test across categories, which is extremely important in this application on cancer, given the presence of competing risks. The running variable is discrete, since it is provided by ONS in quarter of birth for confidentiality reasons. Previous studies have shown that in the presence of discrete running variables, it is still possible to apply continuity-based RD methods (Oreopoulos, 2006). However, recent evidence has cautioned about the use of cluster robust standard errors in this case, since it underestimates statistical uncertainty, leading to biased results (Kolesár & Rothe, 2018). Indeed, robust standard errors are larger than cluster robust standard errors when RD is estimated on small and moderate windows, so the effects are estimated using only heteroskedastic-robust standard errors.

A series of robustness checks were also performed to confirm the results. The analysis was repeated by varying the bandwidth (one year, one year and half, three years), then, the ITTs were estimated using a unique binary variable using the rd.robust package in Stata (Calonico, Cattaneo, Farrell, & Titiunik, 2017). The main analyses for the RD with categorical data were conducted in R using the code provided by Xu (2017).

4.3.3 Descriptive statistics

The analytical sample included two groups affected by the educational reforms in 1947 and 1972. The first group comprises 109,214 individuals born between 1926 and 1940, and the second group consists of 152,376 individuals born between 1949 and 1964. The date of birth is available by quarter of birth, and in the case of the second reform, the birth year refers to the school year (i.e. year starting in September and ending in August). The first cohorts subject to the reforms are April 1933 and September 1957, given the aggregation by quarter of birth and the different classification of birth year across the two cohorts, the first cohort to be affected in the 1947 reform is the second quarter of birth in 1933, while in the 1972 reform it is the first quarter of birth in 1957.

Table 1 presents the main characteristics of the two samples. The sample size one year before and one year after is 13,886 and 20,522, respectively, therefore the sample size to evaluate the second educational reform is very similar to that of Davies and colleagues (2018). In the 1926–1940 subsamples, the number of individuals affected by cancer between 1971 and 2015 (only first cancer registration is included) was 36,543 and among those, 17,805 died from cancer. In the 1949–1964 subsamples, the number of cancer registrations was 15,086 and among those, there were 3,318 cancer deaths. Deaths from cancer represent around 35 per cent of all deaths in both samples, with the number of cancer registrations being much higher in the 1926-1940 subsample, since it has an older population structure. Table 1 also reports educational attainment in the two samples, separated for males and females. The different educational distribution is evident in the two subsamples, most of the first subsample have no qualifications, while those in the second sample are spread more equally across the different educational groups. Table 2 presents cancer incidence by site for each subsample, with most cancers concentrated in a few certain sites, such as breast (6,249), lung (4,919), colon-rectum (4,919), and prostate (3,799).

Table 1 Descriptive Statistics

	1926–1940 Sample		1950–1964 Sample	
	Male	Female	Male	Female
N. observations	55,271	53,943	77,811	74,565
Cancer registrations	19,448	17,095	5,535	9,551
Age at cancer (per cent, standard deviation (SD) in parenthesis)	69.2 (9.5)	67.5 (11.05)	51.1 (9.4)	46.1 (10.5)
Cancer deaths	9,782	8,023	1,583	1,735
Age at cancer death (per cent, SD in parenthesis)	68.5 (9.6)	68.5 (10.5)	50.2 (9.9)	49.5 (10)
Deaths cancer-free	16,195	11,751	3,950	1,755
Any death	28,974	21,980	5,693	3,680
Less than qualification (per cent)	77.6	83.5	24.1	25.3
Secondary (per cent)	12.7	11.6	44.6	45.4
Tertiary (per cent)	7.9	3.1	20.4	20.3
Missing education (per cent)	1.7	1.8	11	9

Source: ONS Longitudinal Study.

Table 2 Cancer Incidence by Cancer Site and Sex

Cancer Site	1926–1940 Sample		1950–1964 Sample	
	Male	*Female*	*Male*	*Female*
Lung	2,727	1,554	348	290
Colon-rectum	2,014	1,507	509	382
Pancreas	389	356	85	59
Bladder	862	299	124	30
Kidney	472	210	183	99
Oral	259	129	257	92
Oesophagus	435	222	135	37
Melanoma	266	284	233	273
Hodgkin	51	28	75	54
Larynx	196	43	64	12
Leukaemia	346	249	137	102
Brain	273	186	148	89
Liver	213	99	71	25
Prostate	3,190		609	
Testis	75		206	
Breast		3,823		2,426
Uterus		687		281
Ovary		679		294
Cervix		315		315

Source: ONS Longitudinal Study.

These educational reforms have affected a large part of the population. The reforms occurred in different contexts, hence the educational distribution in the two samples is quite diverse with very few people with a higher education in the first subsample. In Figure 1 and Figure 2, the proportion of those attaining no qualification is presented across quarter of birth cohorts. Unfortunately, the age at leaving school is not available from the census. Other studies with detailed information about school leaving age found that the reforms had a clear effect in producing a discontinuous increase in school leaving age across birth cohorts (Clark & Royer, 2013; Davies et al., 2018). The first reforms appear to have only increased the age at leaving school, while the second reforms produced a discontinuous change in the proportion of individuals with no qualifications, probably due to the closer proximity of compulsory schooling age (16) to qualifications, such as General Certificate of Education (GCE) Ordinary Level, CSE, and GCSE (Braakmann, 2011).

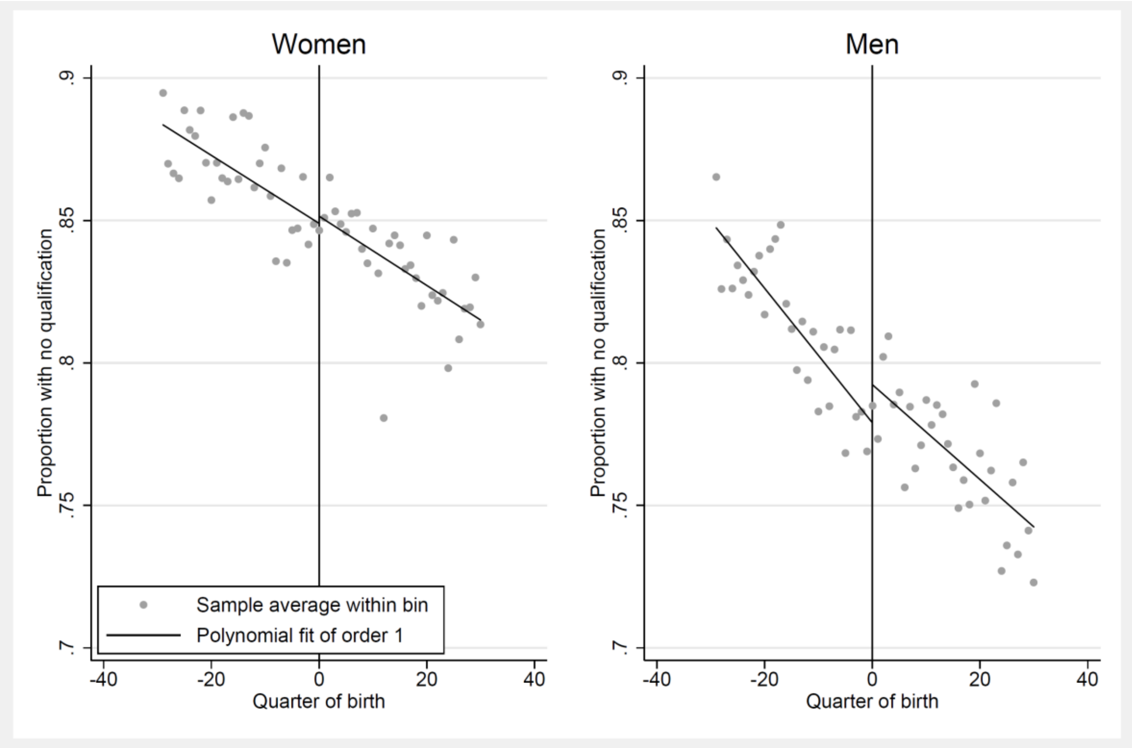

Figure 1 First stage estimates examining the impact of the compulsory schooling changes on less than a qualification for the 1947 reform. *Source:* ONS Longitudinal Study.

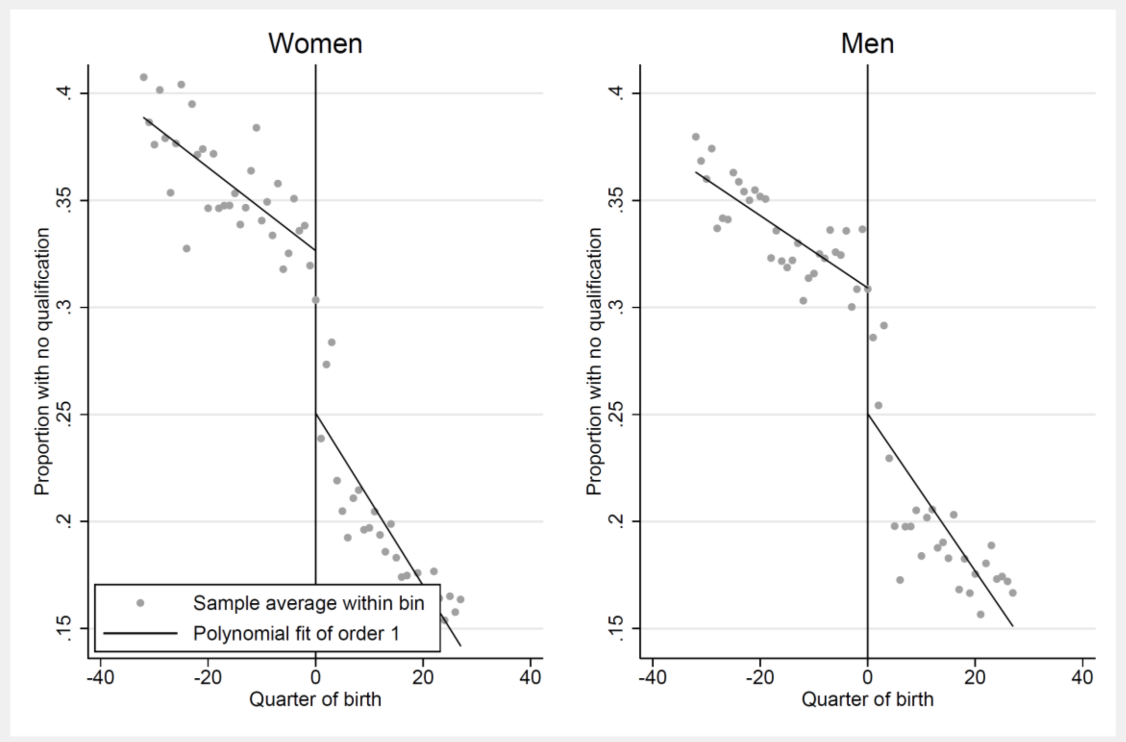

Figure 2 First stages estimates examining the impact of the compulsory schooling changes on less than a qualification for the 1972 reform. *Source:* ONS Longitudinal Study.

4.4 Results

4.4.1 The association between education and cancer

Linear probability models were estimated where the dependent variables are whether an individual gets cancer, whether the cancer is preventable or not, and whether or not an individual dies from cancer, as presented in Table 3 and Table 4. A clear educational gradient is evident for cancer incidence and cancer mortality, with better educated individuals having a lower probability of getting cancer, as well as dying from cancer. However, the associations appear to be weaker in the case of cancer incidence, where only tertiary education is associated with a lower probability of getting cancer. More consistent associations across the educational spectrum are evident regarding the incidence of the ten most preventable cancers. Men, in particular, are less likely to get one of the ten most preventable cancers as their educational level increases. For example, having tertiary education for men significantly decreases the probability of getting one of the preventable cancers by 5.25 per cent in the first subsample and by 0.6 per cent in the second subsample. For women, this gradient in preventable cancer is much less evident. Similarly, for cancer mortality after diagnosis, tertiary education decreases the probability of cancer mortality, with the decrease ranging from 1.1 per cent to 5.4 per cent.

The associations between education and cancer incidence for the most common cancer sites were examined and are shown in Table 5 and Table 6. As expected from previous literature (Sidorchuk et al., 2009), lung cancer incidence is lower for individuals with secondary and tertiary education compared to no qualifications. The decrease in probability ranges between 0.3 per cent for females with secondary education in the second subsample and 3.9 per cent for males with tertiary education in the first subsample. Moreover, no significant association was found between colorectal cancer incidence and education in both subsamples. Previous studies have shown the presence of a socioeconomic gradient in colorectal cancer risk in US and Canada, but to a much lower extent in Europe (Aarts et al., 2010; Manser & Bauerfeind, 2014). However, to my knowledge, there are no previous studies on the association between colorectal cancer incidence and education in England and Wales. These results suggest the absence of an educational gradient in colorectal cancer in England and Wales.

Prostate and breast cancers show a positive association with education, with higher educated individuals having a higher probability of prostate and breast cancers. In particular, for the most educated men, the probability of prostate cancer increased by 2.2 per cent in the first time period and breast cancer by 0.5 per cent in the second time period compared to those with no qualifications. Previous studies have suggested different reasons for this positive relationship between these cancer sites and education

(Palme & Simeonova, 2015; Steenland et al., 2002), including participation in screening programmes and overdiagnosis of low-risk prostate cancers.

Table 3 Linear Probability Model for Education and Cancer Incidence and Mortality after Diagnosis (1926–1940 Sample)

	Cancer at Any Age		10 Most Preventable Cancers		Cancer Death	
	Male	Female	Male	Female	Male	Female
Education (reference = No Qualification)						
Secondary Education	-0.00289	-0.00514	-0.0308***	-0.00981+	-0.0420***	-0.0306***
	(0.00718)	(0.00522)	(0.00402)	(0.00498)	(0.00495)	(0.00365)
Tertiary Education	-0.0144+	-0.0379**	-0.0521***	-0.0257**	-0.0514***	-0.0542***
	(0.00850)	(0.0120)	(0.00512)	(0.00950)	(0.00539)	(0.00720)
Birth cohort (reference = 1926–1930)						
1930–1935	-0.0357***	-0.0316***	-0.0329***	-0.0227***	-0.0462***	-0.0431***
	(0.00555)	(0.00458)	(0.00371)	(0.00322)	(0.00535)	(0.00402)
1936–1940	-0.0851***	-0.0674***	-0.0553***	-0.0458***	-0.0918***	-0.0805***
	(0.00720)	(0.00543)	(0.00349)	(0.00354)	(0.00535)	(0.00427)
Constant	0.396***	0.353***	0.179***	0.189***	0.234***	0.195***
	(0.00350)	(0.00363)	(0.00241)	(0.00241)	(0.00378)	(0.00298)
N	54,326	52,991	54,326	52,991	54,326	52,991

Note: Standard errors in parentheses, + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. *Source:* ONS Longitudinal Study.

Table 4 Linear Probability Model for Education and Cancer Incidence and Mortality after Diagnosis (1950–1964 Sample)

	Cancer at Any Age		10 Most Preventable Cancers		Cancer Death	
	Male	Female	Male	Female	Male	Female
Education (reference = No Qualification)						
Secondary Education	0.00305	-0.00312	-0.00278^{+}	-0.00228	-0.0109***	-0.0167***
	(0.00285)	(0.00347)	(0.00147)	(0.00201)	(0.00150)	(0.00226)
Tertiary Education	-0.00659*	-0.0171***	-0.00666***	-0.00405^{+}	-0.0148***	-0.0192***
	(0.00289)	(0.00370)	(0.00170)	(0.00204)	(0.00162)	(0.00220)
Birth cohort (reference = 1950–1954)						
1955–1959	-0.0469***	-0.0408***	-0.0142***	-0.0241***	-0.0162***	-0.0162***
	(0.00520)	(0.00417)	(0.00220)	(0.00262)	(0.00218)	(0.00185)
1960–1964	-0.0771***	-0.0633***	-0.0256***	-0.0374***	-0.0254***	-0.0233***
	(0.00463)	(0.00395)	(0.00190)	(0.00249)	(0.00173)	(0.00166)
Constant	0.117***	0.175***	0.0395***	0.0791***	0.0427***	0.0490***
	(0.00436)	(0.00388)	(0.00200)	(0.00210)	(0.00197)	(0.00215)
N	69,277	67,901	69,277	67,901	69,277	67,901

Note: Standard errors in parentheses, + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. *Source:* ONS Longitudinal Study.

Table 5 Linear Probability Model for Education and Cancer Incidence by Sites (1926–1940 Sample)

	Lung		Colon-rectum		Prostate	Breast
	Male	Female	Male	Female	Male	Female
Education (reference = No Qualification)						
Secondary Education	-0.0244^{***}	-0.0168^{***}	-0.00159	-0.00117	0.0131^{***}	0.0144^{***}
	(0.00193)	(0.00172)	(0.00252)	(0.00199)	(0.00338)	(0.00367)
Tertiary Education	-0.0386^{***}	-0.0194^{***}	-0.00125	-0.00228	0.0217^{***}	0.000116
	(0.00239)	(0.00316)	(0.00277)	(0.00426)	(0.00462)	(0.00822)
Birth cohort (reference = 1926–1930)						
1930–1935	-0.0173^{***}	-0.00614^{**}	-0.00434^{*}	-0.00716^{**}	0.00230	-0.00218
	(0.00248)	(0.00183)	(0.00199)	(0.00219)	(0.00225)	(0.00239)
1936–1940	-0.0239^{***}	-0.0144^{***}	-0.0109^{***}	-0.0151^{***}	-0.00141	-0.00211
	(0.00207)	(0.00188)	(0.00180)	(0.00181)	(0.00212)	(0.00219)
Constant	0.0695^{***}	0.0384^{***}	0.0421^{***}	0.0356^{***}	0.0545^{***}	0.0712^{***}
	(0.00181)	(0.00125)	(0.00145)	(0.00143)	(0.00126)	(0.00165)
N	54,326	52,991	54,326	52,991	54,326	52,991

Note: Standard errors in parentheses, + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. *Source:* ONS Longitudinal Study.

Table 6 Linear Probability Model for Education and Cancer Incidence by Sites (1950–1964 Sample)

	Lung		Colon-rectum		Prostate	Breast
	Male	Female	Male	Female	Male	Female
Education (reference = No Qualification)						
Secondary Education	-0.00369***	-0.00335***	0.000949	-0.000354	0.00273***	0.00431**
	(0.000771)	(0.000714)	(0.000944)	(0.000753)	(0.000780)	(0.00151)
Tertiary Education	-0.00481***	-0.00462***	-0.000309	-0.000813	0.00308**	0.00544**
	(0.000720)	(0.000716)	(0.000998)	(0.000786)	(0.00103)	(0.00186)
Birth cohort (reference = 1950–1954)						
1955–1959	-0.00338***	-0.00335***	-0.00436***	-0.00260***	-0.00976***	-0.0132***
	(0.000703)	(0.000725)	(0.00118)	(0.000651)	(0.00176)	(0.00221)
1960–1964	-0.00588***	-0.00489***	-0.00860***	-0.00495***	-0.0149***	-0.0207***
	(0.000593)	(0.000711)	(0.00107)	(0.000714)	(0.00162)	(0.00204)
Constant	0.0107***	0.00962***	0.0109***	0.00835***	0.0148***	0.0427***
	(0.000840)	(0.000877)	(0.00108)	(0.000689)	(0.00160)	(0.00176)
N	69,277	67,901	69,277	67,901	69,277	67,901

Note: Standard errors in parentheses, + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Source:* ONS Longitudinal Study.

The proportion of people without a qualification decreased after the educational reforms. Indeed, in the two samples the proportion of people with no qualifications decreased by 4.7 and 3.4 per cent for men and women, respectively, after the 1947 educational reform, and 13.7 and 16.3 per cent for men and women, respectively, after the 1972 reform. Therefore, the change in proportion of those without any educational qualification is much more marked after the 1972 reform. Table A1 in Appendix presents the linear probability models using only a binary variable for not having an educational qualification. A rough calculation of the impact of the reforms on cancer incidence and mortality is represented by the multiplication of the coefficients in the linear probability models with the changing proportion of individuals without an educational qualification. Comparing the first subsample with the second one, it appears that despite observing a larger reduction in the probability of cancer incidence and mortality, the decrease in the proportion of people without a qualification is much lower. For example, the changes in proportion of people without qualification leads to changes in the probability of cancer death after the 1947 educational reform by 0.2 per cent for men and 0.1 per cent for women in the whole period. Conversely, the reduction in proportion of people without educational qualification is larger for the second subsample but the probability of cancer incidence and death is also much lower given the younger age structure. Therefore, the probability of cancer incidence after the 1972 reform changes only by 0.2 per cent for men and 0.3 per cent for women as consequence of the reduction of people without educational qualifications. In conclusion, the magnitude of the changes in probability of different cancer outcomes seems to be small given the changes in proportion of people without an educational qualification.

4.4.2 The effect of education on cancer

Overall, education is associated with a lower probability of cancer, as well as a higher mortality from cancer. However, the causal role of education needs to be explored. The results presented in Tables 7–11 show the ITT estimates, which are the effect of treatment eligibility for cancer incidence, survival and mortality.

Table 7 shows that education does not appear to be causally related to cancer incidence following the two reforms. Being eligible for the 1947 reform does not significantly decrease the probability that individuals would experience cancer (Panel B). Similarly, Panel A presents the results from the competing risk of dying without experiencing cancer, showing that being eligible for any of the two reforms does not imply a lower probability of death with no cancer. Moreover, Panel C presents an insignificant joint test of significance of education across outcomes (death and cancer), showing that education reduces neither cancer-free mortality nor cancer incidence. In addition, the equivalence of effects for mortality without cancer and cancer incidence is not rejected, it is not possible to distinguish the two effects. These results suggest an absence of a causal role of one additional year of compulsory schooling for the probability of getting cancer for both reforms. When the analysis was repeated distinguishing between preventable and unpreventable cancers, similarly there was a lack of evidence for a causal effect of an additional year of compulsory schooling on the incidence of preventable cancers (Table 8). However, in the case of the second reform, education significantly increased the probability of a preventable cancer for men, but once the bias was corrected for using robust standard error, the coefficient lost significance. Similar results are evident in Table 11 by examining three specific cancer sites, such as lung, prostate and breast.

The effect of education on survival after cancer diagnosis is shown in Table 9. Short-term and long-term mortality were compared after diagnosis for the two reforms, with education causally decreases the probability of death between 5 and 10 years after cancer diagnosis among women affected by the 1947 reform. However, the significance of ITT disappeared when robust standard errors were computed. Moreover, no effect on survival was evident for individuals remaining in school as a consequence of the 1972 reform for both men and women. Similarly, the joint significance and equivalence of the effect of education across short-term and long-term mortality after cancer diagnosis disappears when robust standard errors were included. In conclusion, the results suggest weak evidence for a causal effect of an additional year of compulsory schooling after the two reforms on survival after cancer diagnosis when robust standard errors were considered. Finally, Table 10 presents the results for mortality with cancer as a main underlying cause of death taking into account mortality from other causes. Therefore, the results are informative of mortality from cancer, rather than survival after a cancer diagnosis. Taken together, also in this case one additional year of compulsory schooling is not causally related to mortality from cancer after both reforms.

The reported results were calculated using an optimal bandwidth calculated by optimising the joint response probabilities of different categories, the choice of an optimal bandwidth allows a single ITT estimate. The optimal bandwidths are symmetrical and range between 13 and 21 quarters depending on the model. However, the robustness of the results is checked with different bandwidths, particularly by using four, six and twelve quarters (one year, one year and half, and three years before and after the threshold) and the results remained consistent. The robustness was also tested using a simple dichotomous indicator for cancer incidence and mortality using a conventional RD design and the results are presented in Appendix (Table A21–A5) and are consistent with the presented estimates.

Table 7 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Incidence

	1926–1940 Sample		1950–1964 Sample	
	Male	Female	Male	Female
Optimal Bandwidth (N. of quarters)	13.3	13.1	15.9	15.0
Reference: Alive without cancer				
Panel A: Death without cancer				
ITTs	0.010	-0.010	-0.002	0.000
t-test P-value	0.570	0.085+	0.941	0.982
Robust t-test P-value	0.725	0.269	0.968	0.988
Panel B: Cancer				
ITTs	0.006	0.001	0.003	0.001
t-test P-value	0.248	0.808	0.264	0.493
Robust t-test P-value	0.475	0.876	0.547	0.656
Panel C: Joint Tests				
Wald-test (Significance) P-value	0.226	0.213	0.530	0.791
Robust Wald-test (Significance) P-value	0.567	0.529	0.832	0.905
Wald-test (Equivalence) P-value	0.715	0.260	0.446	0.538
Robust Wald-test (Equivalence) P-value	0.821	0.470	0.682	0.689

Notes: Standard errors in parentheses, + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. *Source:* ONS Longitudinal Study.

Table 8 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Incidence by Preventability

	1926–1940 Sample		1950–1964 Sample	
	Male	Female	Male	Female
Optimal Bandwidth (N. of quarters)	15.1	17.6	13.8	12.1
Reference: Alive without cancer				
Panel A: Death without cancer				
ITTs	0.015	**-0.014**	0.000	-0.002
t-test P-value	0.310	**0.024***	0.523	0.525
Robust t-test P-value	0.584	**0.297**	0.694	0.648
Panel B: Cancer preventable				
ITTs	0.006	0.006	**0.006**	0.001
t-test P-value	0.225	0.313	**0.012***	0.371
Robust t-test P-value	0.513	0.644	**0.126**	0.521
Panel C: Cancer non-preventable				
ITTs	-0.001	0.008	-0.002	-0.002
t-test P-value	0.561	0.259	0.749	0.567
Robust t-test P-value	0.754	0.606	0.844	0.681
Panel D: Joint Tests				
Wald-test (Significance) P-value	0.119	0.111	0.078	0.692
Robust Wald-test (Significance) P-value	0.640	0.734	0.466	0.861
Wald-test (Equivalence) P-value	0.938	0.050*	0.246	0.499
Robust Wald-test (Equivalence) P-value	0.982	0.527	0.589	0.700

Notes: Standard errors in parentheses, + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. *Source:* ONS Longitudinal Study.

Table 9 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Survival After Cancer Diagnosis

	1926–1940 Sample		1950–1964 Sample	
	Male	Female	Male	Female
Bandwidth (N. of quarters)	18.6	21.0	22.5	12.3
Reference: Alive after 10 years				
Panel A: Death after cancer diagnosis within 5 years				
ITTs	0.018	-0.026	-0.006	0.010
t-test P-value	0.129	0.081+	0.516	0.316
Robust t-test P-value	0.430	0.385	0.781	0.569
Panel B: Death after cancer diagnosis between 5 and 10 years				
ITTs	-0.005	**-0.020**	-0.004	0.004
t-test P-value	0.828	**0.002****	0.956	0.420
Robust t-test P-value	0.910	**0.132**	0.981	0.650
Panel C: Joint Tests				
Wald-test (Significance) P-value	0.307	0.000	0.810	0.413
Robust Wald-test (Significance) P-value	0.727	0.149	0.962	0.752
Wald-test (Equivalence) P-value	0.503	0.974	0.542	0.622
Robust Wald-test (Equivalence) P-value	0.519	0.998	0.794	0.780

Notes: + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. *Source:* ONS Longitudinal Study.

Table 10 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Mortality

	1926–1940 Sample		1950–1964 Sample	
	Male	Female	Male	Female
Bandwidth (N. of quarters)	16.8	17.7	14.3	13.1
Reference: Alive				
Panel A: Death from cancer				
ITTs	0.008	-0.002	-0.011	0.017
t-test P-value	0.981	0.669	0.453	0.193
Robust t-test P-value	0.988	0.786	0.615	0.381
Panel B: Death from other causes				
ITTs	0.005	-0.018	-0.013	-0.002
t-test P-value	0.311	0.086+	0.141	0.987
Robust t-test P-value	0.504	0.270	0.326	0.991
Panel C: Joint Tests				
Wald-test (Significance) P-value	0.528	0.226	0.221	0.427
Robust Wald-test (Significance) P-value	0.759	0.542	0.509	0.681
Wald-test (Equivalence) P-value	0.624	0.250	0.936	0.234
Robust Wald-test (Equivalence) P-value	0.747	0.462	0.957	0.423

Notes: + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. *Source:* ONS Longitudinal Study.

Table 11 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Incidence by Site

	Lung				Prostate		Breast	
	1926–1940 Sample		1950–1964 Sample		1926–1940 Sample	1950–1964 Sample	1926–1940 Sample	1950–1964 Sample
	Male	Female	Male	Female	Male	Male	Female	Female
Optimal Bandwidth (N. of quarters)	20.3	16.8	15.2	15.5	14.4	15.8	18.5	9.8
Reference: Alive without cancer								
Panel A: Death without cancer								
ITTs	0.008	-0.012	-0.002	-0.001	0.012	-0.010	-0.002	0.000
t-test P-value	0.824	0.098	0.816	0.564	0.725	0.072	0.249	0.982
Robust t-test P-value	0.929	0.375	0.895	0.668	0.854	0.307	0.653	0.986
Panel B: Cancer by site								
ITTs	0.001	-0.003	0.000	0.001	-0.002	0.009	**-0.003**	0.002
t-test P-value	0.854	0.245	0.639	0.626	0.910	0.163	**0.043***	0.198
Robust t-test P-value	0.941	0.533	0.790	0.725	0.953	0.435	**0.403**	0.338
Panel C: Other cancers sites								
ITTs	0.002	0.013	0.002	0.002	0.008	-0.001	0.003	-0.002
t-test P-value	0.510	0.075	0.364	0.429	0.308	0.881	0.601	0.479
Robust t-test P-value	0.792	0.342	0.607	0.558	0.593	0.932	0.840	0.596
Panel D: Joint Tests								
Wald-test (Significance) P-value	0.819	0.124	0.767	0.764	0.601	0.203	0.128	0.561
Robust Wald-test (Significance) P-value	0.985	0.648	0.947	0.890	0.917	0.690	0.816	0.767
Wald-test (Equivalence) P-value	0.872	0.090	0.747	0.609	0.628	0.100	0.481	0.399
Robust Wald-test (Equivalence) P-value	0.978	0.502	0.910	0.761	0.881	0.480	0.896	0.599

Notes: + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001. *Source:* ONS Longitudinal Study.

4.5 Conclusions

The analyses conducted in this study provide a comprehensive view of the causal effect of education on cancer incidence, survival and mortality using quasi-experimental designs in England and Wales. The natural experiments exploit the exogenous increase in education subsequent to the 1947 and 1972 reforms. One of the main contributions of this work is the examination, for the first time, of the effect of education on survival after cancer diagnosis. Overall, there is no evidence for a protective effect of one additional year of compulsory schooling on cancer outcomes following both educational reforms. Similarly, little evidence for the causal effect of education on cancer risk has been found in the literature (Davies et al., 2018; Leuven et al., 2016), but some previous studies have shown consistent lower mortality from overall cancer and lung cancer after some educational reforms in US and Sweden (Glied & Lleras-Muney, 2008; Lager & Torssander, 2012). The null results from this study are consistent with previous research that reported null effects of education on health, healthy behaviours and health knowledge in England and Wales after the two analysed reforms (Braakmann, 2011; Clark & Royer, 2013; Johnston, Lordan, Shields, & Suziedelyte, 2015).

This study contributes to the existing literature in at least two important ways. First, it provides additional evidence on the causal effect of education on health, focusing on cancer outcomes. Mixed evidence has been found concerning the effect of an increase in compulsory schooling on health (for review, see Mazumder 2011). The current work offers a more in-depth analysis by examining the effect on cancer risk and survival in England and Wales in two different historical periods. Previous studies have not found unanimous consensus regarding the effect of the increase in compulsory schooling after the 1947 and 1972 reforms. Clark and Royer (2013) found a null effect of the reforms on mortality and several health outcomes, whereas there is other evidence supporting the protective role of education for different health outcomes following the two reforms (Davies et al., 2018; Silles, 2009). The results of the present study suggest the absence of a causal effect of an additional year of compulsory schooling for cancer incidence, mortality, and survival following the two compulsory school reforms.

Second, this analysis offers weak evidence of a causal effect of a one-year increase in compulsory schooling following both the 1947 and 1972 reforms on a range of cancer-related outcomes such as incidence, survival, and mortality. In particular, the causal effect of education on survival after cancer diagnosis was investigated for the first time. To date, studies regarding the effects of education on cancer outcomes have focused mostly on Nordic countries, where the compulsory school reforms were accompanied by other changes in the educational systems, such as students' tracking (Lager & Torssander, 2012; Leuven et al., 2016; Palme & Simeonova, 2015). In a recent review, Galama and colleagues (2018) suggested that changes in students' tracking systems, rather than increases in years of schooling, affected smoking prevalence. This could be a potential explanation for the different effects found concerning education on cancer incidence in Nordic countries compared to England and Wales. Similar to our work, Davies and colleagues (2018) found no effect of education on cancer incidence due to the 1972 educational reform using UK Biobank data.

The finding of no protective effects of one additional year of compulsory schooling on cancer outcomes could be expected for several reasons. First, the historical and social context where the reforms took place might have prevented the emergence of educational inequality in cancer. The universal healthcare coverage of NHS in England and Wales, as well as the little diffusion of information about cancer prevention, might have played a role for the absence of results. The *fundamental cause* perspective suggests the presence of information and resources that more educated individuals could utilise to gain health benefits (Phelan et al., 2004). Montez and Friedman (2015) have also highlighted the context dependence for the causal effect of education on health. The emergence of causal association is highly dependent on the socioeconomic and institutional context, so those affected by the 1947 reform experienced a disruption of their education by the war. Moreover, the harmful effect of smoking has become widely known only after the 1964 Surgeon's General Report and in the UK specifically in 1957, “Tobacco smoking and cancer of the lung. Statement by the Medical Research Council” (Alberg et al., 2014). Therefore, those affected by the 1947 reform have spent at least some part of their adult life without being completely aware of the health hazards deriving from smoking. Concerning other risk factors, Davies and colleagues (2018) found that the 1972 reform decreased alcohol consumption and sedentary behaviours. However, these increases in preventative behaviours do not seem to be large enough to be reflected on cancer outcomes.

Furthermore, a large degree of heterogeneity exists in cancer aetiology and survival by site. Cancer includes different types of illnesses and it is unclear in which way risk factors influence cancer outcomes in various sites. Therefore, the heterogeneity by cancer site might prevent the identification of a unique effect of education on cancer. In the present study, we tried to distinguish between preventable and unpreventable cancers. Despite the presence of an element of chance, scholars advocate for the important leading role of primary prevention for cancer, prevention is considered the relevant channel for decreasing the burden of disease (Ashford et al., 2015; Colditz & Wei, 2012; Marmot, 2018). However, the *social determinants of health* framework require increasing attention to the *causes of the causes*. Indeed, the pattern of risky health behaviours might be determined by social environment and involuntary exposure to carcinogens. For example, the social environment in which individuals grow up and live are not often a matter of personal choice, and some unhealthy behaviours might represent ways of coping with stress (Baum, Garofalo, & Yali, 1999). Similarly, the lower educated might have a higher occupational exposure to harmful substances independently of their individual lifestyle choices. However, Boffetta (2004), after reviewing studies on occupational and environmental causes of cancer, concluded that they were attributable to only 1–2 per cent of total cancers.

Moreover, the reforms might not be a good approximation for estimating the causal effect of education for three reasons. First, education has increased equally for all individuals in the same generation, second in a specific point in the educational distribution (age at school left from 14 to 15, and from 15 to 16) and third, the increase in years of compulsory schooling was not directly reflected in a higher likelihood of obtaining educational qualifications. Firstly, this exogenous increase in the education of a whole generation is not necessarily connected to a higher socioeconomic position in society because many are being lifted simultaneously. Secondly, the reforms have raised schooling in the middle of the educational distribution (from 14 to 15 in the 1947 reform and from 15 to 16 in the 1972 reform). Therefore, the absence of causality is to be understood locally and it is not possible to exclude that there might be the presence of a causal effect of education on cancer at different points of the educational distribution. For example, Buckles and colleagues (2013) found evidence for a causal effect of college education on cancer mortality in the US. Third, the absence of an effect might be due to the fact that the reform did not lead to a higher likelihood of an educational qualification. For example, Clark and Royer (2013) have found little impact of the reforms on educational qualifications. Braakmann (2011) used the 1972 compulsory school reform in conjunction with the month of birth to study the effect of educational qualifications on health and health-related behaviours. Their identification rests on the higher probability for only those born in February (2–3 per cent more likely) to obtain the first level of qualification in England and Wales, CSE. However, even using educational qualifications, the study was unable to identify any causal effect of educational qualifications on health-related outcomes. Finally, in the English and Welsh contexts, education is not necessary the most representative measure for socioeconomic standing in society. Occupational social class could be a better indicator for socioeconomic standing, since education has decreased the importance over time for the social class positions of adults in Great Britain once their class origins are taken into account (Jackson, Goldthorpe, & Mills, 2005).

Nonetheless, the analysis is not free from limitations. To begin with, this study was unable to explore the mechanisms through which education might influence cancer outcomes, specifically, information about prevention and screening was not available. Examining health behaviours could help to test for cancer prevention. We attempted to tackle this aspect using incidence information by cancer sites, but it was still unclear how different risk factors affect cancer incidence and some of the effects might counteract each other. Moreover, cancer registries collect data on the stage of cancer at diagnosis only since 1993, thereby limiting the ability to test the mediation of screening.

Second, the different state of knowledge about cancer across sites of growth and time prevents us from testing the role of resource flexibility as predicted by the *fundamental cause* framework. For example, Lyratzopoulos et al. (2011) examined the changing socioeconomic inequalities in breast and colorectal cancer in England and Wales, highlighting the role of innovations in medical treatments in shaping inequalities in survival. Moreover, the analysis by preventability used 2015 as reference point for the most preventable cancers, but the aetiology of these cancers may have taken place when knowledge about risk factors was not established and widespread, therefore limiting the extent to which education might have mattered. Finally, a local effect at the bottom of the educational distribution was estimated, but the presence of a protective effect of education on cancer at other points of the educational distribution cannot be excluded.

A better understanding of the causes of educational inequalities in cancer is extremely important given the increasing number of people affected. Our focus on both incidence and survival provides insight on the overall relationship between education and cancer in England and Wales. This research could help inform policy-makers of the effect of increasing educational attainment for an increasingly important health outcome. Evidence regarding this area of research is still lacking, and for greater societal benefit, more research is needed to examine the relationship between cancer and education. In particular, the investigation of the role of educational policy, such as changing years of schooling and tracking systems, for health inequalities should focus on an increasingly widespread non-communicable disease like cancer.

Acknowledgements

The permission of the Office for National Statistics to use the Longitudinal Study is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information & User Support (CeLSIUS). CeLSIUS is supported by the ESRC Census of Population Programme under project ES/K000365/1. The authors alone are responsible for the interpretation of the data. This work contains statistical data from ONS, which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the

statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

References

Aarts, M. J., Lemmens, V. E. P. P., Louwman, M. W. J., Kunst, A. E., & Coebergh, J. W. W. (2010). Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. *European Journal of Cancer*, *46*(15), 2681–2695.

Alberg, A. J., Shopland, D. R., & Cummings, K. M. (2014). Epidemiology in history the 2014 surgeon general's report: commemorating the 50th anniversary of the 1964 report of the advisory committee to the US Surgeon general and updating the evidence on the health consequences of cigarette smoking. *American Journal of Epidemiology*, *179*(4), 403–412.

Albrow, R., Kitchener, H., Gupta, N., & Desai, M. (2012). Cervical screening in England: the past, present, and future. *Cancer Cytopathology*, *120*(2), 87–96.

Arendt, J. N. (2005). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review*, *24*(2), 149–160.

Ashford, N. A., Bauman, P., Brown, H. S., Clapp, R. W., Finkel, A. M., Gee, D., … Sass, J. B. (2015). Cancer risk: role of environment. *Science*, *347*(6223), 727–727.

Baum, A., Garofalo, J. P., & Yali, A. M. (1999). Socioeconomic status and chronic stress. Does stress account for SES effects on health? *Annals of the New York Academy of Sciences*, *896*, 131–144.

Boffetta, P. (2004). Epidemiology of environmental and occupational cancer. *Oncogene*, *23*(38), 6392–6403.

Boffetta, P., Hashibe, M., La Vecchia, C., Zatonski, W., & Rehm, J. (2006). The burden of cancer attributable to alcohol drinking. *International Journal of*

Cancer, *119*(4), 884–887.

Bor, J., Moscoe, E., Mutevedzi, P., Newell, M.-L., & Bärnighausen, T. (2014). Regression discontinuity designs in epidemiology. *Epidemiology*, *25*(5), 729–737.

Braakmann, N. (2011). The causal relationship between education, health and health related behaviour: evidence from a natural experiment in England. *Journal of Health Economics*, *30*(4), 753–763.

Braaten, T., Weiderpass, E., Kumle, M., & Lund, E. (2005). Explaining the socioeconomic variation in cancer risk in the Norwegian Women and Cancer Study. *Cancer Epidemiology Biomarkers and Prevention*, *14*(11 I), 2591–2597.

Braaten, T., Weiderpass, E., & Lund, E. (2009). Socioeconomic differences in cancer survival: the Norwegian Women and Cancer Study. *BMC Public Health*, *9*, 178.

Brown, K. F., Rumgay, H., Dunlop, C., Ryan, M., Quartly, F., Cox, A., … Parkin, D. M. (2018). The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *British Journal of Cancer*, (2), 1.

Brunello, G., Fabbri, D., & Fort, M. (2013). The causal effect of education on body mass: evidence from Europe. *Journal of Labor Economics*, *31*(1), 195–223.

Buckles, K., Hagemann, A., Malamud, O., Morrill, M. S., & Wozniak, A. K. (2013). *The effect of college education on health* (NBER Working Paper 8328). Cambridge, MA: National Bureau of Economic Research.

Caliendo, M., Tatsiramos, K., & Uhlendorff, A. (2013). Benefit duration, unemployment duration and job match quality: a regression-discontinuity approach. *Journal of Applied Econometrics*, *28*(4), 604–627.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust.

Software for regression-discontinuity designs. *Stata Journal*, *17*(2), 372–404.

Clark, D., & Royer, H. (2013). The effect of education on adult mortality and health: evidence from Britain. *American Economic Review*, *103*(6), 2087–2120.

Clegg, L. X., Reichman, M. E., Miller, B. A., Hankey, B. F., Singh, G. K., Lin, Y. D., … Edwards, B. K. (2009). Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. *Cancer Causes & Control : CCC*, *20*(4), 417–435.

Colditz, G. A., & Wei, E. K. (2012). Preventability of cancer: the relative contributions of biologic and social and physical environmental determinants of cancer mortality. *Annual Review of Public Health*, *33*(1), 137–156.

Coleman, M. P., Forman, D., Bryant, H., Butler, J., Rachet, B., Maringe, C., … Richards, M. A. (2011). Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the international cancer benchmarking partnership): an analysis of population-based cancer registry data. *The Lancet*, *377*(9760), 127–138.

Coleman, M. P., Rachet, B., Woods, L. M., Mitry, E., Riga, M., Cooper, N., … Estève, J. (2004). Trends and socioeconomic inequalities in cancer survival in England and Wales up to 2001. *British Journal of Cancer*, *90*(7), 1367–1373.

Cookson, R., Propper, C., Asaria, M., & Raine, R. (2016). Socio-economic inequalities in health care in England. *Fiscal Studies*, *37*(3–4), 371–403.

Crosnoe, R., & Riegle-Crumb, C. (2007). A life course model of education and alcohol use. *Journal of Health and Social Behavior*, *48*(3), 267–282.

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics, 29*(1), 1–28.

Davies, N. M., Dickson, M., Davey Smith, G., van den Berg, G. J., & Windmeijer, F. (2018). The causal effects of education on health outcomes in the UK Biobank. *Nature Human Behavior*, *2*, 117–125.

Devereux, P., & Hart, R. A. (2010). Forced to be rich? Returns to compulsory schooling in Britain. *The Economic Journal*, *120*(549), 1345–1364.

Donaldson, M. S. (2004). Nutrition and cancer: a review of the evidence for an anti-cancer diet. *Nutrition Journal*, *3*(1), 19.

dos Santos Silva, I. (2004). Time trends in cancer incidence and mortality. In D. Kuh, Y. Ben Shlomo, & S. Ezra (Eds.), *A Life Course Approach to Chronic Disease Epidemiology* (pp. 281–305). Oxford University Press.

Du, X. L., Lin, C. C., Johnson, N. J., & Altekruse, S. (2011). Effects of individual-level socioeconomic factors on racial disparities in cancer treatment and survival. *Cancer*, *117*(14), 3242–3251.

Elstad, J. I., Torstensrud, R., Lyngstad, T. H., & Kravdal, O. (2012). Trends in educational inequalities in mortality, seven types of cancers, Norway 1971-2002. *European Journal of Public Health*, *22*(6), 771–776.

Ferraro, K. F., Schafer, M. H., & Wilkinson, L. R. (2016). Childhood disadvantage and health problems in middle and later life: early imprints on physical health? *American Sociological Review*, *81*(1), 107–133.

Fletcher, J. M. (2015). New evidence of the effects of education on health in the US: compulsory schooling laws revisited. *Social Science and Medicine*, *127*, 101–107.

Foreman, K. J., Barac, A., Williams, H. C., Sufiyan, M. B., Safiri, S., Sinke, A. H., … Alizadeh-Navaei, R. (2018). Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016. *JAMA Oncology*, *4*(11), 1553.

Fort, M., Schneeweis, N., & Winter-Ebmer, R. (2016). Is education always reducing fertility? Evidence from compulsory schooling reforms. *Economic Journal*, *126*(595), 1823–1855.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., … Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*, *186*(9), 1026–1034.

Fuchs, V. (1982). Time preference and health: an exploratory study. In V. Fuchs (Ed.), *Economic Aspects of Health* (Vol. I, pp. 93–120). University of Chicago Press.

Galama, T. J., Lleras-Muney, A., & van Kippersluis, H. (2018). *The effect of education on health and mortality: a review of experimental and quasi-experimental evidence* (NBER Working Paper 24225). Cambridge, MA: National Bureau of Economic Research.

Geruso, M., & Royer, H. (2018). *The impact of education on family formation: quasi-experimental evidence from the UK* (NBER Working Paper 24332). Cambridge, MA: National Bureau of Economic Research.

Glied, S., & Lleras-Muney, A. (2008). Technological innovation and inequality in health. *Demography*, *45*(3), 741–761.

Greenwald, P., Clifford, C. K., & Milner, J. A. (2001). Diet and cancer prevention. *European Journal of Cancer*, *37*, 948–965.

Grenet, J. (2013). Is extending compulsory schooling alone enough to raise earnings? Evidence from French and British compulsory schooling laws. *Scandinavian Journal of Economics*, *115*(1), 176–210.

Gulliford, M. C. (1996). The reliability of cancer registry records. *Quality in Health*

Care, *5*, 120–127.

Gulliford, M. C., Bell, J., Bourne, H. M., & Petruckevitch, A. (1993). The reliability of cancer registry records. *British Journal of Cancer*, *67*(4), 819–821.

Hayward, M. D., & Gorman, B. K. (2004). The long arm of childhood: the influence of early-life social conditions on men's mortality. *Demography*, *41*(1), 87–107.

Hemminki, K., & Li, X. (2003). Level of education and the risk of cancer in Sweden. *Cancer Epidemiology , Biomarkers and Prevention*, *12*(August), 796–802.

Henley, N. C., Hole, D. J., Kesson, E., Burns, H. J. G., George, W. D., & Cooke, T. G. (2005). Does deprivation affect breast cancer management? *British Journal of Cancer*, *92*(4), 631–633.

Hermann, S., Rohrmann, S., Linseisen, J., Nieters, A., Khan, A., Gallo, V., … Riboli, E. (2010). Level of education and the risk of lymphoma in the European prospective investigation into cancer and nutrition. *Journal of Cancer Research and Clinical Oncology*, *136*(1), 71–77.

Ho, J. Y., & Fenelon, A. (2015). The contribution of smoking to educational gradients in U.S. life expectancy. *Journal of Health and Social Behavior*, *56*(3), 307–322.

Imbens, G. W. (2010). Better LATE than nothing: some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature*, *48*(2), 399–423.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: a guide to practice. *Journal of Econometrics*, *142*(2), 615–635.

Jackson, M., Goldthorpe, J. H., & Mills, C. (2005). Education, employers, and class mobility. *Research in Social Stratification and Mobility*, *23*, 3–33.

James, J. (2015). Health and education expansion. *Economics of Education Review*, *49*, 193–215.

Janssen-Heijnen, M., & Coebergh, J. W. W. (2003). The changing epidemiology of lung cancer in Europe. *Lung Cancer*, *41*(3), 245–258.

Jarvis, M., & Wardle, J. (2005). Social patterning of individual health behaviours: the case of cigarette smoking. In M. G. Marmot & R. G. Wilkinson (Eds.), *Social Determinants of Health* (2nd ed., pp. 224–237). Oxford University Press.

Johnston, D. W., Lordan, G., Shields, M. A., & Suziedelyte, A. (2015). Education and health knowledge: evidence from UK compulsory schooling reform. *Social Science and Medicine*, *127*, 92–100.

Kawachi, I., & Kroenke, C. (2006). Socioeconomic Disparities in Cancer Incidence and Mortality. In D. Schottenfeld & J. F. Fraumeni (Eds.), *Cancer Epidemiology and Prevention* (Fourth, Vol. 1, pp. 174–188). Oxford University Press.

Kilpelaïnen, T. P., Talala, K., Raitanen, J., Taari, K., Kujala, P., Tammela, T. L. J., & Auvinen, A. (2016). Prostate cancer and socioeconomic status in the finnish randomized study of screening for prostate cancer. *American Journal of Epidemiology*, *184*(10), 720–731.

Klassen, A. C., & Smith, K. C. (2011). The enduring and evolving relationship between social class and breast cancer burden: a review of the literature. *Cancer Epidemiology*, *35*(3), 217–234.

Kogevinas, M. (1990). *Longitudinal study: socio-demographic differences in cancer survival*. London: H.M.S.O.

Kogevinas, M., & Porta, M. (1997). Socioeconomic differences in cancer survival: a review of the evidence. *IARC Scientific Publications*, *15*(138), 177–206.

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277–2304.

Kravdal, H. (2014). Widening educational differences in cancer survival in Norway.

European Journal of Public Health, *24*(2), 270–275.

Kravdal, Ø. (2000). Social inequalities in cancer survival. *Population Studies*, *54*(1), 1–18.

Lager, A. C. J., & Torssander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proceedings of the National Academy of Sciences*, *109*(22), 8461–8466.

Landais, C. (2013). Assessing the welfare effects of unemployment assessing the welfare effects of unemployment benefits using the regression kink design. *American Economic Journal: Economic Policy*, *7*(7589), 243–278.

Lawless, L., Drichoutis, A. C., & Jr, R. M. N. (2013). Time preferences and health behaviour: a review. *Agricultural and Food Economics*, *1*(17), 1–19.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355.

Lee, D. S., & Lemieux, T. (2013). Regression discontinuity designs in social sciences. In H. Best & C. Wolf (Eds.), *The SAGE Handbook of Regression Analysis and Causal Inference* (pp. 301–326). London: SAGE Publications Ltd.

Leon, D. A. (1988). *Longitudinal study: social distribution of cancer: a report on the relationship between socio-demographic factors and the incidence of cancer, based on data collected in the OPCS Longitudinal Study*. London: H.M.S.O.

Leuven, E., Plug, E., & Rønning, M. (2016). Education and cancer risk. *Labour Economics*, *43*, 106–121.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*, *35*(Extra Issue: Forty Years of Medical Sociology: The State of the Art and Directions for the Future), 80–94.

Liu, L., Cozen, W., Bernstein, L., Ross, R. K., & Deapen, D. (2001). Changing relationship between socioeconomics status and prostate cancer incidence. *Journal of the National Cancer Institute, 93*(9), 705–709.

Lleras-muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies, 72*, 189–221.

Loeb, S., Bjurlin, M. A., Nicholson, J., Tammela, T. L., Penson, D. F., Carter, H. B., … Etzioni, R. (2014). Overdiagnosis and overtreatment of prostate cancer. *European Urology, 65*(6), 1046–1055.

Lundqvist, A., Andersson, E., Ahlberg, I., Nilbert, M., & Gerdtham, U. (2016). Socioeconomic inequalities in breast cancer incidence and mortality in Europe—a systematic review and meta-analysis. *European Journal of Public Health*, *26*(5), 804–813.

Lyratzopoulos, G., Barbiere, J. M., Rachet, B., Baum, M., Thompson, M. R., & Coleman, M. P. (2011). Changes over time in socioeconomic inequalities in breast and rectal cancer survival in England and Wales during a 32-year period (1973-2004): the potential role of health care. *Annals of Oncology*, *22*(7), 1661–1666.

Ma, J., Xu, J., Anderson, R. N., & Jemal, A. (2012). Widening educational disparities in premature death rates in twenty six states in the United States, 1993-2007. *PLoS ONE*, *7*(7), 1–10.

Mackenbach, J. P. (2012). The persistence of health inequalities in modern welfare states: the explanation of a paradox. *Social Science and Medicine*, *75*(4), 761–769.

Mackenbach, J. P., Huisman, M., Andersen, O., Bopp, M., Borgan, J. K., Borrell, C., … Kunst, A. E. (2004). Inequalities in lung cancer mortality by the educational level in 10 European populations. *European Journal of Cancer, 40*(1), 126–135.

Mackenbach, J. P., Kulhánová, I., Bopp, M., Deboosere, P., Eikemo, T. A., Hoffmann, R., … White, C. (2015). Variations in the relation between education and cause-specific mortality in 19 European populations: a test of the “fundamental causes” theory of social inequalities in health. *Social Science and Medicine*, *127*, 51–62.

Mackenbach, J. P., Stirbu, I., Roskam, A. R., Schaap, M. M., Menvielle, G., Leinsalu, M., & Kunst, A. (2008). Socioeconomic inequalities in health in 22 European countries. *The New England Journal of Medicine*, *358*(23), 2468–2481.

Manser, C. N., & Bauerfeind, P. (2014). Impact of socioeconomic status on incidence, mortality, and survival of colorectal cancer patients: a systematic review. *Gastrointestinal Endoscopy*, *80*(1), 42–60.e9.

Marmot, M. G. (2010). *Fair Society, Healthy Lives. The Marmot Review*. Retrieved from http://www.cpa.org.uk/cpa_documents/FairSocietyHealthyLivesExecSummary.pdf

Marmot, M. G. (2018). Health equity, cancer, and social determinants of health. *The Lancet Global Health*, *6*, S29.

Martikainen, P., & Valkonen, T. (2000). Diminishing educational differences in breast cancer mortality among Finnish women: a register-based 25-year follow-up. *American Journal of Public Health*, *90*(2), 277–280.

Mazumder, B. (2011). *The Effects of Education on Health and Mortality*. *Nordic Economic Policy Review*.

McCormack, V. A., & Boffetta, P. (2011). Today’s lifestyles, tomorrow’s cancers: trends in lifestyle risk factors for cancer in low- and middle-income countries. *Annals of Oncology*, *22*(11), 2349–2357.

McTiernan, A. (2008). Mechanisms linking physical activity with cancer. *Nature*

Reviews Cancer, *8*(3), 205–211.

Menvielle, G., Boshuizen, H., Kunst, A. E., Dalton, S. O., Vineis, P., Bergmann, M. M., … Bueno-De-Mesquita, H. B. (2009). The role of smoking and diet in explaining educational inequalities in lung cancer incidence. *Journal of the National Cancer Institute*, *101*(5), 321–330.

Menvielle, G., & Kunst, A. (2008). Social inequalities in cancer incidence and cancer survival: lessons from Danish studies. *European Journal of Cancer*, *44*(14), 1933–1937.

Menvielle, G., Leclerc, A., Chastang, J. F., & Luce, D. (2006). Social inequalities in breast cancer mortality among French women: disappearing educational disparities from 1968 to 1996. *British Journal of Cancer*, *94*(1), 152–155.

Montez, J. K., & Friedman, E. M. (2015). Educational attainment and adult health: under what conditions is the association causal? *Social Science and Medicine, 127*, 1–7.

Moscoe, E., Bor, J., & Bärnighausen, T. (2015). Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *Journal of Clinical Epidemiology*.

Naik, H., Qiu, X., Brown, M. C., Eng, L., Pringle, D., Mahler, M., … Liu, G. (2016). Socioeconomic status and lifestyle behaviours in cancer survivors: smoking and physical activity. *Current Oncology*, *23*(6), e546–e555.

Nechuta, S., Paneth, N., & Velie, E. M. (2010). Pregnancy characteristics and maternal breast cancer risk: a review of the epidemiologic literature. *Cancer Causes and Control*, *21*(7), 967–989.

Omran, A. R. (1971). The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly, 49*(4), 509–538.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, *96*(1), 152–175.

Palme, M., & Simeonova, E. (2015). Does women's education affect breast cancer risk and survival? Evidence from a population based social experiment in education. *Journal of Health Economics*, *42*, 115–124.

Phelan, J. C., Link, B. G., Diez-Roux, A., Kawachi, I., & Levin, B. (2004). Fundamental causes' of social inequalities in mortality: a test of the theory. *Journal of Health and Social Behavior*, *45*(1), 265–285.

Pollock, A. M., & Vickers, N. (1995). Reliability of data of the Thames Cancer Registry on 673 cases of colorectal cancer: effect of the registration process. *Quality in Health Care*, *4*(3), 184–189.

Pollock, A. M., & Vickers, N. (1998). Deprivation and emergency admissions for cancers of colorectum, lung, and breast in south east England: ecological study. *Bmj*, *317*(9322050), 245–252.

Potischman, N., Troisi, R., & Vatten, L. (2004). A life course approach to cancer epidemiology. In D. Kuh, Y. Ben Shlomo, & S. Ezra (Eds.), *A Life Course Approach to Chronic Disease Epidemiology* (pp. 260–280). Oxford University Press.

Rachet, B., Maringe, C., Nur, U., Quaresma, M., Shah, A., Woods, L. M., … Coleman, M. P. (2009). Population-based cancer survival trends in England and Wales up to 2007: an assessment of the NHS cancer plan for England. *The Lancet Oncology*, *10*(4), 351–369.

Rennie, K. L., & Jebb, S. A. (2005). Prevalence of obesity in Great Britain. *Obesity Reviews*, *6*(5), 191–192.

Rubin, M. S., Clouston, S., & Link, B. G. (2014). A fundamental cause approach to the study of disparities in lung cancer and pancreatic cancer mortality in the United States. *Social Science and Medicine, 100,* 54–61.

Sant, M., Aareleid, T., Berrino, F., Bielska Lasota, M., Carli, P. M., Faivre, J., … Steward, J. A. (2003). EUROCARE-3: survival of cancer patients diagnosed 1990–94—Results and commentary. *Annals of Oncology*, *14*(SUPPL.5), 61–118.

Sasco, A. J., Secretan, M. B., & Straif, K. (2004). Tobacco smoking and cancer: a brief review of recent epidemiological evidence. *Lung Cancer*, *45 Suppl 2*, S107–S112.

Shack, L., Jordan, C., Thomson, C. S., Mak, V., & Møller, H. (2008). Variation in incidence of breast, lung and cervical cancer and malignant melanoma of skin by socioeconomic group in England. *BMC Cancer*, *8*, 1–10.

Sidorchuk, A., Agardh, E. E., Aremu, O., Hallqvist, J., Allebeck, P., & Moradi, T. (2009). Socioeconomic differences in lung cancer incidence: a systematic review and meta-analysis. *Cancer Causes & Control*, *20*(4), 459–471.

Silles, M. a. (2009). The causal effect of education on health: evidence from the United Kingdom. *Economics of Education Review*, *28*(1), 122–128.

Sloggett, A., Young, H., & Grundy, E. (2007). The association of cancer survival with four socioeconomic indicators: a longitudinal study of the older population of England and Wales 1981-2000. *BMC Cancer*, *7*, 20.

Song, M., & Giovannucci, E. (2016). Preventable incidence and mortality of carcinoma associated with lifestyle factors among white adults in the United States. *JAMA Oncology*, *2*(9), 1154–1161.

Steenland, K., Henley, J., & Thun, M. (2002). All-cause and cause-specific death rates by educational status for two million people in two American Cancer Society cohorts, 1959-1996. *American Journal of Epidemiology*, *156*(1), 11–21.

Steenland, K., Rodriguez, C., Mondul, A., & Calle, E. E. (2004). Prostate cancer incidence and survival in relation to education (United States). *Cancer Causes and Control*, *15*(9), 939–945.

Strand, B. H., Kunst, A., Huisman, M., Menvielle, G., Glickman, M., Bopp, M., … Mackenbach, J. P. (2007). The reversed social gradient: higher breast cancer mortality in the higher educated compared to lower educated. A comparison of 11 European populations during the 1990s. *European Journal of Cancer*, *43*(7), 1200–1207.

Sueyoshi, G. (1995). A class of binary response models for grouped duration data. *Journal of Applied Econometrics*, *10*(4), 411–431.

Swanson, J. M. (2012). The UK Biobank and selection bias. *The Lancet*, *380*(9837), 110.

Vaneenwyk, J., Campo, J. S., & Ossiander, E. M. (2002). Socioeconomic and demographic disparities in treatment for carcinomas of the colon and rectum. *Cancer*, *95*(1), 39–46.

Venkataramani, A. S., Bor, J., & Jena, A. B. (2016). Regression discontinuity designs in healthcare research. *Bmj*, i1216.

Wagener, D. K., & Schatzkin, A. (1994). Temporal trends in the socioeconomic gradient for breast cancer mortality among US women. *American Journal of Public Health*, *84*(6), 1003–1006.

Wang, A., Clouston, S. A. P., Rubin, M. S., Colen, C. G., & Link, B. G. (2012). Fundamental causes of colorectal cancer mortality: the implications of informational diffusion. *The Milbank Quarterly*, *90*(3), 592–618.

Wolin, K. Y., Carson, K., & Colditz, G. A. (2010). Obesity and cancer. *The*

Oncologist, *15*(6), 556–565.

Woods, L. M., Rachet, B., & Coleman, M. P. (2006). Origins of socio-economic inequalities in cancer survival: a review. *Annals of Oncology*, *17*(1), 5–19.

Xu, K. L. (2017). Regression discontinuity with categorical outcomes. *Journal of Econometrics*, *201*(1), 1–18.

Zaninotto, P., Head, J., Stamatakis, E., Wardle, H., & Mindell, J. (2009). Trends in obesity among adults in England from 1993 to 2004 by age and social class and projections of prevalence to 2012. *Journal of Epidemiology and Community Health*, *63*(2), 140–146.

Appendix

Table A1 Linear Probability Model for Educational Qualification and Cancer Incidence and Mortality (1926–1940 Sample and 1950–1964 Sample)

	Cancer at Any Age				Cancer Death			
	1926–1940 Sample		1950–1964 Sample		1926–1940 Sample		1950–1964 Sample	
	Male	Female	Male	Female	Male	Female	Male	Female
Education (reference = Secondary Education)								
No Qualification	0.00729	0.0120*	-0.00000173	0.00751*	0.0456***	0.0356***	0.0122***	0.0175***
	(0.00541)	(0.00528)	(0.00263)	(0.00330)	(0.00322)	(0.00355)	(0.00146)	(0.00218)
Birth cohort (reference = 1926–1930)								
1930–1935	-0.0357***	-0.0315***	-0.0469***	-0.0407***	-0.0462***	-0.0431***	-0.0162***	-0.0162***
	(0.00555)	(0.00458)	(0.00519)	(0.00415)	(0.00535)	(0.00401)	(0.00217)	(0.00184)
1936–1940	-0.0851***	-0.0674***	-0.0769***	-0.0629***	-0.0918***	-0.0805***	-0.0254***	-0.0233***
	(0.00720)	(0.00543)	(0.00462)	(0.00394)	(0.00535)	(0.00426)	(0.00173)	(0.00166)
Constant	0.389***	0.341***	0.117***	0.168***	0.188***	0.160***	0.0305***	0.0315***
	(0.00544)	(0.00610)	(0.00452)	(0.00348)	(0.00413)	(0.00375)	(0.00167)	(0.00160)
N	54,326	52,991	69,277	67,901	54,326	52,991	69,277	67,901

Notes: Standard errors in parentheses, + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Source:* ONS Longitudinal Study.

Table A2 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Outcomes for Men (1926–1940 Sample)

	Cancer	Cancer Death	Death	Death from cancer
Conventional	0.0171	0.0129	0.00593	0.0129
	(0.0175)	(0.0141)	(0.0173)	(0.0141)
Bias-corrected	0.0205	0.0155	0.00555	0.0155
	(0.0175)	(0.0141)	(0.0173)	(0.0141)
Robust	0.0205	0.0155	0.00555	0.0155
	(0.0207)	(0.0167)	(0.0207)	(0.0167)
N	55,271	55,271	55,271	55,271
Effective N left	7,156	7,156	7,156	7,156
Effective N right	8,115	8,115	8,115	8,115
Bandwidth left	8.636	8.385	8.243	8.385
Bandwidth right	8.636	8.385	8.243	8.385

Notes: Standard errors in parentheses, $^{+}$ $p < 0.10$, * $p < 0.05$, standard errors in parentheses.
Source: ONS Longitudinal Study.

Table A3 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Outcomes for Women (1926–1940 Sample)

	Cancer	Cancer Death	Death	Death from cancer
Conventional	0.00706	0.00643	-0.0143	0.00643
	(0.0159)	(0.0135)	(0.0155)	(0.0135)
Bias-corrected	0.00655	0.00805	-0.0150	0.00805
	(0.0159)	(0.0135)	(0.0155)	(0.0135)
Robust	0.00655	0.00805	-0.0150	0.00805
	(0.0189)	(0.0159)	(0.0185)	(0.0159)
N	53,943	53,943	53,943	53,943
Effective N left	8,655	6,957	6,957	6,957
Effective N right	9,739	7,973	7,973	7,973
Bandwidth left	10.07	8.111	8.160	8.111
Bandwidth right	10.07	8.111	8.160	8.111

Notes: Standard errors in parentheses, $^{+}$ $p < 0.10$, * $p < 0.05$, standard errors in parentheses.
Source: ONS Longitudinal Study.

Table A4 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Outcomes for Men (1950–1964 Sample)

	Cancer	Cancer Death	Death	Death from cancer
Conventional	0.00865	0.00522	0.00219	0.00522
	(0.00769)	(0.00373)	(0.00571)	(0.00373)
Bias-corrected	0.0105	0.00636^{+}	0.00355	0.00636^{+}
	(0.00769)	(0.00373)	(0.00571)	(0.00373)
Robust	0.0105	0.00636	0.00355	0.00636
	(0.00896)	(0.00417)	(0.00654)	(0.00417)
N	77,811	77,811	77,811	77,811
Effective N left	8,999	10,111	12,651	10,111
Effective N right	10,698	11,828	14,559	11,828
Bandwidth left	7.603	8.083	10.79	8.083
Bandwidth right	7.603	8.083	10.79	8.083

Notes: Standard errors in parentheses, $^{+}$ $p < 0.10$, * $p < 0.05$, standard errors in parentheses.
Source: ONS Longitudinal Study.

Table A5 Intent-to-Treat Estimates: The Causal Effect of Education on Cancer Outcomes for Women (1950–1964 Sample)

	Cancer	Cancer Death	Death	Death from cancer
Conventional	-0.000443	0.000461	-0.00134	0.000461
	(0.00861)	(0.00436)	(0.00439)	(0.00436)
Bias-corrected	-0.00110	-0.000139	-0.00145	-0.000139
	(0.00861)	(0.00436)	(0.00439)	(0.00436)
Robust	-0.00110	-0.000139	-0.00145	-0.000139
	(0.0101)	(0.00511)	(0.00520)	(0.00511)
N	74,565	74,565	74,565	74,565
Effective N left	13,410	10,943	12,200	10,943
Effective N right	15,168	12,555	13,921	12,555
Bandwidth left	11.64	9.074	10.27	9.074
Bandwidth right	11.64	9.074	10.27	9.074

Notes: Standard errors in parentheses, $^{+}$ $p < 0.10$, * $p < 0.05$, standard errors in parentheses.
Source: ONS Longitudinal Study.

Chapter 5

Conclusions

5.1 Summary and discussion

This doctoral research makes at least three significant contributions to the existing scholarship on SES inequality in health and mortality. First, the last part of life can be conceptualised as a life stage and is informative about differences in disability among advantaged and disadvantaged socioeconomic groups. Only wealth appears to be linked with less-disabled trajectories at the end of life in England. Therefore, these results suggest the need for deeper reflection on the meaning of different socioeconomic indicators in old age. Second, this research highlights the importance of changing demographic factors, such as marriage, when studying educational inequality in health over time. In particular, there is an increasingly relevant role of marital status for the rising educational inequality in health over time in United States. Finally, the fourth chapter has provided weak evidence for a causal role of an additional year of schooling for cancer incidence, mortality and survival in England and Wales. Population ageing could possibly lead to the emergence of new health inequalities therefore it is essential to evaluate the causal impact of socioeconomic factors on emerging diseases.

The association between socioeconomic status and health varies greatly during the life-course. Some scholars have suggested the presence of a disappearing association between socioeconomic position and health in old age. Chapter 2, published in

Population Studies, focussed on socioeconomic inequalities in disability trajectories leading to death in England. To contribute to the empirical evidence concerning the existence of a disappearing association between socioeconomic status and health in old age, this study examined the last years of life, conceptualising the final part of life as a life-course phase in which differences across socioeconomic groups are present. This research reveals that the disability pathways leading to death depend on socioeconomic circumstances. Moreover, there was a considerable amount of heterogeneity in the pathways to death across different socioeconomic indicators. In particular, a socioeconomic gradient for the presence and level of disability appears to be evident only for wealth. On the contrary, there are no clear significant differences between income quartiles as well as educational groups.

Second, educational inequality in health and mortality is widening over time in the United States, which may be explained by the changing group composition of educational groups. Therefore, the second chapter investigated the changing self-reported health gap among different socioeconomic groups over time, focusing on the intersection of education with marital status. The findings suggested that when studying health inequalities, it is important to consider the association of socioeconomic status with other family processes. Moreover, the analysis revealed an increasingly relevant role of marital status over time for educational inequalities in self-reported health. Therefore, how individual-level circumstances in terms of marital arrangements influence macro-level patterns of educational inequalities in health was examined.

Third, epidemiological transition and population ageing have increased the relevance of non-communicable diseases for population dynamics. The final part of this research aimed to identify socioeconomic differences in cancer using a quasi-experimental design, investigating the causal role of education on cancer incidence and mortality in England and Wales. Education might affect cancer risk and survival through greater access to resources in terms of prevention, early detection and treatments. Exogenous increases in education produced by the compulsory schooling law changes in 1947 and 1972 in England and Wales were exploited. The two reforms were useful for the identification of the causal effect of education on cancer risk, survival, and mortality. This analysis provided a comprehensive picture of the causal effect of education on cancer outcomes in England and Wales. The results suggest the absence of a causal impact of one additional year of compulsory education for cancer incidence and mortality. To the best of my knowledge, this was the first time the impact of education on overall survival after cancer diagnosis was analysed.

The massive educational expansion in the past century represented a radical societal change, with many more people having access to education, generating a considerable increase in human capital. At the same time, advances in technology and the standard of living increased life expectancy, which, in turn, produced an older population structure. Moreover, the *second demographic transition* and educational expansion have had an extensive impact on living arrangements, partnerships and educational attainment of individuals. In this thesis, socioeconomic inequalities in health are analysed from different perspectives; these all represent important aspects to take into account when studying the socioeconomic gradient on health and mortality.

5.2 Limitations

As discussed in previous section, the three studies in this thesis advanced the literature in several ways, nonetheless, there were some limitations. First, information to construct the disability pathways to death were only available every two years in the ELSA data. The construction of health-related trajectories at the end of life might benefit from having information on shorter intervals between observations. If richer data become available, it would be theoretically possible to observe more diverse trajectories preceding death. Moreover, although the severity of disability was explored, detailed data on the underlying medical conditions leading to disability were not available. Data on hospitalisation and biological samples might be more informative on the underlying health conditions leading to disability. In particular, certain diseases might lead to some trajectories, simultaneously being related to more or less advantaged socioeconomic conditions. Therefore, a greater level of detail in terms of frequency of observations and underlying medical conditions would be an important value added for research in the terminal health-related trajectories.

Second, the need for a more accurate discussion about the meaning of education emerges from all chapters. There was no educational gradient in the health-related pathways to death in chapter 2, which may be due to the way education is related to socioeconomic circumstances in England and Wales. Moreover, there was little variation in educational attainment in the birth cohorts analysed. Chapter 3 focussed on trends over time. It becomes especially important to reflect more in the literature on the changing meaning of education, as higher education has increasingly different implications for labour market prospects and demographic behaviours. It is also important to discuss the changing meaning of education over time when analysing trends across educational groups. Finally, the weak causal evidence for education on cancer in chapter 4 raises additional questions on the efficacy of compulsory schooling reforms on population health. Qualitative differences, such as academic tracking, school funding or class size, in education rather than an additional year of compulsory schooling could be important for health outcomes in later life.

Third, chapter 3 analysed the role of marriage for educational inequalities in self-reported health. Two aspects deserve more attention for future research. First, marriage has changed meaning over the course of the past century. When looking at trends over time, it is essential to explore how the meaning of marriage has changed as well as the processes of assortative mating. However, it was not possible to fully take into account these aspects. Future research should discuss more the way in which the meaning of marriage has changed and what are the possible consequences for these changes in meaning. Second, health is measured with self-reported health, which includes a general evaluation of health status highly predictive of functional limitation and mortality, providing an overall picture of physical well-being despite rising some reliability concerns. However, it would be interesting to explore different health dimensions and examine the role of marital status composition on several aspects of health.

Finally, chapter 4 examined the effect of education on cancer outcomes. It is important to move from associations to causal links when studying the emergence of new health inequalities. Moreover, the weak evidence for a causal effect might be due to the change in education analysed being only a one-year increase in compulsory education. Changes in the mechanisms leading to inequalities in health could be determined by educational qualifications obtained, rather than years spent in compulsory schooling. It is important to advance the understanding of the different aspects of educational attainment that matter for health outcomes. The *fundamental cause* perspective predicts the emergence of health inequalities when more advantaged socioeconomic groups can benefit from information and resources to prevent and treat diseases effectively. However, since the role of knowledge is pivotal for creating benefits for more advantaged socioeconomic groups, in the case of scarcity of information about the mechanisms producing cancer outcomes, it is especially hard to test the *fundamental cause* framework. As information about cancer aetiology and efficacy of treatments become more widespread, then the *fundamental cause* framework could help to further explain socioeconomic differences in cancer outcomes. Moreover, the *life-course* perspective is still underdeveloped, with a limited number of studies focusing on risk factors in adult phases of the life-course, while much more research concentrates on the role of prenatal and early life factors in development of cancer in some sites. Finally, the role of public policy to enhance prevention is essential in limiting socioeconomic differences in cancer outcomes.

5.3 Direction of future research

Population ageing in the next years calls for greater attention to the connection between SES and population health. In particular, from a theoretical perspective, the integration of different outlooks concerning health inequalities can help in advancing the understanding. For example, the *fundamental cause*, the *life-course* and *social determinants of health* approaches represent complementary and integrated ways of analysing health inequalities. The combination of different perspectives could produce new evidence to better understand causes and create mitigating policies for socioeconomic differentials in health.

Three directions emerged from this thesis as important for future research. First, it is important to devote more attention to the longitudinal aspect of population ageing, particularly, to study the socioeconomic gradient in the dynamics of onset and recovery from illness. More and more people will need to deal with chronic illness and disability over the course of their life. Therefore, it is essential to understand how the mechanisms of recovery and management of chronic conditions differ by socioeconomic groups. If differences exist, the analysis of the reasons behind these disparities needs to be understood.

Second, as my work on the role of marital status for increasing health inequality has emphasised, socioeconomic differences in health need to be analysed together with the demographic correlates of SES. This thesis considered marital status, but education is associated with different demographic behaviours, for example, geographical mobility and fertility. The mediating role of these aspects could be very important for understanding the changing socioeconomic gradient in health over time. Moreover, in addition to education it would also be worth exploring the role of demographic correlates of other SES indicators.

Third, the epidemiological transition has increased the relevance of non-communicable and chronic diseases. Therefore, as new illnesses become common, it is important to study whether new health inequalities emerge. Chapter 4 focussed on the causal role of education for cancer outcomes. It would be interesting to expand the analysis of causal effects to other increasingly common diseases due to the population ageing, such as dementia, Alzheimer's, and Parkinson's disease. Studying the socioeconomic gradient in specific ageing-related diseases can be informative about the caring needs across the socioeconomic spectrum. Therefore, future research needs to examine whether population ageing implies the emergence of new unexplored inequalities in the onset and progression of diseases and the new mechanisms leading to socioeconomic differentials.