The London School of Economics and Political Science

Essays in Public and Health Economics

Thomas Minten

Thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy.

London, October 2021

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 48,329 words (including footnotes and appendices but excluding bibliography).

Statement of Conjoint Work

I confirm that Chapter 2, "The Social Determinants of Choice Quality: Evidence from Health Insurance in the Netherlands", was jointly co-authored with Benjamin Handel (UC Berkeley), Jonathan Kolstad (UC Berkeley) and Johannes Spinnewijn (LSE). A previous version of this paper has appeared as NBER Working Paper 27785. This statement is to confirm that I contributed 25 percent of this work.

I confirm that Chapter 3, "The Income Gradient in Mortality during the Covid-19 Crisis: Evidence from Belgium", was jointly co-authored with André Decoster (KU Leuven) and Johannes Spinnewijn (LSE). This paper has been published in the *Covid Economics* issue 50, and is forthcoming in *The Journal of Economic Inequality.* This statement is to confirm that I contributed 33 percent of this work.

Abstract

While citizens in rich countries have indisputably become healthier and richer on average, there is a general sense that this progress has not benefited everyone equally and that health and economic inequality has increased. This thesis contributes to the literature on the measurement and causes of such trends in inequalities by using newly available administrative data in the low countries Belgium and the Netherlands to document and analyze three separate dimensions of contemporary health and economic inequality.

The first chapter analyzes the evolution of migrants' descendants' educational outcomes and incomes in the Netherlands, with a focus on second and third generations migrants from Morocco, Suriname and Turkey. While gaps between natives' and migrants' descendants remain large, gaps are generally smaller for later generations, and are overall decreasing. Moreover, using migrant-of-entry fixed effects, a positive effect of the length of stay of migrant families in the Netherlands on the test scores of migrants' children is established and continues after fifty years. I complement these findings with a discussion of migrants' mobility patterns and the role of intermarriage in economic integration.

The second chapter concerns the choice quality of insurance contracts by individuals in the Netherlands. We study a specific attribute of the health insurance purchase decision all Dutch inhabitants make: the choice of the size of the deductible. We find that individual choice quality is strongly correlated with the education level and professional sector. Moreover, there is a strong correlation between the decision quality of an individual and those of his/her connections, as we find within-firm, location and family impacts on decision making. We document that such inequality in choice quality leads to substantial differences in financial outcomes, and evaluate alternative policies.

The third chapter analyzes the distributional pattern of mortality in Belgium during the first wave of the Covid-19 pandemic in 2020. Using population-wide administrative data, we find that there is a significant negative income gradient in excess mortality, with excess deaths in the bottom income decile more than twice as high as in the top income decile. However, compared to the inequality in mortality in normal times, the income gradient in all-cause mortality is only marginally steeper.

Acknowledgments

Looking back at my LSE years, I will fondly remember the seemingly endless freedom to pursue my curiosity, the extensive reading into different literatures, the energy that co-authoring brings, and the long coding sessions. Yet, I will also remember the difficult and long journey to develop many research ideas into a successful thesis. This work would not have been possible without the support I received from many people.

My deepest appreciation goes out to my supervisor Johannes Spinnewijn, who has been instrumental in the writing of this thesis. His encouragement and support at all the critical stages of the PhD was invaluable, and I am grateful for the academic opportunities he provided me with.

I was lucky to have great co-authors during my PhD. Working with Ben Handel, Jon Kolstad and Johannes Spinnewijn offered me a chance to see the frontier of economics research in action. A silver lining of the Corona pandemic was working with André Decoster and Johannes Spinnewijn on a Covid-19 paper. I enjoyed working with and learning from Maria Ventura, Chloé de Meulenaer, Daniel Privitera and William Parker, as well as with Michael Stepner and co-authors.

Special thanks go to Alan Manning, whose helpful feedback and insights guided my research on migration. I am grateful to Alwyn Young for helpful advice during the writing of my MRes thesis. Additionally, I am thankful for the input from several other members of the Economics Department at the LSE, and from helpful economists at various conferences. I am indebted to Joep Konings and especially Jan Eeckhout, for providing me with research opportunities and stimulating me to do graduate work in economics. I thank UPF for hosting a research visit in Barcelona. I am grateful to ESRC, STICERD, and NBER-IFS for funding my research, and to Statistics Netherlands, Statbel and KSZ for providing microdata.

Derek, Kangchul, Nicolas, Hugo, Mengshan, Joy, Giel, Karlien, Silke, Gemma, Andrea, Sarah, Joe, Adams, Shan, Inés and many others have brightened up my PhD days. Finally, I am, more than to anyone else, grateful to my friends and my family back home, for the many visits to London, and for their support throughout my PhD.

Contents

List of Tables

List of Figures

Chapter 1

Long-Run Economic Assimilation of non-Western Immigrants in the Netherlands

Abstract[1]: This paper analyzes the evolution of migrant descendants' economic outcomes in the Netherlands, with a focus on the second and third generations migrants from Morocco, Suriname and Turkey (MST). While gaps in educational achievement and incomes between those with a native and those with a migrant background generally shrink with each generation, these gaps remain large. Using administrative data on nearly all migration movements, I investigate whether time spent by the extended migrant family in the Netherlands leads to better educational outcomes for children. Using migrant-of-entry fixed effects, I find that even after more than fifty years, all else equal, being born in a family with a one year longer stay in the Netherlands causes an increase of 0.24 percentiles in primary school test scores. I complement these findings with an analysis of intergenerational mobility of migrants, following the methods of Chetty et al. (2020*b*), and find a steady state income gap of 8.5 percentiles for descendants of MST migrants. However, this steady state income gap is likely an overestimate, as accounting for individuals with heterogeneous native-MST backgrounds shrinks the migrant descendants' gap in the third generation by 36.4%.

[1]I thank my supervisor Johannes Spinnewijn as well as my advisor for this project Alan Manning, who both offered me plenty of constructive feedback and support. I also thank Alwyn Young, Ceren Ozgen, Derek Pillay, Kilian Russ, Nathaniel Hendren, Vincenzo Scrutinio and many seminar participants at the LSE for excellent feedback. I am grateful for financial support from STICERD and ESRC.

I Introduction

14.4% of the working population in the Netherlands has at least for 50% a non-Western background in 2017, a percentage that increases to 19.32% when looking at minors in the Netherlands, and to 21.7% when looking at newborns. Yet, we know surprisingly little about the evolution of economic outcomes of these non-Western descendants in the Netherlands. Are economic outcomes trending towards the averages of those with a native background, or can we predict a long-run native-migrant gap and persistent differences in educational attainment and wages? Naturally, the educational and income levels of large parts of the population are important for the future of the Dutch economy and the sustainability of its welfare state. Additionally, there is a strong equity and social justice dimension to these socioeconomic differences between children with a native background and children of non-Western migrants, and they are an often cited driver for tensions between ethnic groups. As such, the importance of improved economic integration of ethnic minorities into host societies has been underwritten by the "Zaragoza Declaration" of the European Commission in 2010.

An optimistic view, following the classical assimilation perspective, is that economic assimilation of migrants and their descendants in Europe is a matter of time and that the migrant-native gap will become smaller over time only to disappear with an extended stay in the host country. In the literature, there is abundant evidence of an *initial catch-up* of first generation migrants' incomes in a variety of settings and host countries (see for example, Akee and Jones (2019), Lubotsky (2007) and Ansala, Aslund and Sarvimaki (2018)).[2] Evidence is, however, more mixed when it comes to the evolution of second and later generations' incomes relative those of natives' children. In some countries in Europe, second generation non-Western migrants close an additional part of the migrant gap (see, eg. Bolotnyy and Bratu (2018) and Algan et al. (2010)), while in other countries second generation migrants do similar or worse than first-generation migrants, depending on the country of origin (see, eg. Algan et al. (2010) and Piton and Rycx (2021)).[3] Moreover, we know that long-standing income

[2]While the first two papers focus on the integration of migrants in the US, Ansala, Aslund and Sarvimaki (2018) show that the integration of migrants is remarkably similar in Sweden and Finland, two countries that have very different migration histories. Several mechanisms cause a relatively large rate of migrant integration. Low initial average employment rates and wages, the adoption of the host country's language, and the establishment of professional networks all make for a relatively fast improvement of economic status.

[3]Both Algan et al. (2010) and Piton and Rycx (2021) cite large differences in the measured performance of immigrants' children, depending on the country of origin. Such findings underwrite a so-called segmented assimilation hypothesis, where certain groups assimilate while others do not (eg., Portes and Zhou (1993)).

differences between ethnic groups have persisted for generations in several societies.[4] The large variety of integration patterns in Europe is seemingly dependent on pair of sending and host countries, and warrants an in-depth analysis of the economic outcomes of non-Western migrant descendants in the Netherlands.[5]

There are several key advantages to studying this issue specifically in the Netherlands. First, while typically not seen as a "migrant's country", the Netherlands has received a steady stream of non-Western migrants since the 1960s, relatively early compared to other high-income Western European countries. Non-Western immigration to the Netherlands has largely consisted of migration from former colonies (eg. Indonesia and Suriname), economic migration (eg. Turkey and Morocco), and more recently predominantly asylum migration (eg. Afghanistan and Iran). This steady stream of migrants is useful for the identification of time trends, as will be discussed later. Second, the availability of high-quality administrative data on all individuals' migration movements going back to the 1940's, together with parent-child links and information on incomes and educational test scores, allows the tracking of integration patterns for all migrant families across time. Third, studying the long-run migrant gap in the Netherlands is interesting since the overall success or failure of economic assimilation of migrant communities is highly debated. The long-run integration of descendants of the three largest migrant communities from Morocco, Suriname and Turkey (MST) will be the focus of this paper.

I start the analysis by documenting contemporary migrant gaps in education for 12 year old primary school children, as well as migrant gaps in personal incomes and wages for 28 to 32 year olds. There are large migrant gaps in primary school test scores and incomes for individuals with non-Western backgrounds. Descendants with backgrounds of Morocco, Suriname and Turkey (MST) face especially large test score gaps of 14.3, 11.2 and 17.0 percentiles, respectively, and income gaps of 13.0, 11.4 and 12.1 percentiles, respectively. Children of Western immigrants have smaller test school gaps of only 0.7 percentiles but sizeable income gaps of 7.5. An interesting finding of the paper is that individuals with

[4] A well-known example is the black-white gap in the United States. Chetty et al. (2020*b*) argue that the current income level-differences between blacks and whites are close to a steady-state and thus likely to persist if intergenerational mobility patterns for whites and blacks stay similar. However, it has to be noted that the black-white gap in the United States is fundamentally different from the migrant-native gap in Europe. Most non-Western migrant families in Europe are still only recent arrivals. Moreover, the institutional environment for migrants in Europe is different from the institutional environment for blacks in the United States. Note that a persistent gap in income in the USA is absent for the descendants of Western immigrants that moved into the USA in the last centuries, as Abramitzky et al. (2021) show that long-run outcomes of descendants of Western immigrants converge quickly to average native outcomes.

[5] While there are various dimensions to the integration of non-Western migrant families in Europe, this paper will only deal with measurable economic integration, more specifically educational attainment and incomes.

a heterogeneous (mixed) background with one native parent and one parent with an MST heritage close on average more than 75% of the migrant gap in tests scores, and 73% of the gap in incomes.

Acknowledging that snapshot migrant gaps change over time and show trends, the main part of the analysis tries to study empirical integration patterns. Reassuringly, gaps in test scores have been decreasing in the last years with 0.56 percentiles each year in 2006-2018 for individuals with MST backgrounds. However, it is less straightforward to disentangle these time trends into integration patterns. To do this, I use two distinct methodologies to make use of historical data to learn about long-run assimilation of immigrants between 1999 and 2018. The first methodology looks at the historical correlation between the time a migrant family resides in the host country and the educational outcomes of the children within that family. By controlling for fixed effects for the first relatives who entered the Netherlands, I use only within-extended family variation to measure the correlation between time in the Netherlands and educational outcomes, so that most (changes in) migrant selection and nationwide trends in migrant outcomes are arguably accounted for. On average, time spent by the extended family in the Netherlands cause improved educational outcomes. The results show that having spent one year more in the Netherlands cause a 0.24 percentile higher primary school test score in families with a non-Western background, and that positive effects are still accumulating decades after the initial entry of these families. However, there is large heterogeneity by groups: while children from Moroccan immigrants seem to get better test scores at a rate of 0.40 per year a family stays in the Netherlands, children with a Surinamese background improve their test scores only by 0.27 per year. Overall, my findings strengthen the intuition that integration is a process that continues for decades, and can differ by background.

The second methodology looks at the economic integration of migrants in an intergenerational mobility framework, similar to the analysis of the black-white income gap in Chetty et al. (2020*b*). Estimating economic mobility in a linear rank-rank regression, I find large disparities in the outcomes between migrants' and natives' children, conditional on parents' incomes. For children of MST backgrounds, there are absolute mobility gaps in primary school test scores of 9.56 and 13.10 at the 1st and 10th parental income decile, respectively, and absolute income mobility gaps of 4.21 and 8.18 percentile at the 1st and 10th parental income decile, respectively. These estimates are large compared with many other estimates of migrant or racial mobility gaps, and they are larger than the gaps in individual income mobility of blacks in the US, as documented by Chetty et al. (2020*b*), and larger than the income mobility gaps for migrants from Jamaica, Trinidad and Tobago and Haiti in the United States (Abramitzky et al. (2021)).[6] Overall, this analysis paints a bleak picture: individuals with MST background have low absolute educational and income mobility, with limited improvement in the mobility estimates across generations.

The last part of the paper uses these mobility estimates to calibrate a model of the evolution of average income percentiles for children with a native background compared to those with a background from MST countries. Additionally, I extend the model from Chetty et al. (2020*b*) by including interethnic couples and children with heterogeneous (mixed) backgrounds, which provides several insights. While there has been an enormous improvement of relative income percentile from the first to the second generation for individuals with a MST background, a migrant-native family income gap of 8.5 is predicted in the long run, conditional on the mobility estimates remaining constant for future generations. The steady state gap is mitigated, however, when taking into account the increasing fraction of children with mixed native-migrant backgrounds who have significantly higher mobility. Extending this model with parameters for intermarriage, the shrinking of the migrant gap is increased by 14.4% in the first generation, and 36.4% in the second generation.

Related Literature. There exists a large literature on the economic integration of migrants, with abundant empirical research focussing on the US and several European countries.[7] It has become clear that panel data is necessary to measure the integration of migrant groups over long periods of time, as it does not suffer from biases associated with repeated cross-sectional data.[8] Several recent studies use high quality administrative panel data on two generations to document migrant integration. In Canada, Aydemir, Chen and Corak (2009) use the Canadian census in 1980 and 2000 and argue that there is no difference in the degree of intergenerational persistence of earnings between natives' and immigrants' descendants. Abramitzky, Boustan and Eriksson (2014), Abramitzky, Boustan and

[6] Western migrants in the US (Abramitzky et al. (2021)), or all types of migrants to Sweden (Bolotnyy and Bratu (2018)), have higher or equal or higher upward mobility compared to natives.

[7] The literature on the economic assimilation of migrants in the US features early empirical papers by Chiswick (1978) and Borjas (1985), while an overview of migrant gaps is given by Dustmann and Frattini (2011). Famously, Borjas (1992) discusses the intergenerational mobility of individuals with different ethnicities and argues that children's outcomes depend on the average human capital of people with the same ethnic background ("ethnic capital").

[8] Abramitzky, Boustan and Eriksson (2014) observe that many papers have used repeated cross-sectional data to measure the integration of migrants across generations. They find that using repeated cross-sectional data is prone to many biases, and argue that in the debate on the assimilation of immigrants in the US, assimilation has been overestimated due to the biases from declining skill level of entering migrants, negatively selected return migration and other composition changes. They argue that panel data should therefore be the gold standard for long-run integration research. Akee and Jones (2019) confirm this view, as they show that by 10 years after arrival in the US, almost 40% of migrants have returned to their home countries.

Eriksson (2020) and Abramitzky et al. (2021) find that descendants of European migrants have fully economically integrated in the US in about two generations.[9] In Europe, studies using two generations are done by Bauer and Riphahn (2007) for Switzerland, Hammarstedt and Palme (2012), Bolotnyy and Bratu (2018) for Sweden, Ansala, Hämäläinen and Sarvimäki (2019) for Finland, Hermansen (2016) for Norway, and Piton and Rycx (2021) for Belgium. These studies suggest that economic integration is highly variable to the specific pair of sending and host countries. For example, while Bolotnyy and Bratu (2018), Ansala, Hämäläinen and Sarvimäki (2019), Hermansen (2016) find equal or higher economic mobility in Sweden and Norway for migrants' versus natives' children, Piton and Rycx (2021) talk about a broken social mobility escalator for migrants' children in Belgium, with significantly lower mobility for non-Western migrants.

The contribution of this paper to the migrant assimilation literature is threefold.

First, this paper documents patterns of migrants' outcomes and economic mobility in the Netherlands, where these patterns are not yet well known, and the successful integration of non-Western migrants is highly debated. As such, this paper contributes to a handful of empirical papers on migrants' economic outcomes in the Netherlands such as Bakker, Dagevos and Engbersen (2017), Zorlu and Hartog (2018), Thijssen et al. (2021), van de Beek et al. (2021) and Falcke, Meng and Nollen (2020).[10]

The second contribution of this paper relates to the insights on the long-run economic integration it provides. The paper shows that integration is a long process, with economic outcomes improving up to 50 years after initial migration. This paper is uniquely positioned in the literature to make statements on long-run integration, with panel data stretching over three generations of migrants. Such long-run panel data of migrants is missing in most papers on migrant integration.[11]

[9] Abramitzky, Boustan and Eriksson (2014) look at the European migrants in the US during age of mass migration, and subvert the long-held view that European migrants assimilated quickly, as they find that differences persist into the second generation. Abramitzky, Boustan and Eriksson (2020) use names of immigrants to document that in the US, the rate of closing the naming gap is around twenty years, both historically and currently. Abramitzky et al. (2021) uses son-father pairings from several decades across the past century to document that second generation immigrants have higher mobility estimates than children of the US-born.

[10] Bakker, Dagevos and Engbersen (2017) document a decreasing refugee gap for the 1st generation. Zorlu and Hartog (2018) argue that knowledge of the local language is a strong driver for the integration of migrants. van de Beek et al. (2021) highlight that the lack of integration of migrants in the Netherlands has severe associated costs for the taxpayer. Thijssen et al. (2021) show that discrimination is an issue in the Netherlands, with lower call-back rates for Turkish job applicants. Falcke, Meng and Nollen (2020) find that non-Western migrants in the Netherlands are less likely to be employed in a that matches their field and level of study.

[11] Indeed, despite the increased availability of administrative data, little empirical evidence has been put forward that investigates the evolution migrant gaps beyond the second generation, or within non-Western families that have been in Europe for over 30 years. This might be in part due to the fact that large-scale migration of non-Western migrants to Europe

Third, this paper contributes to the literature on children with parents from two distinct ethnic backgrounds. Mostly due to data limitations, empirical findings on such children are relatively scarce in the literature on integration, with exceptions by Fryer et al. (2012) and Tegunimataka (2020) providing a mixed picture about the economic outcomes of these children.[12] In the Netherlands, I find that children with one native and one migrant parent close up to 70% of the migrant gap. This finding has important implications in a model of the migrant descendants' gap that includes interethnic coupling, which is often overlooked in empirical studies on migrant integration.

The rest of the paper proceeds as follows. Section II briefly describes non-Western migration to the Netherlands and gives a description of the data. Section III provides descriptive statistics on the contemporary migrant gap in the Netherlands, and discusses its recent evolution. Section IV and V then present the main results, with Section IV analyzing the link between length of stay in the Netherlands and migrant descendants' outcomes, and Section V discussing migrants' economic mobility in an intergenerational approach similar to Chetty et al. (2014*b*) and Chetty et al. (2020*b*). Finally, Section VI concludes.

II Setting and Data

II.A Setting: Non-Western Migrants in the Netherlands

While in the 19th and early 20th century the Netherlands had largely been a country of emigrants, since 1961 this trend has reversed and immigration flows exceeded emigration flows except in 1967 (Zorlu and Hartog (2001)). Figure 1.1 indeed shows that after 1960 several tens of thousands non-Western migrants moved to the Netherlands.[13] Over time, these immigration flows have caused a sizeable migrant population of non-Western origin in the Netherlands to accumulate. Currently, the shares of individuals from non-Western backgrounds living in the Netherlands are comparable with other Western European countries. 14.4% of the working population in the Netherlands has at least for 50% a non-Western background in 2017, a percentage that increases to 19.32% of minors in the

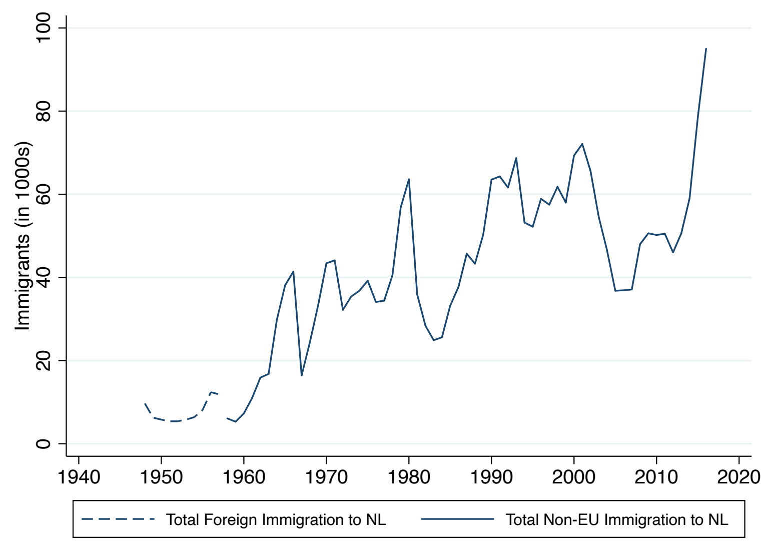
is a relatively recent phenomenon. For example, Scandinavian countries with traditionally good access to administrative data, have only been accepting non-Western migrants in large numbers since since the late 1980s. Note that in some cases where longer-run integration is measured, data on last names is used, which provides imperfect measurement.

[12]Other ways intermarriages affect the migrant gap are discussed in van Ours and Veenman (2010), who argue that intermarriage also positively affects the educational outcomes of children, and Bisin, Topa and Verdier (2004) and Bisin and Verdier (2000).

[13]For a more detailed description of the immigration flows in this particular period, see Zorlu and Hartog (2001).

Netherlands, and to 21.7% of newborns.

FIGURE 1.1: IMMIGRATION TO THE NETHERLANDS 1948-2016



Notes: This figure plots the number of total foreign immigrants to the Netherlands for 1948-1957, and the total number of non-EU immigrants to the Netherlands for 1958-2016. Source: CBS Statline.

While non-Western migration to the Netherlands has been very diverse, in this paper I focus mostly on the three largest migrant communities, and will distinguish them from people of other non-Western or Western backgrounds. The three largest ethnic minority communities include people with Moroccan, Surinamese and Turkish (abbreviated MST) backgrounds and constitute 53.3% of all children with non-Western backgrounds in the Netherlands in 2017. These communities have been formed by colonial ties as well as economic migration. Large migration waves from Surinam and the Dutch Antilles happened after the decolonisation of Surinam in 1975. In the 1960s, the Netherlands recruited and attracted spontaneous "guest workers" from Italy, Spain, Portugal, Turkey, Greece, Morocco, Yugoslavia and Tunisia. These migration waves were often followed by chain migration, which led to a considerable yearly stream of migrants from Morocco, Suriname and Turkey.[14]

Important to note is that non-Western migration to the Netherlands was relatively rare before 1960 (Lucassen and Penninx (1997), Zorlu and Hartog (2001)). Individuals with recent non-Western heritage can thus, for the most part, linked back to a "first" immigrant in the family tree who has come to the Netherlands after 1960. This fact is made use of in the data, which include all migration movements

[14]Due to different birth rates or sustained migration flows, the size of ethnic communities in the Netherlands is very different by ethnic background.

for individuals who at one point after 1995 were registered as an inhabitant in the Netherlands.

II.B Data and Sample

This paper uses individual-level data from Statistics Netherlands (CBS), linking several administrative registers that provide information on all citizens' income, wealth, education, employment as well as various demographic variables. Important for this paper is the detailed information on ethnic background through child-parent links, together with the universe of migration movements of people who were in the Netherlands at one point after 1995. The datasets originate from different governmental organizations, and are maintained at the Central Bureau for Statistics.[15]

Below, the most important demographic variables are described. Three variables were selected to evaluate the socio-economic status of each individual as they are available for (almost) the entire population within a certain age bracket: primary school test scores in childhood, and primary income and wages in adulthood.

Main Ethnic Background. In order to make the most accurate representation of the ethnic background of all individuals in the Netherlands, I created a procedure that defines someone's ethnic background based on the country of birth of his/her ancestors. In short, for each individual a family tree is constructed, up until the great-grandparents, and the available information of the earliest ancestors within one's family tree is used to gauge someone's background.[16] Someone's ethnic background is defined to be that background where the mode of the earliest ancestors were born. In several cases this does not give a unique background. If there are 50/50% splits, the non-Dutch country is chosen. If that still does not gives one unique solution, the paternal great-grandfather's background is chosen.[17] Distinction is made between Dutch, Moroccan, Surinamese, Turkish, Other Non-Western and Other

[15] All data was linked using anonymized personal identifiers in a secured remote-access environment at the CBS.

[16] While there is nearly universal coverage of data on the country of birth of parents, for many individuals with a non-native background, information on the country of birth of their grandparents and great-grandparents is missing. If only a part of the family tree is observed, the earliest visible ancestors in the tree were used to derive the background. The procedure is as follows. If the country of birth of any of the great-grandparents is not observed, the country of birth of the grandparents is then used to define someone's background. Similarly, if the grandparents are not observed, the countries of birth of the respective parents is used, which are always observed as everyone in the Dutch registers is included with country of birth of both parents. This procedure is followed for both sides of someone's family.

[17] Note that the earliest ancestors are weighted relative to the position in the family tree. Moreover, if at least 50% of an individual's background is non-Dutch, someone's background is defined to be the relevant non-Dutch country. For other individuals whose background is equally split between two or more groups, the background with the highest prevalence among the great-grandparents is picked. If this does not give one outcome, the ethnicity that is most prevalent on the father's side is chosen. Note that this procedure links individuals to only one ethnic background. Hence, individuals only appear once in the data.

Western backgrounds. As the focus of this paper is on the three largest migrant communities in the Netherlands, in several analyses individuals with Moroccan, Surinamese and Turkish (abbreviated MST) backgrounds are pooled together.

The prevalence of interethnic relationships has given rise to an increasingly large share of the population that is of mixed heritage. This is shown by Appendix Figure A1.6, which shows the fraction of mixed heritage for each migrant community by birth year. For all children born in the Netherlands in 2017, 40% has at least one grandparent who is born outside North-Western Europe. In the analyses, special attention is given to individuals with a 50%-50% background with one parent from the Netherlands and the other from Morocco, Suriname or Turkey.[18]

First Migrant in Family. To study the long-run integration of descendants of migrants throughout time, it is important to be able to link every individual to the individual in his/her extended family that was the first to enter the Netherlands. To obtain the first migration movement in the extended family, the migration movements of someone's parents, grandparents and great-grandparents are taken into consideration.[19] If there are two family members who migrate at the same (earliest) time, the oldest male migrant is selected to be the first migrant in the country. Because the migration history is only available for individuals who were alive in the Netherlands at any time after 1995, there is no complete picture of everyone's family migration history. However, links to at least one migration movement is constructed for more than 93% of the 12-year-old children of Moroccan, Surinamese or Turkish background in the sample.

Primary School Test Percentile. To rank children in terms of educational outcomes, the within-cohort-gender percentile score on the Cito test in primary school is calculated, ie. within each cohort-gender group percentiles are assigned based on the overall test score. The Cito test is a standardized test used by the large majority of primary schools in the Netherlands in 2006-2018, and is administered in the last year of primary school, when pupils are typically aged 12. The Cito test includes language, mathematical and a study skills sections. The test is of significant importance for young children's future, as it is used to give recommendation for tracks in high school.[20]

[18]In this paper individuals with a mixed, heterogeneous, ethnic background are still considered as "descendants of migrants". However, they are also "descendants of natives". This has an impact on the migrant-native gap, as will be discussed in Section V.

[19]Since non-Western migration before 1940 is rare, looking at only three generations will probably include the large majority of migration moves to the Netherlands from non-Western countries.

[20]The high school track recommendation is based on the overall result on the Cito test as well as a discretionary opinion of the teachers. Interestingly, and relevant to this paper, is that in the years of study, there was positive discrimination

Income and Wage. To rank individuals in terms on income and wage over time, within age-gender percentiles for each individual within his/her cohort are constructed. The income definition in this paper is called "primary" income by Statistics Netherlands and includes labor income, employer social security contributions, self-employment income and capital income. It can therefore be seen as pre-tax pre-transfer income as it does not include government transfers (e.g., UI, DI or pensions). Lastly, the hourly wage is based on a calculation from the universal contract data *spolisbus* held at Statistics Netherlands, and wage percentiles are similarly created based on the ranking within cohort-gender groups.

II.C Summary Statistics

In this paper, I measure the migrant gap and integration for two slices of the Dutch population. The first subgroup includes all children who have taken the standardized Cito test at the end of primary school and the second subgroup includes all individuals who are aged between 28 and 32 in 2018, and whose income is observed. These two distinct samples are called the primary school sample and the income sample, respectively.

Primary School Sample. This sample includes all children who take a Cito test in the last year of primary school during the period between 2006 and 2018.[21] Panel A of Table 1.1 shows summary statistics this sample, split out by ethnic background. There are three things to note from this table that relate to the later analysis in this paper. First, the table makes it clear that the vast majority of children with with Moroccan, Surinamese or Turkish background are second or third generation immigrants, as above 90% of them are born in the Netherlands. Moreover, and important for Section IV, for more than 93% of these children, at least one migration move in the family is observed. Second, the table shows the stark difference in socio-economic environments in which native and non-Western children are brought up. While the average parent of native children have more than 13 years of schooling, parents of Moroccan and Turkish children on average have less than 9 years of schooling. These parental schooling disparities also translate into large household income level differences. The average parental income of native children (62,584 euros) is approximately double that of Moroccan (29,512 euros) and Turkish (32,369 euros) parents. Third, the average earliest registered migration

against most children with an immigrant background: they were more frequently advised than children with a Dutch background to follow a more difficult track than the Cito test score would guide them to.

[21] For children who have taken the Cito test twice within this period, the observation with the highest score is selected.

move in the family for the MST backgrounds is in the 1970's. This means that the average child in the sample with an MST background lives in a family that has been in the Netherlands for over 40 years on average at the time these children take the Cito test, as Cito test results were obtained in between 2006 and 2018.

Income Sample. This sample includes all individuals that are between 28 and 32 years old in 2018, and whose personal primary income is observed. Panel B of Table 1.1 shows summary statistics of the individuals within this sample, split out by ethnic background. Notable is that, from this group, around 75% of individuals with backgrounds from Morocco, Suriname and Turkey (MST) are born in the Netherlands, so that still a majority of these individuals are second- or third-generation immigrants.

Table 1.1: Summary Statistics

A. Primary School Test Sample

	Ethnic Background					
	Dutch	Moroccan	Surinamese	Turkish	Other Western	Other Non-Western
Demographics						
Age when Taking Primary School Test	12.3	12.5	12.4	12.6	12.4	12.5
Male	50%	49%	50%	50%	50%	49%
Migration Background						
Born in Netherlands	99%	96%	94%	96%	83%	79%
0 Foreign Born Parents	97%	1%	9%	3%	23%	1%
1 Foreign Born Parents	3%	13%	33%	21%	50%	29%
2 Foreign Born Parents	0%	85%	59%	76%	26%	69%
Registered Migration Move in Family	23%	99%	84%	98%	88%	84%
Earliest Registered Migration Move in Family	1971	1975	1978	1976	1981	1990
Education and Income						
Pct Score on Primary School Test	52.79	35.53	39.54	32.98	51.16	43.61
Mother Years of Education	13.0	6.8	11.5	7.5	12.2	9.7
Father Years of Education	13.3	7.9	11.6	8.9	12.5	10.6
Parent Family Income 2003	62,584	29,512	47,582	32,369	57,287	35,069
Pct Parent Family Income 2003	55.2	28.0	43.1	28.2	50.1	31.7
Number of Observations	1,408,375	66,425	45,291	64,568	104,272	97,449

B. Income Sample

	Ethnic Background					
	Dutch	Moroccan	Surinamese	Turkish	Other Western	Other Non-Western
Demographics						
Age	30.0	30.0	30.0	29.9	30.0	30.0
Male	51%	49%	49%	53%	48%	50%
Migration Background						
Born in Netherlands	99%	72%	77%	70%	34%	21%
0 Foreign Born Parents	97%	0%	1%	0%	10%	0%
1 Foreign Born Parents	3%	5%	19%	4%	23%	9%
2 Foreign Born Parents	0%	95%	80%	96%	68%	90%
Registered Migration Move in Family	18%	87%	73%	82%	36%	34%
Earliest Registered Migration Move in Family	1967.3	1973.5	1976.6	1975.1	1973.3	1986.7
Education and Income						
Years of Education	13.2	11.7	12.4	11.7	12.4	11.5
Personal Income in 2018	31,898	19,184	23,046	22,604	27,413	20,090
Pct Personal Income in 2018	52.4	39.4	41.0	40.3	44.9	39.0
Hours Worked in 2018	1,466	908	1,162	1,083	1,331	970
Hourly Wage in 2018	16.9	15.1	14.8	14.9	15.5	14.8
Salary Income in 2018	28,731	16,672	21,389	18,756	25,093	18,366
Self-Employment Income in 2018	2,973	2,470	1,603	3,767	2,207	1,651
Financial Income in 2018	194	43	54	82	113	73
Parental Income						
Parent Family Income 2003	66,861	22,646	47,369	30,262	59,043	33,580
Pct Parent Family Income 2003	53.6	20.1	40.2	25.0	47.6	28.4
Number of Observations	753,951	30,488	28,581	35,279	128,373	97,241

Notes: The table in Panel A shows summary statistics for all primary school children who took the Cito test in primary school between 2006-2018, split out by ethnic background. The table in Panel B shows summary statistics for all Dutch citizens who are between 28 and 32 years of age in 2018. For both tables, migration background and parental education are measured in 2018.

III Descriptives on the Migrant Gap

The summary statistics of the different ethnicities' educational and income outcomes in Table 1.1 already give a glance of a considerable migrant gap. Before moving on to the main analyses of this paper, this Section paints a more detailed picture of the contemporary migrant gap in the Netherlands, with special attention given to different generations of immigrants and individuals with mixed ethnic backgrounds. Afterwards, this Section presents the evolution of the migrant gap in the last decades.

III.A The Current Migrant Gap

Primary School Test Score Gap. Differences in socio-economic achievement between children with different ethnic backgrounds are already present early in childhood. This is shown in Table 1.2, which uses data available from 2013-2018 to calculate the differences in primary school test school averages by ethnic background. Column A1 shows that children with backgrounds from Morocco, Suriname and Turkey (MST) face especially large test score gaps of 14.3, 11.2 and 17.0 percentiles, respectively. While children with other non-Western backgrounds do on average 7.3 percentiles worse than those with a Dutch background, children of Western immigrants have smaller test school gaps of only 0.7 percentiles.

Diving deeper into Table 1.2 provides insight in the heterogeneity of the gaps across subgroups within different ethnic backgrounds.

First, all results are split by gender in Column B1 and C1. It is not the case that gaps are universally larger for boys or girls, as the gap for boys with Surinamese background is larger than that for girls, but the reverse holds for boys from Moroccan and Turkish background. Girls with a Turkish background face on average the largest test score gap among all gender-background groups, with an average difference with those with a native background of 17.4 percentiles.

Second, columns A2, A3 and A4 (and B2, B3, B4 and C2, C3, C4) include only children with a 100% homogeneous ethnic background and show the gaps by migrant generation.[22] In this table, 1st generation migrants are children born in the foreign country, who migrated to the Netherlands at some point in their life. 2nd generation migrants are born in the Netherlands to two 1st generation migrants, and 3rd generation migrants are born in the Netherlands to two 2nd generation migrants.[23] Important

[22] In this paper, an individual has a homogeneous ethnic background if all great-grandparents come from the same country.

[23] Note that by only selecting these restrictive migrant generations, not all individuals with a certain homogeneous background are used in Columns 2-4.

Table 1.2: Primary School Test Score Gap in 2013-2018

	Pooled Genders					Boys					Girls				
Dependent Variable:	(A1)	(A2)	(A3)	(A4)	(A5)	(A1)	(A2)	(A3)	(A4)	(A5)	(B1)	(B2)	(B3)	(B4)	(B5)
Primary School Test Score	All	1st Gen	2nd Gen	3rd Gen	Mixed 50% Dutch	All	1st Gen	2nd Gen	3rd Gen	Mixed 50% Dutch	All	1st Gen	2nd Gen	3rd Gen	Mixed 50% Dutch
Ethnic Background (ommitted: 100% Dutch)															
Moroccan	-14.3***	-17.6***	-14.9***	-10.3***	-8.2***	-13.6***	-17.0***	-14.1***	-9.5***	-10.6***	-15.0***	-18.1***	-15.7***	-11.1***	-5.6***
	(0.2)	(1.1)	(0.2)	(1.6)	(1.1)	(0.2)	(1.5)	(0.3)	(2.3)	(1.5)	(0.2)	(1.7)	(0.3)	(2.2)	(1.6)
Surinamese	-11.2***	-19.8***	-13.8***	-14.9***	-4.2***	-11.6***	-23.0***	-14.4***	-16.1***	-4.8***	-10.7***	-17.5***	-13.1***	-13.8***	-3.6***
	(0.2)	(1.2)	(0.3)	(1.2)	(0.6)	(0.3)	(1.7)	(0.5)	(1.8)	(0.9)	(0.3)	(1.6)	(0.5)	(1.7)	(0.9)
Turkish	-17.0***	-18.1***	-18.4***	-13.7***	-4.7***	-16.6***	-17.2***	-17.9***	-12.3***	-0.3	-17.4***	-19.2***	-18.9***	-15.4***	-8.2***
	(0.2)	(1.2)	(0.2)	(0.9)	(1.4)	(0.3)	(1.6)	(0.3)	(1.3)	(2.1)	(0.3)	(1.7)	(0.3)	(1.4)	(1.8)
Other Western	-0.7***	-8.2***	-5.3***	-13.3***	4.2***	-0.7***	-8.3***	-5.9***	-15.6***	4.5***	-0.6***	-8.1***	-4.7***	-11.0***	3.9***
	(0.2)	(0.5)	(0.5)	(1.6)	(0.3)	(0.2)	(0.7)	(0.6)	(2.2)	(0.4)	(0.2)	(0.6)	(0.6)	(2.3)	(0.4)
Other Non-Western	-7.3***	-15.1***	-8.1***	-12.1***	0.4	-7.0***	-15.8***	-7.7***	-16.6**	0.9	-7.6***	-14.5***	-8.5***	-9.5*	-0.2
	(0.2)	(0.4)	(0.2)	(4.4)	(0.4)	(0.2)	(0.6)	(0.3)	(7.3)	(0.6)	(0.2)	(0.6)	(0.3)	(5.4)	(0.5)
Constant	52.5***	52.3***	52.2***	52.2***	52.3***	52.1***	51.9***	51.8***	51.8***	51.9***	52.8***	52.7***	52.6***	52.6***	52.7***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	659,447	398,980	442,081	383,783	414,774	329,631	200,116	221,532	192,637	207,551	329,816	198,864	220,549	191,146	207,223
R-squared	0.026	0.005	0.027	0.001	0.001	0.024	0.005	0.025	0.001	0.001	0.028	0.005	0.030	0.001	0.001

Notes: This table shows the gap in primary school test percentiles by gender and ethnic background. Individuals with a Dutch ethnic background are omitted, so that the coefficients present the difference in average test score of the respective ethnic group compared with the individuals with a native Dutch background. Columns A2, B2 and A3, B3 and A4, B4 show gaps for children with a homogenous ethnic background from the first, second and third generation, respectively. Column A5, B5 show gaps for children with a split ethnic background, with 50% of their great-grandparents being Dutch, and 50% of their great-grandparents having the respective migrant background. Robust standard errors in parentheses. $***$ p<0.01, $**$ p<0.05, $*$ p<0.1

to recognize is that within broad categories such as Western and non-Western backgrounds, there are many compositional differences that will cause differences in outcomes across generations. Columns A2-A4 show that in most cases, contemporary migrant gaps are smaller for later generations. A clear example of this pattern are Moroccan boys, for which the contemporary gap is 17 for 1st generation migrants, 14.1 for 2nd generation migrants and 9.5 for 3rd generation migrants.

Third, column A5 shows the gaps between children with a Dutch background and children with a 50%-50% mixed heritage from a specific migrant background and the Netherlands. The vast majority of these children with heterogeneous backgrounds have significantly higher scores than their counterparts with a homogeneous non-Western ethnic background. Indeed, individuals with a mixed non-Western background with one native parent and one parent with an MST background close on average more than 75% of the gap in tests scores. Lastly, children from parents with both a native and another Western background have on average 4.2 percentile *higher* scores.

Personal Income Gap. Table 1.3 is analogous to Table 1.2, but has as dependent variable personal income, and makes use of a sample with all 28 to 32 year olds in 2018. Column A1 shows that individuals with backgrounds of Morocco, Suriname and Turkey have average personal income gaps of 13.0, 11.4 and 12.1, which are largely similar to the primary school test score gaps. While for test school gaps, there was no clear gender difference in migrant gaps, this is not true for income gaps that are on average larger for boys than for girls. Men with Moroccan and Surinamese backgrounds have gaps that are 6.7 and 7.1 percentiles larger than their female counterparts. This pattern is true for both the first and second generation.

To put these figures in perspective with the black-white gap in the US, comparable estimates of black-white racial gaps in individual income by Chetty et al. (2020*b*) are 17.6 percentiles for males and 4.8 percentiles for females.[24] Compared to the black-white differential in the US, the gaps for migrants in the Netherlands therefore are slightly smaller for males, but larger for females.

Personal income percentile gaps at age 28-32 are larger for males and smaller for females when compared to percentile gaps in primary school testing, with the notable exception of those with a Turkish background. Two other findings from the primary school test scores from Table 1.2 translate to personal incomes at age 28-32 as well. Migrant gaps are generally smaller for the second than for the

[24] Chetty et al. (2020*b*) measure the black-white gap in individual income for individuals who are between 31-37 years old in 2014-2015.

Table 1.3: Personal Income Gap of 28-32 Year Olds in 2018

Dependent Variable:	Pooled Genders				Men				Women			
Dependent Variable:	(A1)	(A2)	(A3)	(A4)	(A1)	(A2)	(A3)	(A4)	(B1)	(B2)	(B3)	(B4)
Personal Income Percentile	All	1st Gen	2nd Gen	Mixed 50% Dutch	All	1st Gen	2nd Gen	Mixed 50% Dutch	All	1st Gen	2nd Gen	Mixed 50% Dutch
Ethnic Background (ommitted: 100% Dutch)												
Moroccan	-13.0***	-19.4***	-11.4***	-11.7***	-16.2***	-20.8***	-15.4***	-18.2***	-9.5***	-17.8***	-7.0***	-4.1
	(0.2)	(0.4)	(0.2)	(2.4)	(0.3)	(0.5)	(0.3)	(3.0)	(0.3)	(0.6)	(0.3)	(3.6)
Surinamese	-11.4***	-14.2***	-12.0***	-8.0***	-14.9***	-17.1***	-16.7***	-10.5***	-7.8***	-11.5***	-7.2***	-5.4***
	(0.2)	(0.4)	(0.3)	(0.7)	(0.3)	(0.5)	(0.4)	(0.9)	(0.3)	(0.5)	(0.4)	(0.9)
Turkish	-12.1***	-15.5***	-11.3***	-7.4***	-12.3***	-16.1***	-11.5***	-7.8**	-12.0***	-15.0***	-11.3***	-7.1*
	(0.2)	(0.3)	(0.2)	(2.4)	(0.2)	(0.4)	(0.3)	(3.0)	(0.3)	(0.5)	(0.3)	(3.9)
Other Western	-7.5***	-9.5***	-7.4***	-3.5***	-10.0***	-12.1***	-11.7***	-6.2***	-5.0***	-6.8***	-2.7***	-0.6
	(0.1)	(0.1)	(0.6)	(0.2)	(0.1)	(0.2)	(0.8)	(0.3)	(0.1)	(0.2)	(0.8)	(0.3)
Other Non-Western	-13.4***	-15.4***	-10.7***	-7.2***	-16.9***	-18.7***	-15.9***	-10.3***	-9.1***	-11.2***	-5.2***	-4.0***
	(0.1)	(0.1)	(0.4)	(0.5)	(0.2)	(0.2)	(0.5)	(0.7)	(0.2)	(0.2)	(0.5)	(0.7)
Constant	52.4***	52.8***	52.7***	52.7***	53.2***	53.7***	53.7***	53.7***	51.6***	51.7***	51.7***	51.7***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)	(0.1)
Observations	956,539	715,994	621,078	602,504	496,712	372,157	321,965	311,407	459,827	343,837	299,113	291,097
R-squared	0.029	0.033	0.013	0.001	0.047	0.051	0.022	0.003	0.015	0.018	0.007	0.000

Notes: This table shows the gap in income percentiles by gender and ethnic background. Individuals with a Dutch ethnic background are omitted, so that the coefficients present the difference in average test income of the respective ethnic group compared with the individuals with a native Dutch background. Columns A2, B2 and A3, B3 show gaps for individuals with a homogenous ethnic background from the first, second generation, respectively. Column A5, B5 show gaps for children with a split ethnic background, with 50% of their great-grandparents being Dutch, and 50% of their great-grandparents having the respective migrant background. Robust standard errors in parentheses. $***$ p<0.01, $**$ p<0.05, $*$ p<0.1

first generation, and individuals with a mixed MST-native background close more than 70% of their background's respective migrant-native gap.

The results so far clearly depict large gaps in both primary school test scores and personal incomes for individuals from MST backgrounds. Moreover, a worrying fact might be that overall, current gaps in primary schools test scores are larger than current personal income gaps at 30. This does bode well for the future gap for these children with MST backgrounds. Nevertheless, as this paper does not give information to what extent primary school test score gaps are predictive of personal income gaps, the possible implications of the larger test score gaps are not clear.[25]

Wage Gaps. Table 1.4 has as the dependent variable personal wage, and uses a sample of all 28 to 32 year olds in 2018. Wage gaps for individuals with Moroccan, Surinamese and Turkish backgrounds are 10.1, 12.4 and 10.6 percentiles, respectively. Compared with the personal income gap, wage gaps are on average smaller for individuals with Turkish and Moroccan backgrounds, but larger for those of Surinamese background. A large gap of 21.1 percentiles is observed for individuals with a Western background, which seems most driven by 1st generation migrants. These findings corroborate the fact that income differences are partly explained by lower working hours. Panel B of Table 1.1 shows that individuals with a native Dutch background on average work 1466 hours in 2018, while those with Moroccans, Surinamese and Turkish background work significantly less, with average contracted hours of 908, 1162 and 1083 hours.

Table 1.4 includes columns that control for highest educational degree. While years of education of the native (13.2) is not that much higher than years of education of those with Moroccan (11.7), Surinames (12.4) and Turkish (11.7) backgrounds, the eventual degrees obtained explain a large part of the wage gap. This can be seen by comparing columns with (A1) and without (A2) controls for the highest obtained degree. When controlling for the highest educational degree, wage gaps are shrunk to only 1.5, 5.6 and 2.9 percentiles for individuals with Moroccan, Surinamese and Turkish backgrounds, respectively. This shows that educational gaps can account for 8.6, 6.8 and 7.7 percentiles of the wage gap for these three groups, respectively. As such, educational differences account for 85.1%, 54.9% and 72.6% of the wage gaps for those with MST backgrounds.

[25] Work on the evolution of migrant gaps across the lifecycle is scarce and a great avenue for future research.

Table 1.4: Hourly Wage Gap of 28-32 Year Olds in 2018

A. Pooled Genders

Dependent Variable: Personal Hourly Wage Percentile	(1) All	(2) All	(3) 1st Gen	(4) 1st Gen	(5) 2nd Gen	(6) 2nd Gen	(7) Mixed 50% Dutch	(8) Mixed 50% Dutch
Ethnic Background (ommitted: 100% Dutch)								
Moroccan	-10.1***	-1.5***	-17.9***	-3.2***	-7.8***	-1.1***	-12.2***	-4.2**
	(0.2)	(0.2)	(0.4)	(0.3)	(0.2)	(0.2)	(2.2)	(1.8)
Surinamese	-12.4***	-5.6***	-16.4***	-7.2***	-13.0***	-6.2***	-8.1***	-3.7***
	(0.2)	(0.1)	(0.4)	(0.3)	(0.2)	(0.2)	(0.7)	(0.5)
Turkish	-10.6***	-2.9***	-13.2***	-3.7***	-10.2***	-3.0***	-4.5*	-0.1
	(0.2)	(0.1)	(0.3)	(0.3)	(0.2)	(0.2)	(2.4)	(1.8)
Other Western	-19.1***	-4.1***	-23.9***	-6.7***	-8.9***	-4.4***	-3.4***	-2.1***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.6)	(0.4)	(0.2)	(0.2)
Other Non-Western	-12.1***	-7.7***	-13.4***	-9.4***	-11.3***	-7.6***	-7.9***	-4.6***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.3)	(0.3)	(0.5)	(0.4)
Constant	55.8***	54.5***	56.2***	55.3***	56.1***	55.7***	56.1***	56.2***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
FE For Highest Educational Degree		X		X		X		X
Observations	962,873	818,286	737,660	599,090	578,117	566,611	561,097	550,099
R-squared	0.071	0.420	0.104	0.434	0.012	0.414	0.001	0.417

B. Men

Dependent Variable: Personal Hourly Wage Percentile	(1) All	(2) All	(3) 1st Gen	(4) 1st Gen	(5) 2nd Gen	(6) 2nd Gen	(7) Mixed 50% Dutch	(8) Mixed 50% Dutch
Ethnic Background (ommitted: 100% Dutch)								
Moroccan	-11.5***	-2.0***	-19.0***	-3.6***	-9.1***	-1.8***	-13.1***	-4.9*
	(0.3)	(0.2)	(0.5)	(0.5)	(0.3)	(0.3)	(3.0)	(2.7)
Surinamese	-13.1***	-6.1***	-16.4***	-7.1***	-14.3***	-7.0***	-9.5***	-4.7***
	(0.2)	(0.2)	(0.5)	(0.5)	(0.3)	(0.3)	(0.9)	(0.7)
Turkish	-9.7***	-1.5***	-13.4***	-3.4***	-8.5***	-1.3***	-5.2	0.4
	(0.2)	(0.2)	(0.4)	(0.4)	(0.3)	(0.2)	(3.2)	(2.6)
Other Western	-21.1***	-3.6***	-25.7***	-5.1***	-10.0***	-6.2***	-3.4***	-2.4***
	(0.1)	(0.1)	(0.1)	(0.2)	(0.8)	(0.6)	(0.3)	(0.3)
Other Non-Western	-12.4***	-7.9***	-13.4***	-9.3***	-13.1***	-8.1***	-8.4***	-4.8***
	(0.1)	(0.1)	(0.2)	(0.2)	(0.5)	(0.4)	(0.7)	(0.6)
Constant	56.7***	55.4***	57.0***	56.1***	57.0***	56.6***	57.0***	57.1***
	(0.0)	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
FE For Highest Educational Degree		X		X		X		X
Observations	506,944	412,918	392,234	301,969	294,116	286,058	285,172	277,564
R-squared	0.088	0.371	0.127	0.385	0.013	0.360	0.001	0.361

C. Women

Dependent Variable: Personal Hourly Wage Percentile	(1) All	(2) All	(3) 1st Gen	(4) 1st Gen	(5) 2nd Gen	(6) 2nd Gen	(7) Mixed 50% Dutch	(8) Mixed 50% Dutch
Ethnic Background (ommitted: 100% Dutch)								
Moroccan	-8.7***	-1.7***	-16.8***	-3.6***	-6.3***	-1.3***	-11.2***	-4.2*
	(0.3)	(0.2)	(0.5)	(0.5)	(0.3)	(0.3)	(3.3)	(2.5)
Surinamese	-11.6***	-5.3***	-16.3***	-7.2***	-11.7***	-5.9***	-6.7***	-2.7***
	(0.3)	(0.2)	(0.5)	(0.4)	(0.4)	(0.3)	(0.9)	(0.7)
Turkish	-12.0***	-5.1***	-13.3***	-4.9***	-12.1***	-5.6***	-3.9	-1.9
	(0.3)	(0.2)	(0.5)	(0.5)	(0.3)	(0.2)	(3.8)	(2.6)
Other Western	-16.8***	-4.0***	-21.7***	-7.0***	-7.8***	-3.0***	-3.4***	-1.8***
	(0.1)	(0.1)	(0.1)	(0.2)	(0.8)	(0.6)	(0.4)	(0.3)
Other Non-Western	-11.9***	-7.7***	-13.9***	-9.9***	-9.5***	-7.2***	-7.3***	-4.4***
	(0.2)	(0.1)	(0.2)	(0.2)	(0.5)	(0.4)	(0.7)	(0.5)
Constant	54.9***	53.7***	55.3***	54.4***	55.2***	54.9***	55.2***	55.4***
	(0.0)	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
FE For Highest Educational Degree		X		X		X		X
Observations	455,931	403,001	345,426	294,839	284,003	278,590	275,926	270,591
R-squared	0.054	0.490	0.080	0.505	0.011	0.490	0.001	0.494

Notes: These tables shows the gap in wage percentiles by gender and ethnic background. Individuals with a Dutch ethnic background are omitted, so that the coefficients present the difference in average percentile of wage of the respective ethnic group compared with the individuals with a native Dutch background. Columns 3, 4 and 5, 6 show gaps for individuals with a homogenous ethnic background from the first, second generation, respectively. Column 7, 8 show gaps for children with a split ethnic background, with 50% of their great-grandparents being Dutch, and 50% of their great-grandparents having the respective migrant background. Panel A presents results for all genders, Panel B and C split the results out by gender. Robust standard errors in parentheses. $***$ p<0.01, $**$ p<0.05, $*$ p<0.1

III.B Recent Evolution of the Migrant Gap

Even though several papers have documented a substantial migrant gap in the Netherlands, very few papers have looked at time trends in the migrant gap. In this section, I show new evidence on the shrinkage of migrant income gaps across time.

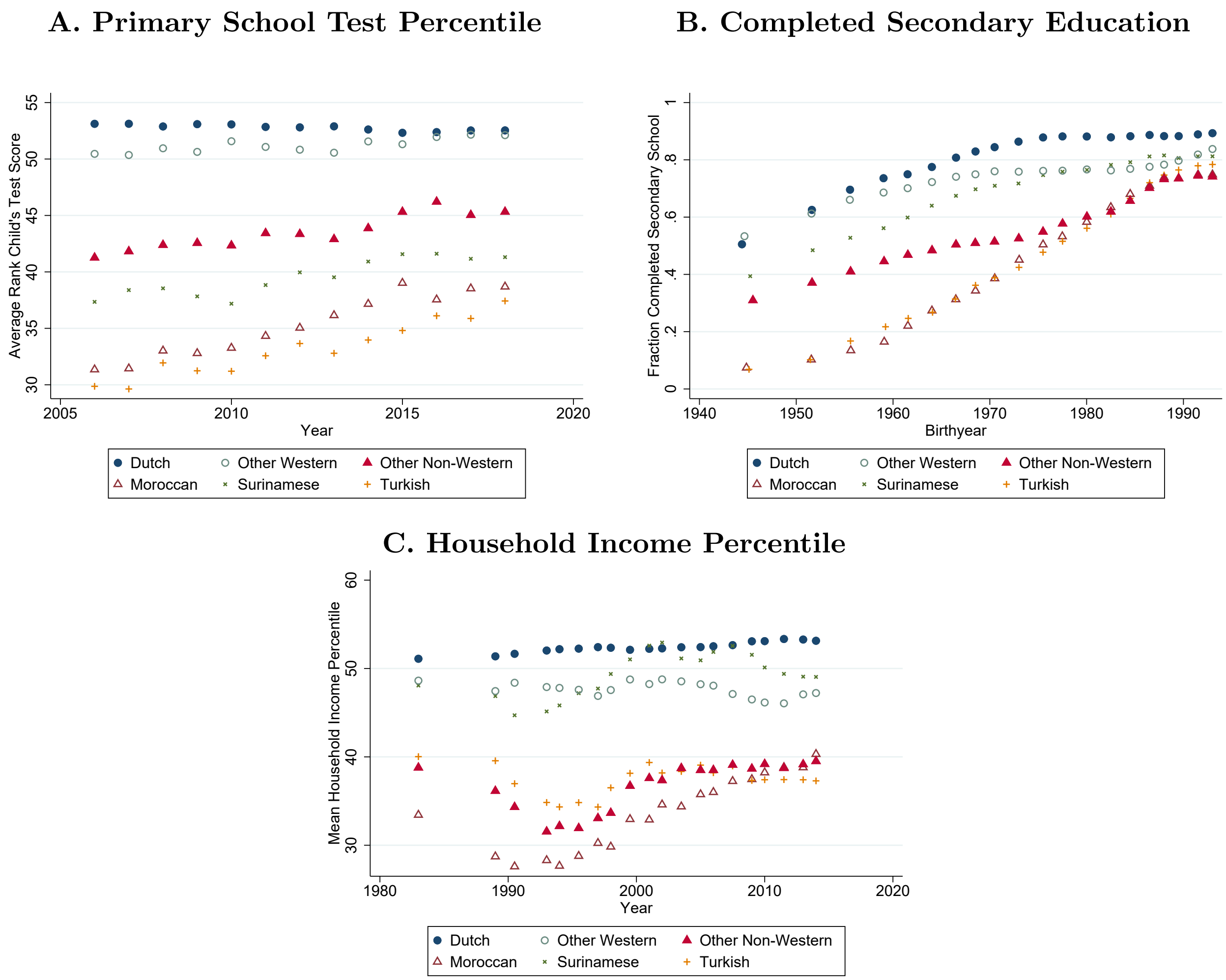
Average migrant gaps in primary school test scores have rapidly decreased between 2006 and 2018, as depicted in Panel A Figure 1.2. This graph shows that over these 12 years, average percentiles of test scores have increased at a yearly rate of 0.70, 0.39 and 0.59 percentiles for children with Moroccan, Surinamese and Turkish backgrounds, respectively. For children with a Moroccan background, the speed of increase is especially high: if linearly extrapolated, the gap remaining in 2018 with natives' descendants would shrink to zero in less than 20 years.

Panel B of Figure 1.2 depicts the changes in high school completion rates for a wider set of birth cohorts. The increase in the high school completion rate of Moroccans and Turkish has been very fast in the 1990s and 2000s (for children born in 1970s and 1980s), with these groups almost catching up to levels of the natives' descendants, yet recently the increases in high school completion rates are getting smaller, stalling convergence. A large fraction of the increase in high school completion rate can likely be traced back to the fact that many first-generation economic migrants did not complete high school in their respective non-Western countries, while education until age 16 is mandatory for children living in the Netherlands.

Finally, Panel C of Figure 1.2 shows that the evolution of the household income percentile gap of 30 to 40 year olds has gone through a U-shaped curve. Indeed, the gap was on average larger in the mid-nineties than in the mid eighties. Between 1990 and 2018, the overall income gap between the ethnically native and ethnically non-Western has been shrinking at a small pace.

Overall, the recent evolution of migrant gaps paint a picture of decreasing migrant gaps. It is important, however, to remark that a decrease of the gap can have plenty of different causes. For example, compositional changes in the migrant group, changed selection of migrant in- and outflows, general trends in all migrants' outcomes, and migrant learning and integration can all have the effect of decreasing migrant gaps over time. The next section will delve deeper into the potential causes of changes in migrant gaps over time, and will investigate in more detail the effect of time spent in the Netherlands on children's outcomes.

Figure 1.2: Migrant Gaps Across Time



Notes: This graph shows the evolution of three different indicators of socioeconomic status, by ethnic background. These graphs are binned scatterplots, with, for each graph, each observation representing an equal number of people. Panel A shows the average primary school test score percentile between 2006 and 2018. Panel B shows the high school completion rate, by birth year. Panel C shows Household Income Percentiles, by the year of measurement.

IV Time in Country and Descendants' Outcomes

Nearly all papers that measure the effect of time spent in a host country on migrant childen's outcomes measure only the effect on children in the first generation (eg. (Böhlmark (2008), Aslund, Bohlmark and Skans (2015) and Lemmermann and Riphahn (2018)). For example, Böhlmark (2008) measures how educational attainment and economic outcomes of a child improves with time spent in Sweden *by that child*. It exploits the fact that first-generation siblings of different ages spend different durations in Sweden before taking a test at a certain age. The difference between outcomes of the younger and older siblings is then defined to be the time-in-country effect.[26]

In the current situation in the Netherlands, however, the majority of children with a migrant background are born in families that have been in the country for many years. As such, there is no difference in the time spent in the country between second or later generation siblings themselves. Yet, there is a difference in the number of years their extended family spent in the host country before these children were born. In this paper, these pre-birth years are called the *time spent by the extended family* in the Netherlands. Little is known about whether the time spent by the extended family in the Netherlands affects the outcomes of the child. Can we expect a child in a family of Moroccan descent that has been in the Netherlands for 40 years to do better compared to a child in a similar family that has been in the Netherlands for merely 20 years? This section investigates this question, first by a naive interpretation of the raw data, and later in a specification that controls for selection effects and time trends. Only the test school sample as defined in Section II.B is used in this section.[27]

IV.A Time of Entry of Family and Test Scores: an Illustration

First, I show why mere correlations between time in country and descendants' primary school test scores can be misleading.[28] Panel A of Figure 1.3 shows the average Cito test scores of 12 year olds, by year of first migrant entry within his/her family.[29] The Figure depicts that, while descendants of migrants who came in the Netherlands before 1960 do better than the average Dutch child, descendants of migrants whose families migrated between 1963-1980 do significantly worse. Such a non-linear pattern is highly suggestive of changes in cohort composition, and provides no clearcut evidence that time in the Netherlands has an effect on children's outcomes. For example, the sharp difference between 1960 and 1963 suggests a change in quality (and/or composition) of migrant inflows. Indeed, Panel B of Figure 1.3 shows that the migrant countries by migrant inflow have drastically changed over the years.

[26] Similar methods are used by Chetty and Hendren (2018), who find that neighbourhoods have causal effects on income, and that outcomes of children tend to converge to the destination outcomes the longer the child has stayed there. Chetty and Hendren (2018) however, correlate a neighbourhood differential with the siblings outcome differential. Such a method cannot be replicated if no information on average departing country educational achievement is known.

[27] The individuals for whom there is no earliest migration movement found are dropped. This includes not more than 6% for children with a background from MST.

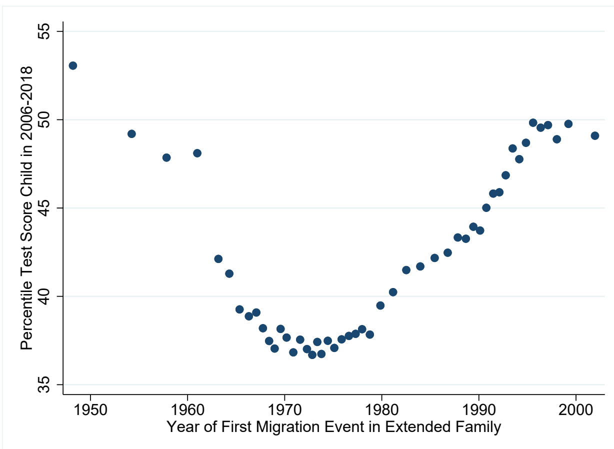
[28] Appendix Figure 1.3 is a histogram and shows the distribution of how long current children's family has been in the Netherlands, for all children in the primary test school sample. It shows that there were two peaks, on in the seventies, ad one in the ninetees. Due to data limitations, individuals are only connected with their parents and four grandparents to measure the earliest recorded move to the Netherlands. More details about the method of connecting the can be found in Section II.B.

[29] The test scores are measured in 2006-2018, and as explained in Section II, and were converted into percentile for each year-gender group.

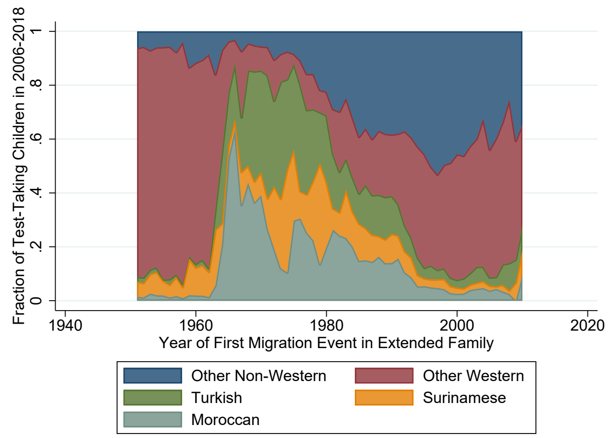
Around 1963, there was a sharp increase in the number of Moroccan and Surinamese immigrants, and Panel A suggests that this has led to lower average test scores for the descendants of migrants of this year.

Figure 1.3: Correlation Time of Entry of Family and Test Scores

A. Test Scores



B. Composition



Notes: Panel A depicts a binned scatterplot of average test score percentiles for children who have a non-Western background but who are born in the Netherlands, based on the year of first migration event within the extended family. The year of first migration event within the extended family is the first migration event in the data for all individuals in the direct lineage of the child up to the grandparents, as explained in Section II.B. The figure in Panel B shows the composition of backgrounds for each of the migration years that are linked to children with primary test scores.

Even the types of migrant inflows from within particular sending countries can change over time, with associated long-lasting effects on the test scores of descendants. An interesting example can be seen in Figure 1.4. This Figure depicts test scores of children of Moroccan and Turkish descent by the year of entry of the first migrant in his/her family tree. Clearly, the test scores of Moroccan descendants of migrants before 1973 are higher than those of descendants of Moroccan migrants who migrated just after 1973. This drop after 1973 might be a late effect of a labour migration policy change, as the Netherlands stopped its official Moroccan-Dutch labour migration agreement in 1973 together with its selection of guest workers by Dutch officials.[30] For Turkey, there is no such large difference in cohort quality after 1973.[31] Interestingly, descendants of recent cohorts of migrants score on average much

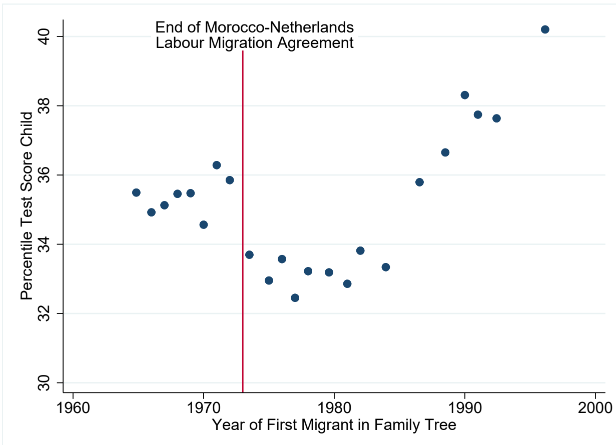
[30]Due to the high demand for low skilled work, Dutch officials went to Morocco and Turkey to select the guest workers before 1973. Migrants coming to Netherlands after 1973 however, could not use the labour migration agreement as an opportunity to work in the Netherlands.

[31]This might be because there remained legal ways for Turks to come to the Netherlands, mainly for Kurdish and

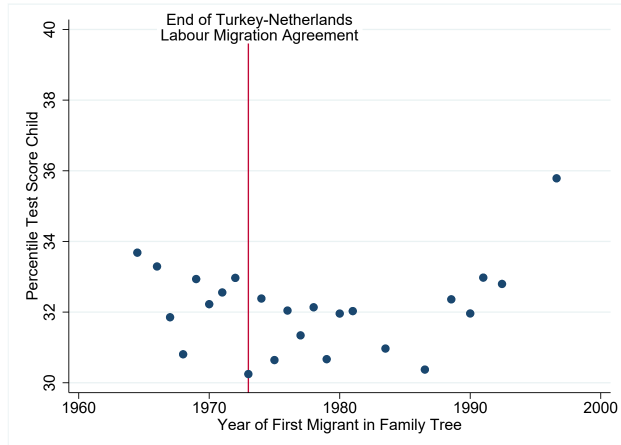
higher than descendants from earlier cohorts.

Figure 1.4: Correlation Time of Entry and Test Scores: Morocco and Turkey

A. Morocco



B. Turkey



Notes: The figures plots binned scatterplots of average test score percentiles of children who have a Moroccan (or Turkish) background but who are born in the Netherlands, by the year of first migration event within the extended family. The year of first migration event within the extended family is the first migration event in the data for all individuals in the direct lineage of the child up to the grandparents, as explained in Section II.B.

Clearly, these documented changes in migrant composition do not allow for a clear relationship between time spent in the host country and test score percentile to be observed in raw data such as in Figure 1.3.[32] The next paragraph will develop a method that can arguably find a reliable way to measure such a relationship.

IV.B Effect of Time in Country and Descendants' Outcomes

This paragraph proposes a general framework to estimate the causal effect of time spent in the host country of a migrant family on the economic (or educational) outcomes of their descendants. In the framework below, the individual outcome y_i at time t linearly depends on the number of years his/her migrant family has been in the country τ_i, a fixed effect α_f for the individual within that family that

Catholic asylum seekers.

[32] Appendix Figure A1.1 is an example where raw data can tell something about the integration of migrants' descendants over time. It plots distributions of outcomes for two generations (grandfather and child) of a constant set of families. The figure depicts that integration is limited, as the shape of the histograms seems only slightly more uniform for the children.

first entered the Netherlands, as well as covariates X_i which include only birth order fixed effects.[33] $\hat{y}_{e,t}$ is the estimate for the average economic outcome for an individual with constant characteristics from ethnicity e at time t, as will be discussed later.

$$y_{i,t} - \hat{y}_{e,t} = \alpha_f + \sum \beta_\tau I(\tau_i = \tau) + \kappa_i X_i + \epsilon_i \tag{1.1}$$

The first-entry fixed effects α_f are used to control for the individual's extended family. Such a specification thus exploits that there are children of different ages within extended families. Indeed, due to the inclusion of these first-entry fixed effects, all variation is from within extended family relations (siblings, or cousins), and follows from the fact that related children who are born in different years have the economic (or educational) measurement taken in different years at the same age.

Controlling for the year of measurement is not possible, because it is is perfectly collinear with relative time τ_i if family fixed effects are included.[34] A potential problem with this is that time trends in test scores for certain groups might bias results. For example, if for some reason unrelated to the time spent in the host country, there is an upward drift over time in test scores for children with a Moroccan background, this upward trend would wrongly be attributed to the time spent in the host country effect. This is similar to the problems faced in Aslund, Bohlmark and Skans (2015) and Van den Berg et al. (2014), who argue that in their setting the changing quality of migrants could lead to biased estimates.

To solve this problem, the predicted score of a median migrant across time is deducted, which is an extension of the methodology used by Aslund, Bohlmark and Skans (2015) and Van den Berg et al. (2014). More specifically, in specification 1.1, background-specific trends in test scores are taken into account by deducting $\hat{y}_{e,t}$, which is the predicted outcome for an individual from the same heritage, born to parents with constant education and income percentile, who have migrated in the year of birth of the child.[35] The identifying assumption for β_τ to represent causal effects of time in the host country on test score outcome is that $\hat{y}_{e,t}$ is an unbiased estimate for the average test score achievement of an

[33] Adding birth fixed effects is a standard method, as used in Böhlmark (2008). These fixed effects account for the fact that there are structural differences in the outcomes of first-born children. The variation that is left therefore only depends on the number of years between the births of children within families.

[34] This problem is a variation of the age-cohort-time problem, as discussed in Borusyak and Jaravel (2017).

[35] In this specification, the outcomes are predicted for an individual with the same heritage, with two parents with a high school education and an income percentile of 35 who migrated in the same year as the child was born. Such specifications are the average of migrant parents in our cohort, as seen in Table 1.1.

individual with the same quality across time. Therefore, $\hat{y}_{e,t}$ should accurately measure the general background-specific trends in test scores unrelated to length of stay. Conditional on $\hat{y}_{e,t}$ and α_f, all selection effects should be accounted for.

The specification 1.1 is estimated on primary school test scores from 2006-2018 in separate regressions for the different categories of non-Western backgrounds in the sample. Appendix Figure A1.2 depicts the distribution of years of first entry for the families of the school children that were used in this estimation. As the children who took the primary school tests were born in between 1994 and 2008, there is a large enough sample size to include in my analysis families that migrated up to 40 years before the birth of the child. Yet, despite the fact that this methodology can obtain estimates for up to 40 years before the birth of these children, it is important to remember that all coefficients are measured within a 12-year time window, so that the measured time in country effects only pertain to this restricted time period.

Figure 1.5 shows the graphical representation of the vector β_τ, as well as estimates for the linear version.[36] Panel A shows that on average, if a child is born in a family with a non-Western background, test score percentiles improve with approximately 9.6 percentiles if the family has resided for 40 years in the Netherlands before birth compared to a similar family that has only just arrived in the Netherlands before birth. The results show that having spent one year more in the Netherlands by a family of non-Western background is causes a 0.24 percentile higher primary school test score, and that positive effects are still accumulating decades after the initial entry of the migrant family. Educational assimilation or "catch-up" is still happening 40 years after the initial migration move.

Panels B-E in Figure 1.5 show that there is large heterogeneity in the effect of time on Cito test scores by ethnic background. Within families with backgrounds from Morocco and Turkey, children's outcomes clearly improve with time at yearly rates of 0.40 and 0.37, respectively. A smaller effect is found for children in families with Surinamese and other Non-Western backgrounds, whose outcomes improve only at a yearly rate of 0.27 and 0.24 percentiles yearly, respectively. The reason for this is unclear, but it might be related to the fact that children with Surinamese and other non-Western backgrounds have smaller initial test score gaps to begin with.

It is important that the estimates in this section are interpreted correctly: the effect of being born in a family that has lived in the Netherlands for a longer time versus a shorter time must be separated

[36] In the linear version, $I(\tau_i = \tau)$ is replaced by the linear variable τ.

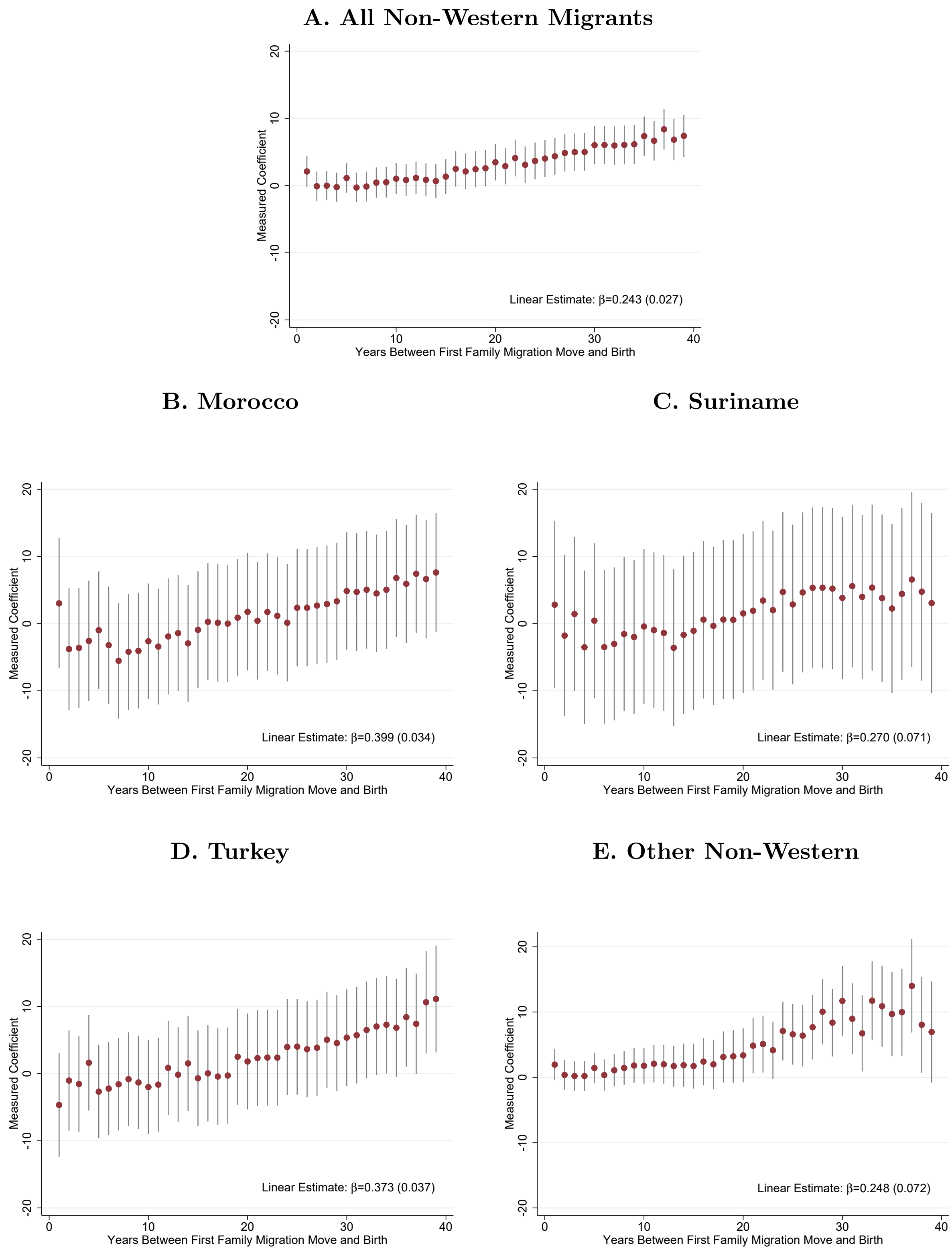
from average trends in migrants' outcomes over time that also apply to new migrants.[37] For example, changes in educational policies so that all children with migrant backgrounds improve on average should not be attributed to a time in country effect, but to a general time trend that is filtered out in the specification due to the deduction of $\hat{y}_{e,t}$.

The filtering out of these general time trends is important, as Appendix Figure A1.5 shows what happens to the results if the background-specific time trends are not accounted for and $\hat{y}_{e,t}$ is excluded from specification 1.1; the measured time in country effect is then much larger for every background. The explanation for this is that average primary school test scores have been trending upwards for all non-Western ethnic backgrounds at a fast pace in between 2006 and 2018, as seen when looking at the trends in test scores in Panel A of Figure 1.2. A large part of these time trends cannot be explained by characteristics of the migrant family, and thus leads to increasing estimates of $\hat{y}_{e,t}$ across time. The interpretation of this paper is that this part of the increase in average test scores has been "secular", and should therefore be left out of the time in country effect.

Overall, the findings in this section show that continued improvements in children's test scores within non-Western families are the norm rather than the exception. These findings strengthen the intuition that integration is a process that continues for decades, and can differ by background. The mechanisms through which children with foreign backgrounds do better if they are born in families that have been for an extended time, are plenty, but were not investigated in this paper. Potential mechanisms include a better ability to navigate the educational system, increased language abilities of parents and extended family, as well as the direct relations with natives through friendships and relationships.

[37] There is a dip for children after two years in the Netherlands. This is counterintuitive, as the interpretation is that children who are born in the first year after the move do better than children born two years after the move.

Figure 1.5: Effect of Time in Country on Percentile Test Scores



Notes: This Figure depicts the estimates of β_τ of model 1.1 when estimating the model on Cito test score percentiles, by ethnic background. These coefficients measure the effect on the child's test score percentile of time spent in the host country by the extended family before the birth of the child. In other worths, these coefficients show the additional increase in test school percentile that every additional year within a country brings, relative to the base year (which is 0). Estimates of the standard errors of point estimates for each year are shown by vertical lines. The linear estimate β included on each graph is the estimate where vector $I(\tau_i = \tau)$ is replaced by a coefficient τ. Results of the same model without controlling for time trends can be found in Appendix Figure A1.5.

V An Intergenerational Perspective on Integration

In this section, the economic and educational mobility of immigrants' and natives' children are analyzed using the intergenerational framework of Chetty et al. (2014*b*) and Chetty et al. (2020*b*). This framework measures educational and income differences, conditional on parental income. First, an explanation of this framework and its application to the economic integration of migrants is discussed. Subsequently, I estimate the model in the Netherlands, and provide estimates for steady-state outcomes. The section ends with an extension of the model which includes individuals with non-homogeneous backgrounds. For expositional reasons, in this section, the observations of all individuals with a Moroccan, Surinamese and Turkish background are pooled.[38]

V.A Conceptual Framework

The model of economic mobility in this section is adapted from Chetty et al. (2020*b*), who introduced the model to explain long-run black-white differential in the United States. It follows the tradition of the canonical intergenerational income models after Galton (1886) and Becker and Tomes (1979).

Consider a discrete time setting where t indexes generations. $y_{i,t,e}$ denotes the family income percentile y of family i of generation t of ethnic background e. The linear relationship between the incomes of different generations within a specific ethnicity e is then written in the following equation.

$$y_{i,t,e} = \alpha_e + \beta_r y_{i,t-1,e} + \epsilon_{i,t} \tag{1.2}$$

The simplicity of equation 1.2 is that one does not need to track the entire income distribution to follow the evolution of mean outcomes by ethnicity. Indeed, as $\epsilon_{i,t}$ denotes an ideosyncratic shock, we can write the means as $\overline{y_{i,t,e}} = \alpha_e + \overline{\beta_e y_{i,t-1,e}}$. Iterating over generations, it is easy to show that the steady state mean outcome by ethnicity evolves to $\overline{y_{e,ss}} = \frac{\alpha_e}{1-\beta_e}$, if mobility estimates α_e and β_e are constant across all generations.

V.B Empirical Mobility Patterns

Before turning to the steady state outcomes, which follows from a model that only uses family income percentiles as variables, I first turn to graphical evidence of mobility patterns in test scores and

[38] The graphs split out by individual country can be found in Appendix Figures A1.3 and A1.4.

individual incomes. Figure 1.6 depicts mobility patterns in primary school test scores and personal incomes using the samples defined in section II.B. The Figure shows binned scatter plots that provide information on both the absolute mobility (typically the level of average children's outcomes at certain parental ranks), as well as relative mobility (the slope of the mobility curves).

Panels A of Figure 1.6, which depicts test score mobility, shows a steeper slope for the natives' descendant group than for the migrants' descendant groups. Every 10 extra percentiles of parental income are associated with 2.4 extra percentiles on the primary school test score for natives' descendants, and 1.5 extra percentiles for individuals with a background from MST.[39] For children of both migrant and native backgrounds, there thus is a large difference between growing up in a poor or a rich family. However, for individuals with an MST background, coming from a rich background proves to be relatively less advantageous, on average.[40] Yet, the starkest finding in this graph is the absolute mobility gap, or the differences in outcomes conditional on parental income. The absolute mobility gap in the 1st and the 10th decile between natives' desdendants and individuals with MST background is 9.56 and 13.10, respectively. Panels A2 and A3 show that average absolute mobility gaps in test scores are larger for girls than for boys.

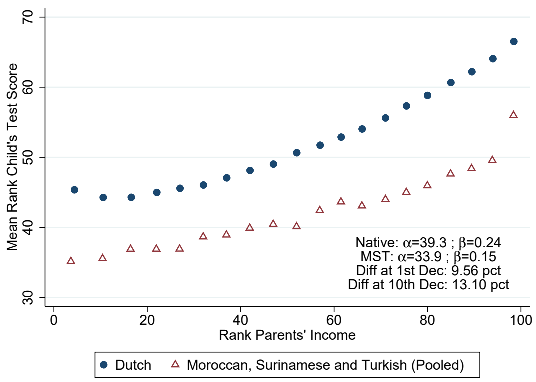
Panel B of Figure 1.6 shows that the findings for income mobility are different from the findings of test score mobility. Parental income percentile has less of an influence on personal income percentile than on the test score percentile. Every 10 extra percentiles of parental income are associated with 1.9 extra percentiles in the personal income ranking for individuals with a native background, and 1.1 extra percentiles for individuals with a background from MST. Additionally, the absolute income mobility differences between those with native and MST backgrounds are now smaller, at 4.21 and 8.18 percentiles in the 1st and 10th decile, respectively. Interestingly, panels B2 and B3 show that the absolute income mobility gaps for MST women are now much smaller than those for men, which is opposite to the finding in test score mobility.

[39] The absolute and relative mobility in this paper is similar to the estimates from Carmichael et al. (2020), who investigate income mobility in the Netherlands.

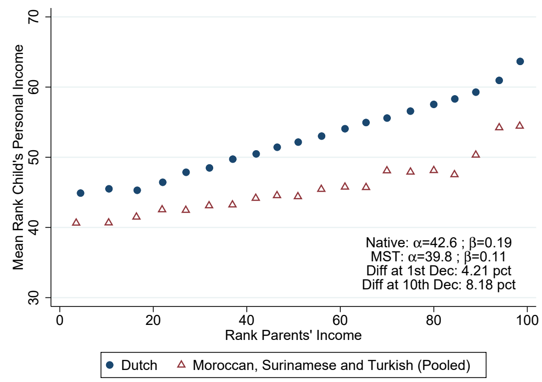
[40] Such a finding would be consistent with the theory of ethnic capital of Borjas (1992). Indeed, he posits that mobility depends on the average income of individuals within the same ethnic group.

Figure 1.6: Mobility Gaps

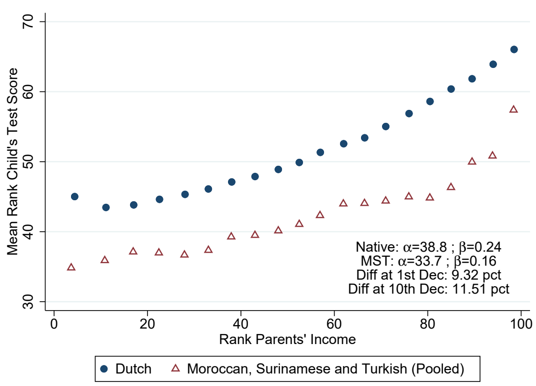
A1. Test Score



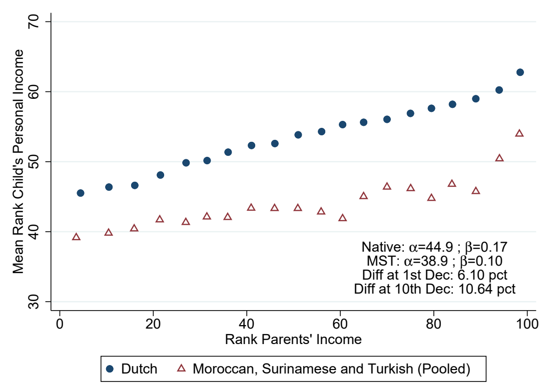
B1. Income



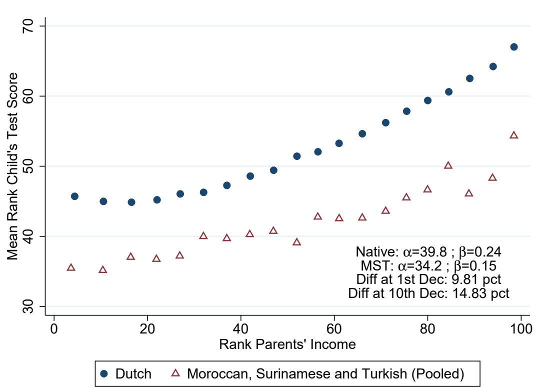
A2. Test Score, Boys



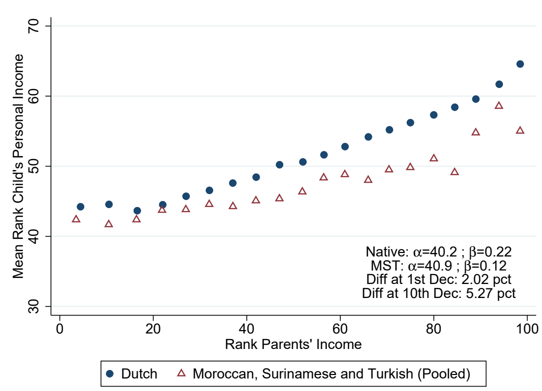
B2. Income, Men



A3. Test Score, Girls



B3. Income, Women



Notes: These Figures depict percentile average outcomes by parental income ventile, over migration background. α and β show estimates of the respective coefficients in equation 1.2. Additionally, the absolute mobility differences between individuals with a native versus a MST background are calculated for the 1st and 10th decile of the parental income distribution. Figures A are estimated on the test score sample and project average percentiles on the Cito test. Figures B are estimated on the income samples found in II.B and project average primary income percentiles. Data on test scores are from 2006-2018, and data on incomes are from 2018. MST stands for individuals with a background from Morocco, Suriname and Turkey.

As a point of comparison, these income mobility gaps are larger than the average absolute black-white mobility gaps in the United States according to Chetty et al. (2020*b*), who find mobility gaps of 4.2 and 5.6 percentiles (at the 25th and 75th percentile). The comparable MST-native income gaps in this paper are 9.56 and 13.10 percentiles (in the 1st and 10th decile). The similar gender trends, with larger income gaps for men than for women, are also observed, but while Chetty et al. (2020*b*) finds no differences in mobility for black versus white women, there is a significant migrant-native income mobility gap for women in the Netherlands. These Dutch findings also contrast with the Swedish findings of Bolotnyy and Bratu (2018), who find similar mobility curves for immigrants' and for natives' children, as well with American findings in Abramitzky et al. (2021), who find higher upward rates of mobility for Western immigrants in the United States. Additionally, these results differ from existing Dutch findings on the descendants of Indonesian immigrants (van Ours and Veenman (2010)), who find that no income mobility differences exist with natives' children after controlling on parental education.

Next, I turn to differences in mobility between different migrant generations with MST backgrounds in Figure 1.7. Intuitively, one could expect mobility curves for children of the first generation migrants to be higher due to the typical under-placement of the parents in the income distribution. This is what is observed, for example, in the US for Asian-American migrants in Chetty et al. (2020*b*), with absolute mobility curves that are higher for the 1st than for 2nd generation migrants. Empirically in the Netherlands, however, Panel A of Figure 1.7 shows that mobility there is no clear difference in mobility curves for different generations of migrants from Morocco, Suriname or Turkey. Later generations of individuals with MST backgrounds seem to have slightly higher mobility than earlier generations, with the average test score mobility gap decreasing from 10.73 to 8.62 percentiles from the first to the third generation, but there is no large convergence of mobility trends towards natives' levels.[41]

Mobility patterns for children with a heterogeneous background show another trend. Figure 1.7 Panel B shows that mobility curves for children with a mixed ethnic background (Dutch and one of Morocco, Suriname and Turkey) are clearly better than for children with homogeneous background from MST. Indeed, on average, children with such an heterogeneous (mixed) background from MST countries close 48.7% of the migrant gap in income mobility of their homogeneous MST counterparts.[42]

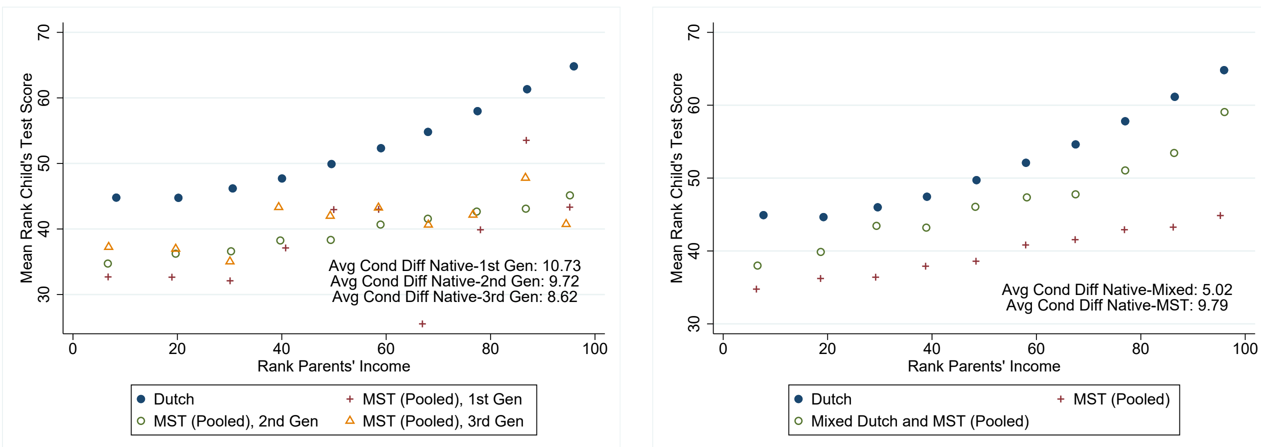
[41]This is also the case when looking at Appendix Figure A1.3, where similar Figures are made for the three different countries separately.

[42]Additionally, children with heterogeneous ethnic backgrounds grow up in households with higher average incomes.

This evidence is surprising, as some evidence has shown that mixed children do not have a particular advantage compared with children from ethnic minorities (Fryer et al. (2012)). Appendix Figure A1.4 shows that especially mixed children with a half of their heritage from Turkey and Morocco are responsible for closing a large part of their migrant gap in economic mobility.

Figure 1.7: Mobility Gaps: by Generations, and Mixed Children

A. by Migrant Generation **B. by Ethnic Makeup**



Notes: These Figures plot percentile Cito scores over parental income deciles. Panel A plots these mobility statistics by migrant generation. Panel B plots these mobility statistics by ethnic make-up, including individuals with a heterogeneous (mixed) background. Included are calculations of average absolute mobility gaps. Data on test scores are from 2006-2018, and data on incomes are from 2018. MST stands for individuals with a background from Morocco, Suriname and Turkey.

V.C Steady State Income Gap

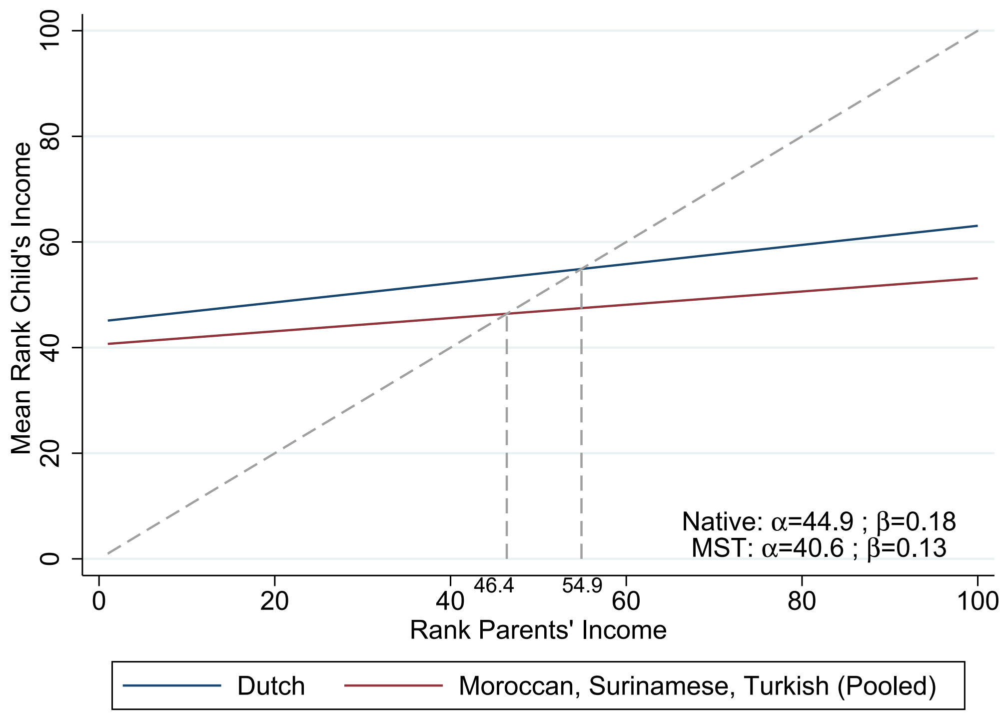
One of the largest advantages of the mobility model of equation 1.2 is that it allows for a simple calculation of long run averages of each group, if one assumes mobility patterns to remain constant in the long run. Figure 1.8 shows the results of an estimation of equation 1.2 on the family income percentiles for two ethnic backgrounds (Dutch, and pooled Moroccan, Surinamese and Turkish). Both the α and β estimates in the regressions of family income percentile for the two groups are presented in the graph. Because these regressions are based on family incomes for both parents and children, the parameters here differ slightly from the parameters from Figure 1.6. Moreover, note that individuals with mixed backgrounds are not included in the estimation sample in this graph. This is to draw

This leads to the observation that these children close about 70% of the unconditional test score and personal income gaps.

comparisons with earlier literature, as well as to allow for the extension of the model later in this Section.

Graphically, the intersection of the 45 degree line and the linear mobility graph for each group is the steady state outcome $\overline{y_{e,ss}}$. This long-run average family income percentile the certain groups will trend towards amounts to 46.4 for individuals with a homogeneous MST background, and 54.9 for individuals with a native background. This leads to an estimate of a steady state gap of 8.5 percentiles between the average percentile of the family income of the native Dutch and that of the selected migrant groups.

Figure 1.8: Steady State Migrant Income Gap



Notes: This figure, based on Chetty et al. (2020*b*), shows how intergenerational mobility of family income determines the steady-state of the income rank by ethnic group, if ethnic groups do not mix. The solid lines represent children's expected family income rank conditional on the parental income rank. The intersection with the 45 degree line and these income mobility graphs represent the steady state income rank by ethnical group. These family incomes are estimated on all individuals between 30-40 years old in 2018 from their respective ethnic group. MST stands for individuals with a background from Morocco, Suriname and Turkey.

This steady state gap compares favourably to the black-white family income gap calculated in Chetty et al. (2020*b*) for the United States. A major reason why the steady state black-white family income gap in the US is predicted to be larger (19.2 percentiles) than the gap this paper predicts for the selected migrant groups in the Netherlands (8.5 percentiles) is due to the many single-parent households among blacks in the US. Indeed, family incomes of blacks in the US are partly suppressed by the low marriage rate, while this is not the case for MST ethnicities in the Netherlands.

It is important to stress that the assumptions for this simple model are quite restrictive, and that predictions are destined to change depending on circumstances. Indeed, despite having seen evidence of no large improvements of the mobility patterns of later generations compared to earlier generations in the snapshot of Panel A of Figure 1.7, there does seem to be a general upward trend in the test score percentiles of minorities within a short period of time, as evidenced by Figure 1.2. Moreover, this model entirely disregards that children may have heterogeneous (mixed) backgrounds, an assumption that I will relax in the next paragraph.

V.D Extension with Heterogeneous (Mixed) Ethnicities

The model of intergenerational mobility by ethnicity as adapted from Chetty et al. (2020*b*) can be misleading as it simplifies the demographics of a country to two distinct homogeneous ethnicities. By doing that, it abstracts away from people with mixed ethnic backgrounds. Yet, the evidence in Panel B of Figure 1.7 suggests that children with a mixed ethnic background close a large part of the mobility gap. Abstracting from ethnically mixed individuals therefore underestimates how well descendants of migrants integrate, and overestimates the long-run migrant gap.

In this section, I extend the model with the possibility to form ethnically mixed families, which changes the migrant gap dynamic.[43] I simulate the family income percentile dynamics up until the 3rd generation, maintaining the assumption that the relationship between children's family income percentile and parents' family income percentile is linear and constant as in the Equation 1.2.

Rather than having two ethnicities, in the extended model with up to three generations there are five distinct ethnic backgrounds that individuals can have. Two categories have a homogeneous background namely natives n, ethnically migrant m, and other categories include only individuals with mixed backgrounds who include 50%-50% native-migrant nm, 75%-25% native-migrant $nnnm$ and 25%-75% native-migrant $nmmm$. With the possibilities to have individuals with a mixed background, it is necessary to model population dynamics aside from income dynamics.[44]

[43] Note that this model is still a simplification, as it does not take into account (1) differential fertility trends and (2) the possibility of more than two ethnicities. Especially incorporating differential fertility trends seems a promising avenue for research, as there is high variability in fertility trends across ethnicities.

[44] Note that these dynamics change with each generation: while in the first generation (when migration occurs), only homogeneous natives n and migrants m are in the model, in the second generation, there are also additionally 50%-50%

I calibrate the following parameters to the data: current estimates of α_e and β_e for the five backgrounds, as well the intermarriage rate $\gamma_{m,n} = 0.15$ where $\gamma_{x,y}$ is the fraction of ethnicity x that forms a couple with ethnicity y.[45] The match rates from the ethnically heterogeneous people are assumed to be proportional with population size, so that $\gamma_{nm,m}$ and $\gamma_{nm,n}$ are equal to the population shares of m and n, respectively.

Intermarriage and homogeneous marriage is assumed to be uniformly distributed across the income distribution within an ethnic group, and each marriage produces two children. The starting income percentiles for each ethnic group are obtained from the parental income percentiles in Table 1.1, which can be mostly seen as the first generation of migrants and their contemporary natives, so that $Y_{1,m} = 26.21$ and $Y_{1,n} = 55.18$. I work with an initial migrant fraction of the population $\theta_{1,m} = \frac{1}{10} = 1 - \theta_{1,n}$. Finally, I adhere a weight of 1 to all descendants, so that migrants' descendants with heterogeneous or homogeneous ethnic backgrounds are equally weighted in the calculation of the average outcomes.[46]

In the first generation the difference in average income rank between natives' descendants and those with a migrant background is $\overline{Y_{1,n}} - \overline{Y_{1,m}} = 28.97$, which comes straight from the assumptions for starting income percentiles. For the second and third generation, the difference in average income ranks between descendants of natives and descendants of migrants can be solved for analytically.[47] The main difficulty rests in the calculation the population sizes of the different groups. For example, with a fraction γ of the $\theta_{1,m}$ first generation migrants marrying homogeneously, there will be $\theta_{1,m}\gamma$ homogeneous descendants of these migrants in the second generation, and $2\theta_{1,m}(1-\gamma)$ heterogeneous (mixed) descendants. This will lead to the following analytic equation for the average income of 2nd generation migrant descendants:

native-migrant nm, who are born out of intermarriages. In the third generation, 75%-25% native-migrant $nnnm$ and 25%-75% native-migrant $nmmm$ join the model.

[45] The homogeneous marriage rate of migrants, $\gamma_{m,m} = 0.85$, is calibrated from the data in the sample used for Figure 1.8. Appendix Figure A1.6 shows the intermarriage rates across the different initial populations, and how this rate has increased over time.

[46] While choosing the weights on heterogeneous children (0.5 or 1) might seem to be an innocuous assumption, the choice has a sizeable effect on the measured income difference between descendants of migrants and natives. This is due to the fact that each "mixed" child has only one parent of each ethnicity, while a "non-mixed" child has two parents from the same ethnicity. Therefore, parents who intermarry will have the same amount of children with a migrant background as two individuals of his or her ethnicity. However, if we think all types of descendants of migrants are equally important as descendants of migrants, this weight of 1 to every descendant is a logical assumption.

[47] Here, the notation is changed slightly and the homogeneous marriage rate of migrants is simplified to $\gamma_{m,m} = \gamma$, as well as the initial fraction of migrants is simplified to $\theta_{1,m} = \theta$.

$$\overline{Y_{2,migrants\_des}} = \underbrace{\frac{2\theta(1-\gamma)}{2\theta(1-\gamma)+\theta\gamma}}_{\text{Mixed Descend.}} Y_{2,nm} + \underbrace{\frac{\theta\gamma}{2\theta(1-\gamma)+\theta\gamma}}_{\text{Homogeneous Descend.}} Y_{2,m}$$

In this equation, the average incomes of mixed individuals in generation 2 can be written as in model 1.2, where it is important to see that intermarriage being uniformly distributed across both parental distributions will lead to the parental distribution being an average of $Y_{1,m}$ and $Y_{1,n}$.

$$Y_{2,m} = \alpha_{nm} + \beta_{nm}\frac{Y_{1,m}+Y_{1,n}}{2}$$

After calculating all population sizes for each population group, as well as obtaining the associated average incomes for each group, the income percentile gap can be written as the difference of two weighted averages of several types of descendants as follows.

$$\overline{Y_{2,native\_des}} - \overline{Y_{2,migrants\_des}} = \frac{(1-2\theta+\gamma\theta)Y_{2,n}+2\theta(1-\gamma)Y_{2,nm}}{1-\theta\gamma} - \frac{2(1-\gamma)Y_{2,mn}+\gamma Y_{2,m}}{2-\gamma} \tag{1.3}$$

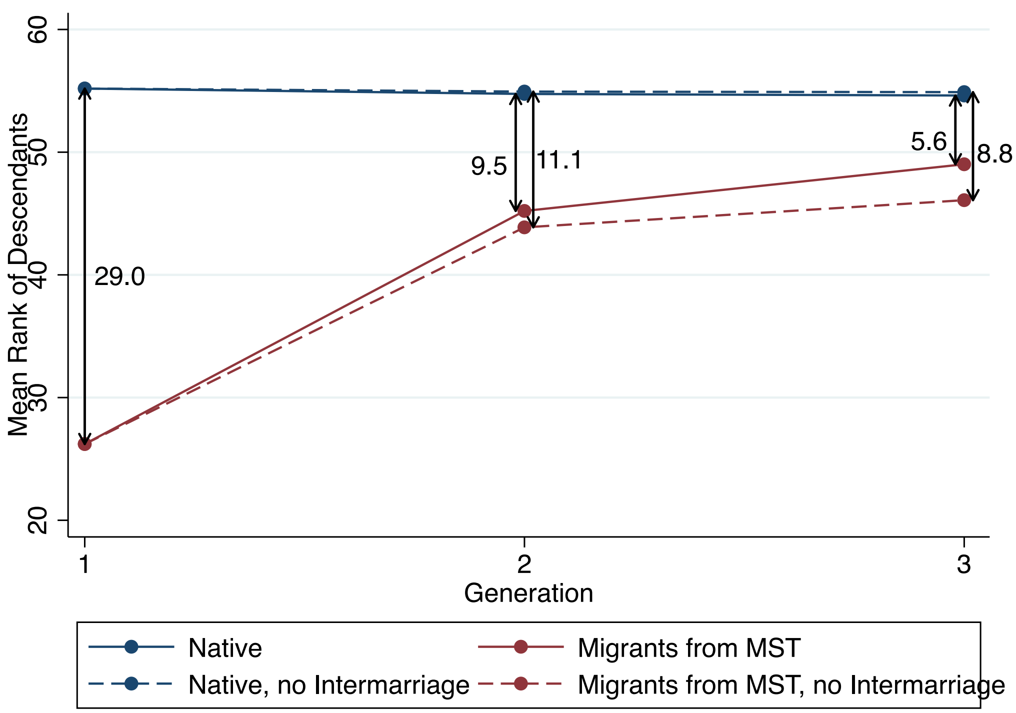
$$= \frac{(2-\gamma)(1-2\theta+\gamma\theta)Y_{2,n}+2(\theta+\gamma-\theta\gamma^2)Y_{2,mn}-\gamma(1-\theta\gamma)Y_{2,m}}{(2-\gamma)(1-\theta\gamma)} \tag{1.4}$$

For the third generation, this difference was calculated computationally. The results from the model follow from mechanical extrapolation of the current relationships of income mobility and interethnic mixing in the data. It is important to see that a model that includes heterogeneous (mixed) children, with the intermarriage rates being strictly positive, will always lead to an ethnic homogenous population. Indeed, over the long run, with random matching, there will a homogeneous population with a homogeneous (mixed) background. Therefore, the migrant or ethnic gap, as defined above, will always reduce to zero over the long run. With intermarriage rates increasing, this is an important insight for those papers who make dramatic predictions about the long-run differences between ethnic groups. Current interracial dating rates are at its highest ever in the US, and the rates in the Netherlands, shown in Figure A1.6 are also increasing. However, with the still relatively low intermarriage rates for MST migrant communities in the Netherlands, predictions for the near future will necessarily include sizeable fractions of the population a background that is of homogeneous ethnicity.

Figure 1.9 shows the expected evolution of the average income ranks of descendants of natives and of migrants from Morocco, Suriname and Turkey (MST), with generation 1 denoting the generation that migrated to the Netherlands. It also shows, across the three generations, the migrant income rank gap, with and without including intermarriage in the model. The dynamics of the model without intermarriage can be interpreted as the slow convergence to the steady state of 8.5 percentiles described earlier in Section V.C.

The largest part of the migrant income gap is closed by moving from the first generation to the second generation, as the gap decreases from 29 percentiles to 11.1 percentiles (or 9.5 in case of the model with intermarriage). Such a large decrease in the migrant gap between the first and the second generation is corroborated by the decreasing of the income rank differences between parents and children in Table 1.1. The children of the second generation close relatively less of the migrant gap, and face a gap of 8.8 (or 5.6 in case of the model with intermarriage) percentiles that is close to the steady state gap of 8.5 (or 0 in case of the model with intermarriage) percentiles.

Figure 1.9: Evolution of Migrant Income Gap



Notes: This figure shows the projected evolution of mean family income rank of descendants of different groups in the Netherlands, based on the model explained in Section V.D. The solid lines are the model where only the ethnically homogenous offspring are included, and abstraction is made of ethnically heterogeneous (mixed) offspring of migrants and natives. The dotted lines show the model's outcomes if all descendants of natives and immigrants are taken into account. The predictions in this graph are based on parameter estimates as described in this section. MST stands for individuals with a background from Morocco, Suriname and Turkey.

The extension in this model, however, shows how much the integration of migrants are understated in a model that excludes intermarriage. Indeed, including the extension that allows for mixed ethnic backgrounds reduces the income gap among migrant descendants in the second generation by 1.6 percentiles, or 14.4%, and in the third generation by 3.2 percentiles, or 36.4%. The differences between both models are large due to the fact that mixed children perform significantly better compared to ethnic minorities, as observed in the previous sections. Additionally, counting every child with a heterogeneous migrant-native background as a descendant of migrants significantly decreases the fraction of migrants' descendants with an ethnically homogeneous migrant background, and therefore also the migrant gap.

VI Conclusion

This paper explored the evolution of educational and income gaps of non-Western migrant communities in the Netherlands. While the gaps are on average smaller for second and third generation migrants than for their parents, large migrant gaps in educational outcomes and incomes are currently prevalent across all age groups in Dutch society. Empirical results in this paper suggest that these gaps will continue to shrink as integration is a long-run process, with descendants of migrants continuing to assimilate as educational outcomes increase even 50 years after the initial migration date.

However, based on historical income mobility patterns, a predicted long-run gap of 8.5 percentiles between descendants of MST migrants and natives is forecasted. Yet, it is important to realize this paper's predictions of future gaps are not set in stone and changing mobility patterns might lead to a long-run gap that is smaller (or larger) than 8.5 percentiles. Moreover, simulations suggest that merely accounting for individuals with heterogeneous native-MST backgrounds shrinks the migrant descendants' gap in the third generation by 36.4%.

This paper hints at two major avenues for follow-up research.

First, improvements in the measurement of long-run migrant assimilation can be made. This paper has looked at migrants' integration through the lens of two different methodologies that both offer only a part of the picture and are necessarily imperfect. Indeed, while the time in country approach is made complex by the appearance of new generations within migrant families, the intergenerational approach measures mobility only across two generations at one specific static time. How these two approaches formally relate to one another is a difficult research question that could prove fruitful and could allow researchers to better disentangle the mechanics of migrant assimilation.

Second, while this paper has been largely descriptive, it can inspire further work aimed at improving the educational and income outcomes for migrants' descendants. Importantly, this paper confirms the widely held view that migrant gaps are already present at early ages which suggests that interventional policies should likely primarily target migrants' descendants at early ages. Moreover, further economic mobility research for migrants might focus on individuals with mixed ethnic backgrounds, as findings in this paper point to their relatively high educational and economic outcomes.

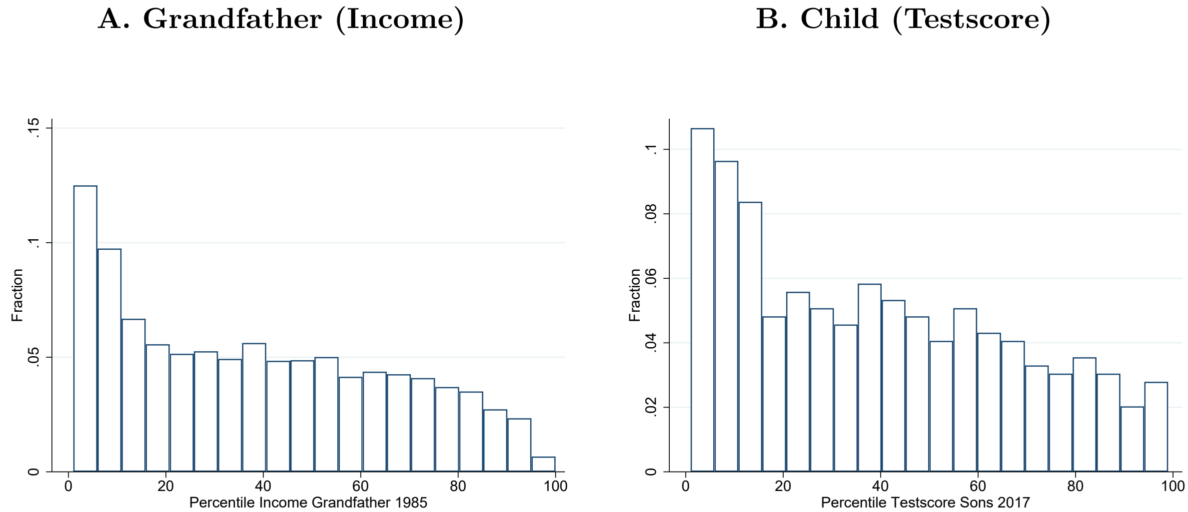
VII Appendix Figures and Tables

Table A1.1: Full Population Summary Statistics

	Ethnic Background					
	Dutch	Moroccan	Surinamese	Turkish	Other Western	Other Non-Western
Demographics						
Age	44.0	30.5	37.2	32.4	42.7	31.0
Male	50%	51%	48%	51%	48%	50%
Migration Background						
Born in Netherlands	99%	59%	53%	55%	50%	33%
0 Foreign Born Parents	98%	4%	7%	5%	8%	1%
1 Foreign Born Parents	2%	11%	18%	13%	36%	13%
2 Foreign Born Parents	0%	84%	75%	82%	56%	86%
Registered Migration Move in Family	13%	73%	58%	69%	40%	42%
Earliest Registered Migration Move in Family	1970	1974	1977	1976	1975	1993
Education and Income						
Years of Education	9.8	6.7	9.0	7.3	9.5	7.3
Family Income 2018	73,673	44,896	57,861	52,588	64,877	46,718
Pct Family Income 2018	52.7	34.2	42.6	37.8	48.4	37.2
Number of Observations	13,277,574	416,717	368,591	433,116	1,634,786	1,050,097

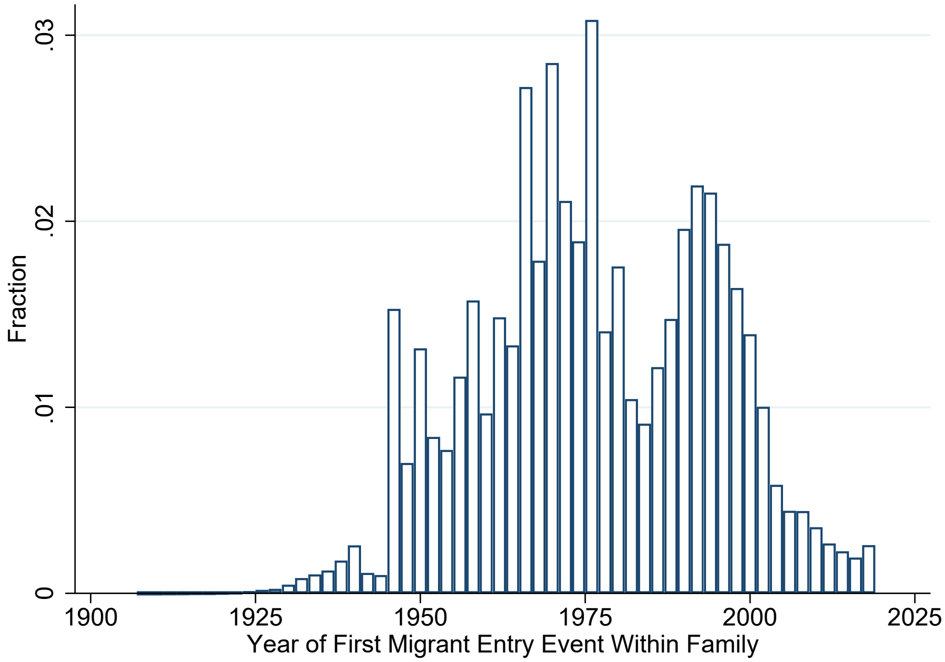
Notes: This table shows summary statistics for the entire population in the Netherlands in 2018, split out by ethnic background.

Figure A1.1: Income Distribution of a Cohort of Immigrants and their Descendants



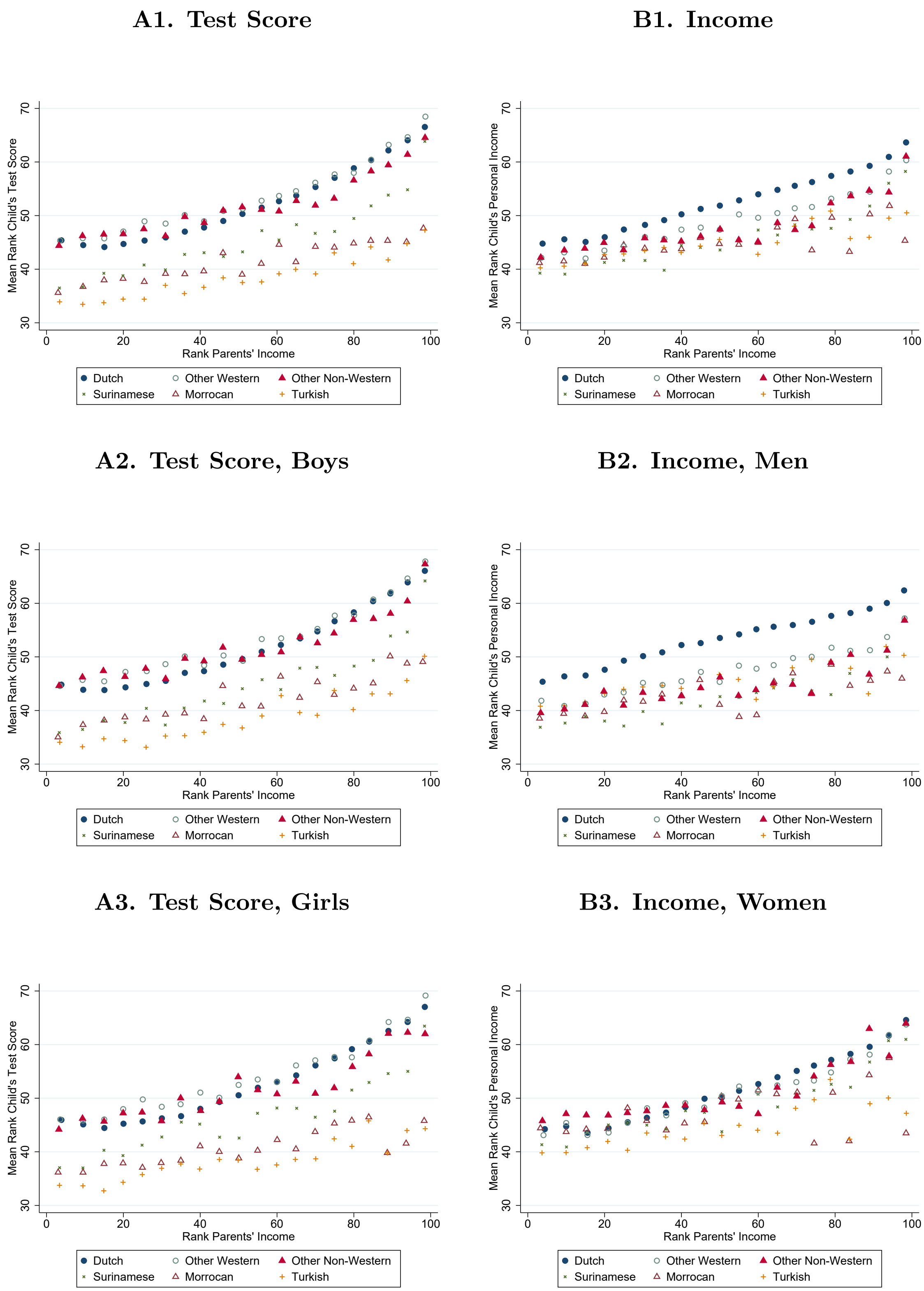
Notes: These figures show evidence of the limited mobility of descendants of migrants, as the test percentile distributions of children are similar to the income distributions of their grandfathers. Only children who took the Cito test between 2006-2018, and whose grandfathers are of age between 28 and 58 in 1985 or 1981, are selected. An additional restriction is that the entire heritage of the children is of either Moroccan, Surinamese or Turkish background, so that individuals with a heterogeneous (mixed) background are not included.

Figure A1.2: Distribution Length of Stay of Families Primary School Sample



Notes: This figure depicts the distribution of years at which the first migration movement was observed in the families of the children in the primary school sample (see Section II). This figure naturally only includes the children that are linkable to at least one migration movement in his/her family tree up to grandparent level. Therefore, it cannot be interpreted as fraction of migration movements, but rather as fraction of migration movements that led to children who took Cito tests.

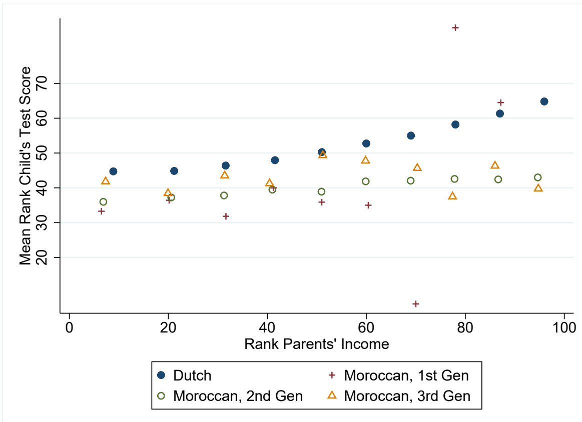
Figure A1.3: Mobility Gaps: by Country



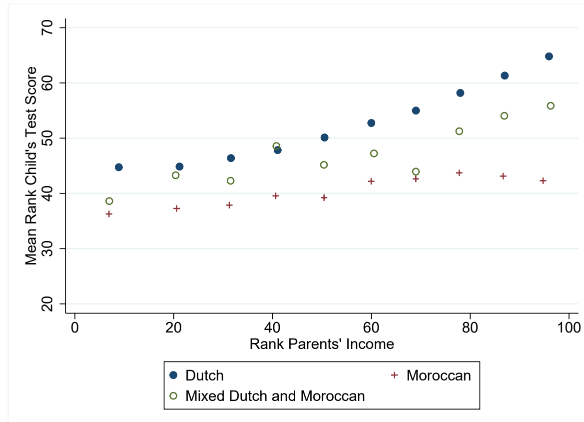
Notes: These Figures are analogous to Figure 1.6 and depict percentile average outcomes by parental income ventile, over migration background. Figures A are estimated on the test score sample and project average percentiles on the Cito test. Figures B are estimated on the income samples found in II.B and project average primary income percentiles. Data on test scores are from 2006-2018, and data on incomes are from 2018. MST stands for individuals with a background from Morocco, Suriname and Turkey.

Figure A1.4: Mobility Gaps: By Detailed Generation and Heterogeneous Background

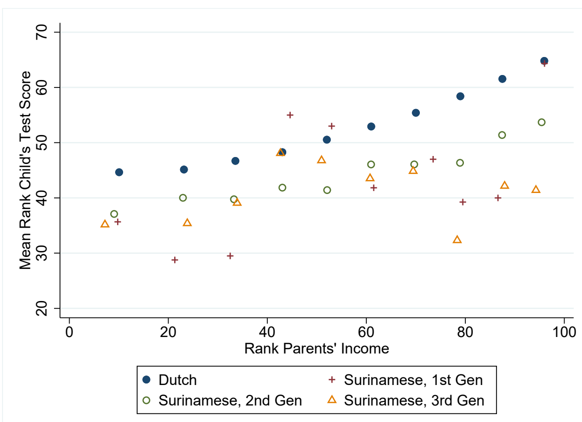
A1. Moroccan by Generation



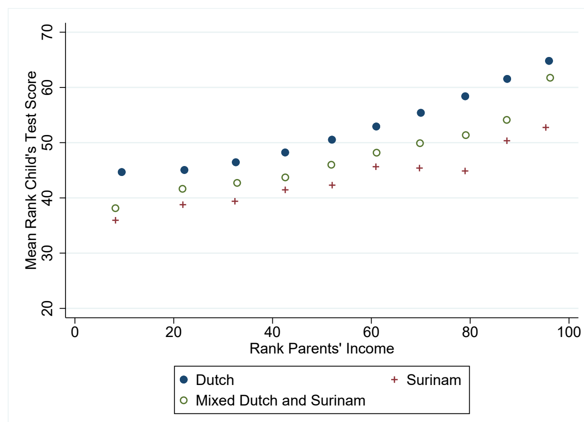
A2. Moroccan including Mixed



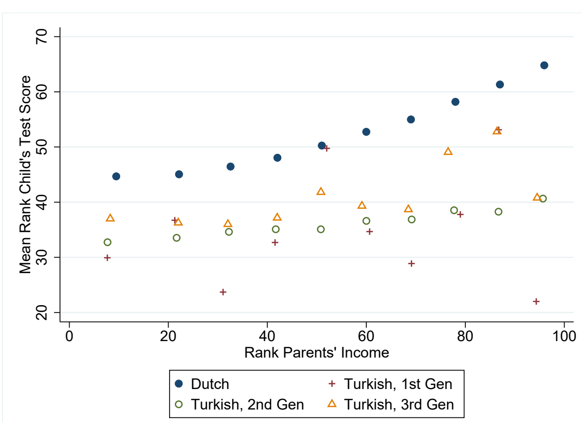
B1. Surinamese by Generation



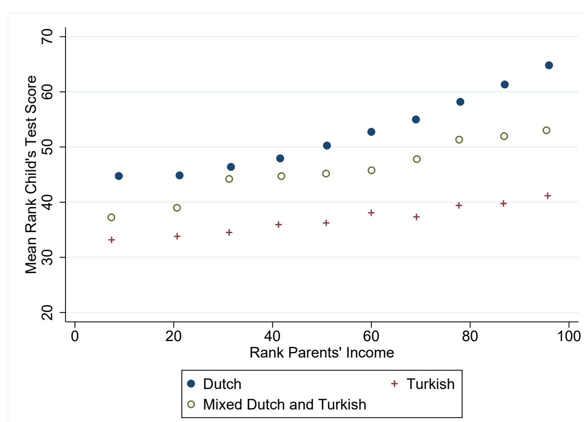
B2. Surinamese including Mixed



C1. Turkish by Generation



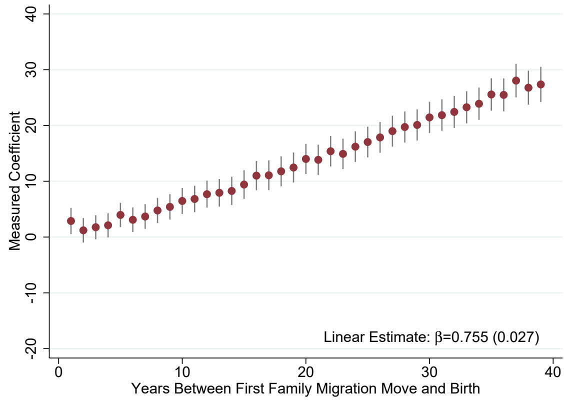
C2. Turkish including Mixed



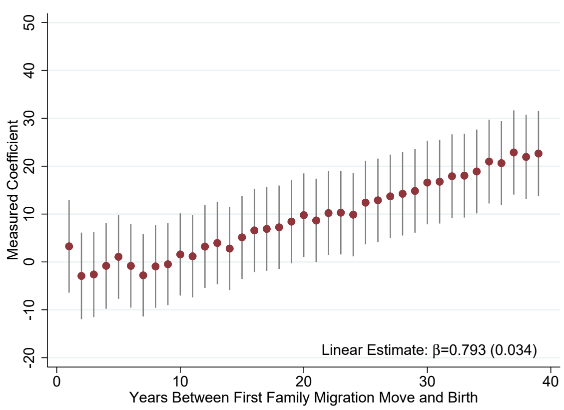
Notes: These Figures are analogous to Figure 1.7 and plot percentile Cito scores over parental income deciles. Panel A plots these mobility statistics by migrant generation. Panel B plots these mobility statistics by ethnic make-up, including individuals with a heterogeneous (mixed) background. Data on test scores are from 2006-2018, and data on incomes are from 2018.

Figure A1.5: Effect of Time in Country on Percentile Test Scores, Not Controlled for Time Trends

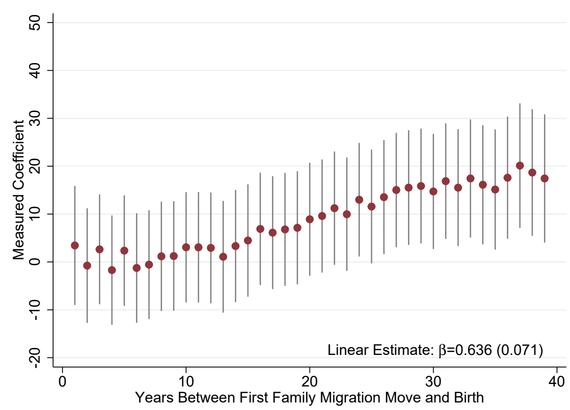
A. All Non-Western Migrants



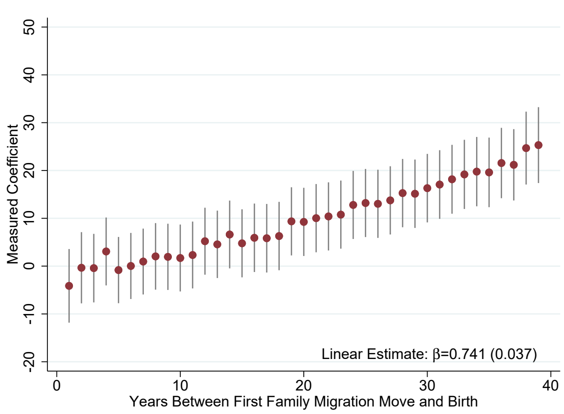
B. Morocco



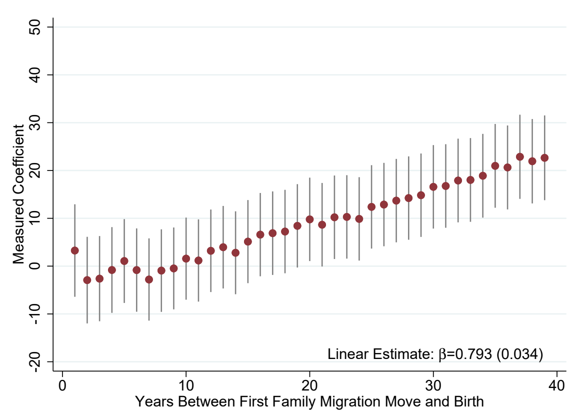
C. Suriname



D. Turkey

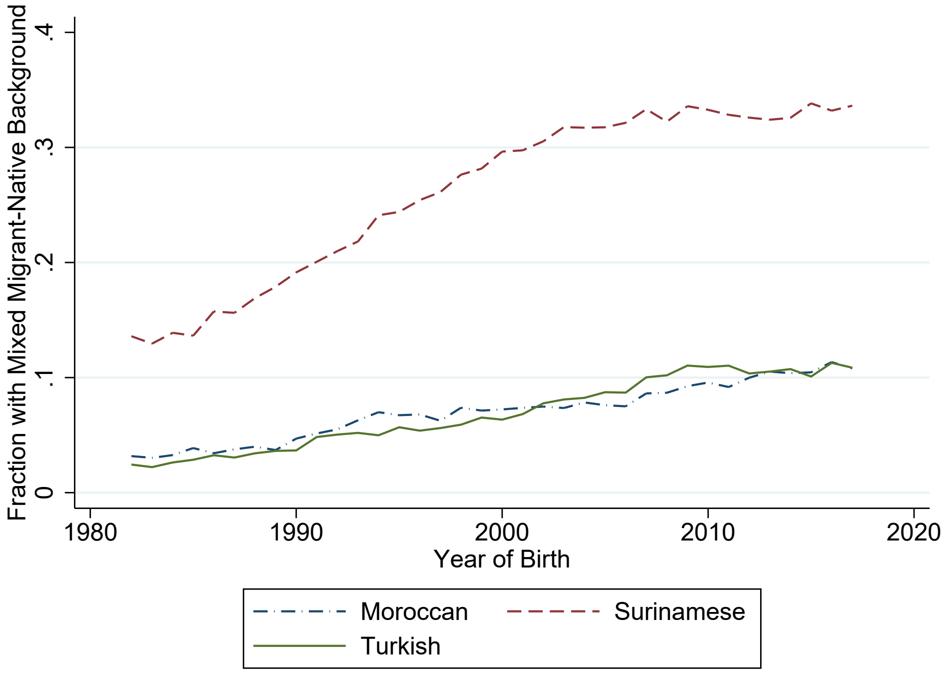


E. Other Non-Western



Notes: This Figure depicts the estimates of β_τ in a version of model 1.1 without $\overline{y}_{cgy}$ when estimating the model on Cito test score percentiles. These coefficients measure the effect on the child's test score percentile of time spent in the host country by the extended family before the birth of the child, when not controlling for time trends. Importantly, not controlling for time trends with increasing trends leads to the overestimation of the effect of time-spent-in-country on test scores. Estimates of the standard errors of point estimates for each year are shown by vertical lines. The linear estimate β included on each graph is the estimate where vector $I(\tau_i = \tau)$ is replaced by a coefficient τ.

Figure A1.6: Share of Children with a Heterogeneous (Mixed) Background



Notes: This figure plots the fraction of children with a mixed migrant-native background as a part of all children with a majority background from that ethnicity, by year of birth.

Chapter 2

The Social Determinants of Choice Quality: Evidence from Health Insurance in the Netherlands

Co-authored by: Benjamin Handel (UC Berkeley), Jonathan Kolstad (UC Berkeley), Thomas Minten (LSE) and Johannes Spinnewijn (LSE).[1]

Abstract: Policy makers increasingly offer choice or rely on markets for the provision of impure public goods like insurance, retirement savings or education. Though choice allows for improved surplus from matching individuals to appropriate products, prior work in these markets has documented choice frictions that have the potential to unwind or even reverse these benefits. We use rich administrative data on health insurance choices, health care utilization and myriad socio-demographic factors for the entire country of the Netherlands to study how insurance deductible choice quality relates to these

[1]We thank the Central Bureau of Statistics of the Netherlands and especially Annemieke Redeman for help with the data. Chloé de Meulenaer, Miguel Fajardo Steinhäuser and William Parker provided excellent research assistance. We thank Jason Abaluck, Anna Aizer, Saurabh Bhargava, Zarek Brot-Goldberg, John Campbell, Raj Chetty, Rebecca Diamond, Keith Ericson, Amy Finkelstein, Sebastian Fleitas, Jihye Jeon, Tim Layton, George Loewenstein, Neale Mahoney, Sarah Miller, Martin Salm, Josh Schwartzstein, Mark Shepard, Jon Skinner, Justin Sydnor and Richard Thaler for their discussions and comments. We thank seminar participants at Amazon, Delaware, Edinburgh University, IO^2 Stanford seminar series, Erasmus University Rotterdam, KULeuven, LSE, Monash University, UPF Barcelona, University of Virginia, University of Wisconsin, BU-Harvard-MIT health seminar, Statistics Netherlands, the 2021 AEA, 2021 ASHE, Essen Health Conference, KVS, the NBER Health Care, the NBER Public Economics, the Paris-London Public Economics, the CEPR Public Economics, the Sloan Economics of Inattention and the UPENN Behavioral Health meetings for excellent comments. We gratefully acknowledge funding by ERC (grant #716485), ESRC and STICERD.

factors. We document that choice quality is low on average but that there is a striking choice quality gradient with respect to socio-economic status. Individuals with higher education levels and more analytic degrees or professions make markedly better decisions, holding constant other key potential factors. Income, net worth, and liquidity are associated with better choices, though to a smaller degree than education. We exploit panel data on individuals' colleagues, neighbors and family members to estimate the causal impacts of peers and one's environment on choices. We find strong impacts on choice quality along each of these three dimensions and show that peer effects accelerate inequality in the sense that more positively influential peer effects are correlated with higher education and income levels. We use our estimates to model the consumer surplus effects of different counterfactual scenarios related to (i) smart defaults and (ii) menu design.

I Introduction

Consumer choice is a central aspect of market function and an important rationale for policy makers who increasingly rely on market solutions that provide choice in the provision of products viewed as public goods, such as retirement investments (see, e.g., Hastings et al. (2013) and Chetty et al. (2014*a*)), schooling (see, e.g., Neilsen (2017)), electricity (see, e.g., Ito (2015)), and health insurance (see, e.g., Enthoven, Garber and Singer (2001)). One important argument for facilitating choice in such markets — rather than a uniform product, whether offered directly by the government or a regulated private firm — is the opportunity to match heterogeneous consumers with products that provide them with greater surplus. Whether consumers are matched with the best products for them, however, hinges on their ability to effectively choose among offerings.

In practice, if consumers make choice errors, as much prior work documents, the welfare gains from greater choice and competition are diminished, or even eliminated. What has been documented less is how these barriers to effective choice vary in the population (e.g., Mullainathan and Shafir (2013), Campbell (2016)). However, to evaluate the welfare implications of choice-based policies, we are concerned not only with the average consumer-product match but with the distribution of choice quality and surplus. In particular, when consumers with lower socioeconomic status are less able to make complex decisions or have less opportunity to engage with those decisions, choice-based policies may increase inequality and be detrimental for social welfare.

In this paper we investigate consumer choice barriers and their social determinants and analyze how the inequality in choice quality affects welfare. We study this in the context of health insurance provision in the Netherlands. The Dutch setting is particularly well suited to studying choice barriers because the financial aspects of insurance contracts are relatively simple, making it more straightforward to assess choice quality. Moreover, we can leverage rich administrative data on the universe of the population of the Netherlands (approximately 12 million people) to study the social determinants of choice quality. Our data includes detailed information on demographics, health status, income, net worth, liquidity, education level, education field, profession, and social networks (work, neighborhood, family). These data are linked to individual health insurance choices in the Dutch market in which private insurers offer products under a set of regulatory constraints on product attributes. The policy approach is similar in spirit to Affordable Care Act in the United States and many other managed competition approaches implemented or discussed in other countries. Beyond the specifics of health insurance, the choice environment shares many features with market based solutions for impure public goods more generally (e.g., retirement savings). Products are offered within limits set by a regulator or market designer. Consumers are expected to make choices over a variety of dimensions of which financial outcomes are a key aspect.

The dimension of choice we focus on is the choice of deductible — the amount in each year a consumer must pay out-of-pocket before insurance payments kick in. All insurance contracts (i.e., for every plan design for every brand) have a baseline default deductible (375 EUR in 2015). Consumers can also elect to switch to five higher deductible options, in 100 EUR increments up to an additional 500 EUR (875 maximum total deductible in 2015). When consumers elect a higher incremental deductible, they get a premium rebate of about half of the incremental deductible amount.[2] Because all insurance brands offer the full range of deductibles for all products offered we can abstract away from brand or other plan characteristics that typically enter utility in choosing insurance products. Instead, we focus on a relatively simple model in which enrollees choose a deductible level based on health risk and risk preferences, though we also discuss the implications of liquidity constraints and price sensitivity in health care demand.

[2] The policy to offer a high-deductible option is a point of ongoing debate in the Netherlands: proponents argue that it allows for improved matching of consumers to deductibles and makes consumers more cognizant of health costs. Opponents have argued that it reduces pooling based on health risk, hurting sick consumers. Based on this research, the inequality in choice barriers has become part of the parliamentary debate too. See the letters addressed to the Parliament by the respective Ministers of Health (Schippers (2016), Van Ark (2021)).

To assess choice quality, we use tools from machine learning to predict health risk as a function of a rich set of variables related to prior medical utilization and consumer characteristics (see, e.g., Einav et al. (2018)). We demonstrate that approximately 60% of consumers would be better off choosing a higher deductible based on predicted health risk. In contrast, only about 10% actually do so in practice. Both when using cross-sectional and/or within-individual variation, we find significant, but small effects of predicted health risk on deductible choice. Even among those for whom we predict health spending to be almost certainly below 375 EUR, take-up of a higher deductible is only about 15% while almost all of these individuals should do so under a neoclassical framework. We show that this large gap between predicted and observed choices (i) cannot be rationalized by reasonable risk preference estimates or standard models of moral hazard and (ii) is not explained by low financial liquidity in our data (see, e.g., Ericson and Sydnor (2018) and Finkelstein, Hendren and Luttmer (2019)).[3] We also discuss the myriad potential choice barriers that can underlie this gap, including passive choice barriers related to choice defaults and active choices barriers from, e.g., limited information or comprehension.[4]

We use these risk projections together with our choice model to classify individuals into those who, in a frictionless environment, should clearly be opting for a high deductible and those who should clearly be opting for a low deductible. We then investigate what specific factors contribute to very low take-up of financially beneficial higher deductibles by predictably healthy consumers. To do this, we estimates a series of models of choice incorporating detailed observables characteristics on human capital, financial status and peer effects. Specifically, we model deductible choice — take-up of the 500 EUR deductible — as a function of health status, observable characteristics and the interaction of the two.

Our empirical analysis identifies a number of strong predictors of heterogeneous choice quality.

---

[3]Our analysis, as well as recent work by Remmerswaal, Boone and Douven (2019), shows evidence of limited moral hazard with respect to the deductible policy we investigate in the Netherlands. We also investigate behavioral hazard a la Baicker, Mullainathan and Schwartzstein (2015) and find limited economizing on potentially under-utilized high-value care (e.g. preventive care, mental health), suggesting that ex ante forecasting of ex post behavioral hazard is not a likely / reasonable explanation for the wedge between ideal individual-level allocations and actual deductible choices.

[4]While disentangling different potential choice frictions is not our goal in this paper, we discuss a collection of factors likely underlying these observed choice patterns. Given that the baseline lower deductible is typically the default option for consumers and that poor choices are predominantly by healthy people not moving away from this default option, micro-foundations underlying default effects are likely to play an important role (see, e.g., Madrian and Shea (2001), Handel (2013), and Chetty et al. (2014*a*)). Recent work by Brot-Goldberg et al. (2021) shows that, for Medicare Part D LIS beneficiaries, default effects are powerful, long-lasting, and due primarily to inattention rather than switching costs. Other potential micro-foundations we discuss include including limited information (see, e.g., Handel and Kolstad (2015*b*), Kling et al. (2012) and Bhargava, Loewenstein and Sydnor (2017)), limited attention and salience (see, e.g., Bordalo, Gennaioli and Shleifer (2012), limited comprehension (Bhargava, Loewenstein and Sydnor (2017), switching costs (Handel (2013)), and first-order risk aversion (see, e.g., Sydnor (2010)). We also discuss several neoclassical alternatives such as, e.g., correlated background risk (e.g., Campbell and Viceira (2002)).

Education level and education field are particularly important, holding constant other factors like income and financial capital. Differences by educational background are almost entirely explained by the interaction with individuals' predicted health status. When predictably healthy, individuals with an education level higher than college are 18 percentage points more likely to choose a high deductible than those with less than high school education. This estimate is 13 percentage points (5 percentage points) for those with a college degree only (high school degree only). Election of the 500 EUR deductible is similar and close to zero for consumers *who are predictably sick* in each group.

We observe approximately 80 distinct education fields. There is a strong positive relationship between being trained in an analytic field and deductible choice quality. Holding all else equal, e.g., statistics majors are 21 percentage points more likely to choose a high deductible when predictably healthy, relative to the collection of other fields. The other top fields in terms of high deductible choice are math, physics, architecture, biology, earth science, philosophy, and medicine. Conversely, all else equal, those with for example training in hair and beauty services or security training are respectively 3 and 6 percentage points less likely to choose the high deductible when predictably healthy than the general population. Similarly to education field, we show that those working in more analytic professions are, all else equal, more likely to choose the high deductible when predictably healthy.

Once we control for education and job type, we find a more modest role for income and financial capital. All else equal, when predictably healthy, someone in the top income quartile is 4 percentage points more likely to choose a high deductible than someone in the bottom income quartile. There are minimal differences across the lowest three income quartiles. For net worth, all else equal, someone in the top quartile (third quartile) is 6 (2) percentage points more likely to choose a high deductible when predictably healthy than some in the bottom quartile. We also investigate, and find minimal effects of, both mortgage debt and general debt. Importantly, we also study the impact of having liquid savings (more than 2000 EUR) on choice. Holding all else equal, when predictably healthy, someone with more liquid savings is 2 percentage points more likely to choose a high deductible than someone with low liquid savings. While these effects are small in magnitude relative to the effects of education (both level and field) they do suggest some impact of these constraints on choice (whether motivated by neoclassical considerations or correlated choice frictions).

Factors like financial capital and, especially, human capital, are determined over a long time horizon and can be the result of a large range of underlying factors. As a consequence, our estimates relating human and financial capital to choices are necessarily non-causal and simply reflect the association between these factors and choices. One potentially important factor with shorter-run variation are social or information networks. The role of peer effects in insurance choice have been studied in specific settings (e.g., Sorenson (2006) studying University of California employees) but not at scale while being able to control for a range of key underlying socio-demographic factors. In a broader population peer effects may affect the average choice quality but could also be an important contributor to inequality. For example, if local peers impact choices, we might expect heterogeneity by geography in choice quality (e.g., urban versus rural). If work peers have large effects on choices, this could further exacerbate the differences in choice quality by job type and firm. Finally, effects within families would be suggestive of important inter-generational transfers of choice capital, either good or bad.

To study peer effects, we leverage the detail of the data that allows us to identify workplace colleagues, neighbors, and family members and their choices. As has been well documented, estimating peer effects poses important empirical challenges, such as, e.g., separately identifying peer effects from correlated unobservable heterogeneity in a peer group (see, e.g., Manski (1999)). We address these challenges using a switcher-design similar in spirit to that described in Abowd, Kramarz and Margolis (1999) to identify the effects of workplace and geographic peers. We estimate a first-stage panel regression with individual fixed effects and firm (location) fixed effects, controlling for predictable differences in health. This framework leverages switchers moving across firms (location) to identify firm (location) fixed effects on deductible choices. In a second-stage, we project these fixed effects onto take-up of high deductible plans within the firm (or location), to estimate the extent to which the fixed effects explain differences in take-up across.

Our results show that within-firm peers have a substantial impact on individual decisions. A 10% increase in the number of co-workers taking a high-deductible causes a 1.4% increases in high-deductible take-up for people switching into the firm. The estimates are strongest for individuals who are predicted to have low health costs — those who benefits from electing a higher deductible. For example, a 10% increase in the number of peers taking up a high-deductible causes a 1.7 % increase in high-deductible take-up for healthy people switching into the firm but a decrease of .5 % for those who are predictably sick. We find similar results for neighborhood peer effects.

We investigate the implications of these peer effects for inequality. Ordering firm peer fixed effects, we find that for the bottom five deciles approximately 20% of employees are college educated. Moving from the sixth to tenth deciles, this proportion increases monotonically from 25% to 40%. Cut a different way, for firms with with a low proportion of college educated ($< 20\%$) the 75th percentile firm peer effect is a one percentage point increase in high deductible choice while for firms with a high proportion of college educated ($> 90\%$) this same statistic is four percentage points. This three percentage point gap is meaningful relative to the overall share of consumers choosing high deductibles.

To investigate this further, we perform a counterfactual analysis that sets all peer effects for colleagues and neighbors to the average of the top decile, equating these effects for all individuals. This leads to a 35% (2.2 pp) increase in high deductible choice for the predictably healthy with less than a high school education and a lower 9% (1.8pp) effect for the predictably healthy with an advanced degree. Taken together, these results show the firm and neighborhood peer effects impact choices meaningfully and also exacerbate inequality.

Differences in choice quality may also arise or persist through inter-generational transfers of human capital. We study the impact of family members on each other's choices leveraging an event-study design. We find that when parents switch their choices, children under 30 living apart from the parent have a 25 percent chance of following their parents and switching. Children over 30 follow their parents' switches, but to a lesser degree, only increasing incremental deductible take-up by 10 percentage points after a parent switch to that deductible. Interestingly, this effect is driven by the take-up response of children who are predictably healthy, but does not differ with their parents' health. This suggests that the primary driver is learning about the parents' decision and considering it in the context of their own health but this does not depend on whether the parent made an effective choice to begin with.

Overall, our results paint a detailed picture of the role of socio-demographic factors in deductible choice and show that a range of factors that are outside of the standard model of insurance choice not only are present, but have large impacts on choices. The final part of the paper turns to welfare and evaluates the desirability of choice-based policies.[5] We combine the disparate factors affecting choices into a measure of choice quality and quantify overall inequality in choices and outcomes. To do so, we use our regression estimates to predict consumer choices as a function of health status. Then, using our model of consumer surplus, we rank consumers in terms of choice quality, conditional on health. The top 5% of decision makers choose the surplus maximizing deductible only 55% of the time, while the

[5]We note that the model does not reflect the potential consumer surplus impacts of additional cost-sharing on health care utilization (see, e.g., Brot-Goldberg et al. (2017)). As mentioned, we find evidence of limited moral hazard with respect to the deductible policy.

remaining 95% of the population make choices that are worse than choosing at random. We assess the underlying correlates of consumers being better (top 5%) or worse (bottom 5%) in terms of the value they extract from deductible choice. For example, the 5% best decision-makers have an average gross income of 105,000 EUR and net worth of about 250,000 EUR, relative to about an income of 40,000 EUR and net worth of 5,000 EUR for the 5% worst decision makers.

A variety of policy options might be employed to address choice quality, and inequality in choices, in the context of the Dutch health insurance market. To shed light on those we use our model estimates to study several counterfactual policies. First, we consider the consumer surplus gains from an optimal allocation of consumers to deductibles. This scenario offers a useful first-best benchmark, but also relates to a plausible policy intervention if regulators were to use a smart default approach. That is, our counterfactual captures what would happen if our cost prediction model were used to default people into plans and they took that advice. The other two counterfactual policies reflect policies in which the choice set is limited to either only the higher deductible option (875 EUR) or only the lower deductible option (375 EUR), essentially eliminating choice.

The average benefit from a smart default policy is an improvement in welfare per enrollee of between 58 and 69 EUR, where the lowest value in the range reflects a high assumed CARA coefficient of 10^{-3} and the largest value reflects risk neutrality. These are small in absolute terms, but high relative to the average money at stake of about 145 EUR. Eliminating choice reduces average welfare, but the impact is smaller. The offered option to take a high deductible increases consumer welfare only by 7 EUR to 8 EUR per person. Only offering the 875 EUR deductible would decrease consumer welfare by 26 to 45 EUR.

These results, like much of the prior literature, ignore the role of inequality in outcomes. To incorporate this, we weight outcomes as a function of income using parameters from the inequality literature (see, e.g., Atkinson (1970)). The value of offering the option to take a high deductible further decreases when using income-dependent welfare weights, since individuals with lower income make worse decisions and have worse health on average. This negative correlation between income and health also reduces the appeal of mandating all individuals in the high-deductible option: with high inequality aversion the social surplus loss from only offering the high-deductible is between 134 and 149 EUR, much larger than the analysis that does not factor in inequality aversion.

Overall, the counterfactual analysis shows that the existence and magnitude of choice frictions dramatically reduce the value of offering the high-deductible option, especially for the low-income individuals, who are both less healthy and make worse choices. Instead, consumers would be much better off if their choices were better directed (e.g., through smart defaults a la Handel and Kolstad (2015*a*) or Abaluck and Adams (2019)). Even with such a directed policy, the high-income consumers have the most to gain, despite their higher quality decisions, because they are healthier on average.

Related Literature This paper relates to several distinct literatures, but is closest to prior work on insurance choice including papers without choice frictions (e.g., Cohen and Einav (2007), Bundorf, Levin and Mahoney (2012), Cardon and Hendel (2001), Einav, Finkelstein and Schrimpf (2010), Einav et al. (2013)) and many with choice frictions.[6] This collection of prior papers make contributions on many dimensions including (i) documenting key micro-foundations underlying choices and (ii) documenting the value (or in some cases lack thereof) that consumers extract from choice in insurance markets.

Relative to this prior work, the choice we study is simpler and the data we have are much deeper and more comprehensive in terms of socio-demographic factors, allowing us to contribute in several key ways. First, we are able to study choice heterogeneity on many potentially important dimensions simultaneously for the same population. Most prior work studies heterogeneity as a function of age, gender, and, in some cases, income. For example, Bhargava, Loewenstein and Sydnor (2017) study how income of employees at a large firm relates to dominated plan choice, finding that lower income is associated with poorer choices. Relative to that paper, which is near the forefront of the literature, we are also able to study education level and type, financial capital, and peer effects for a large representative sample. One key implication, e.g., is that education level and type are more predictive of choice quality than income. Perhaps most notably, Fang, Keane and Silverman (2008) use MCBS and HRS survey data

[6]See, e.g., Sydnor (2010), Abaluck and Gruber (2011), Ketcham et al. (2012), Barseghyan et al. (2013), Ericson (2014), Handel and Kolstad (2015*b*), Ketcham, Lucarelli and Powers (2015), Polyakova (2016), Abaluck and Gruber (2016*a*), Handel, Kolstad and Spinnewijn (2019), Abaluck and Gruber (2016*b*), Ho, Hogan and Scott Morton (2017) Ketcham, Kuminoff and Powers (2019), Abaluck and Adams (2019), Brot-Goldberg et al. (2021)). The literature on insurance choice with choice frictions / behavioral choice foundations is summarized in the chapter by Handel and Schwartzstein (2019). Our paper also relates to two papers on the voluntary deductible in the Netherlands specifically. Van Winssen, Van Kleef and Van de Ven (2015) show that the overall voluntary deductible take-up in the Netherlands is low and a large share of individuals would have gained by taking a higher deductible. They use data from a single insurer to find that a voluntary deductible is most profitable for mostly young, male, and healthy individuals, but do not study individual choice quality. Van Winssen, Van Kleef and Van de Ven (2016) discuss several potential reasons from the behavioral economics literature to explain the low overall take-up. While these papers have documented the sub-optimally low take-up of the voluntary deductible, our paper links individual choices of the entire Dutch population with granular data on socio-economic and educational registries and employer-employee links to provide an in-depth analysis of determinants of choice quality.

to study choice of Medigap supplemental coverage as a function of surveyed education level, income, wealth, risk aversion, financial planning, and questions that relate to cognitive ability. They find, e.g., that controlling for measures of cognitive ability is important for explaining advantageous selection into Medigap. Relative to our work, they study a smaller surveyed sample of approximately 10,000 seniors with the need to impute health information across surveys and without variables related to education type, peer groups, and several other factors we study. Second, relative to the prior work above with richer dimensions of heterogeneity, we have a more representative dataset for a society, which spans the range of the support for the socio-demographic factors we study. Studies of choice in Medicare Part D noted above typically have the largest / most representative samples, of seniors, but those are also the studies that have more limited measures of socio-demographic heterogeneity. Conversely, studies with richer heterogeneity such as Bhargava, Loewenstein and Sydnor (2017) and Fang, Keane and Silverman (2008) have more limited samples, coming from one employer or from a smaller surveyed sample.

The prior work on peer effects in health insurance choice is thin with the most notable prior work by Sorenson (2006), who studies peer effects in the insurance choices of employees at a large firm.[7] Outside of health insurance, where there are very limited papers studying peer effects at scale, there are some notable papers with very strong identification of peer effects and their underlying mechanisms for smaller samples including, e.g., for (i) mortgage refinancing by teachers (Maturana and Nickerson (2018)) (ii) firm performance under executives (Shue (2013)) (iii) housing purchases (Bailey et al. (2018)) and (iv) education (Epple and Romano (2011)). Our analysis studies peer effects on multiple dimensions at scale for an entire country and links those peer effects to measures of choice quality and inequality.

Our analysis also relates to papers that study choice quality and the incidence of consumer frictions in other domains (e.g., Allcott, Lockwood and Taubinsky (2019)). Most notably, there a number of papers that study choice quality and default effects in retirement savings. Chetty et al. (2014*a*) study retirement savings in Denmark using granular nationwide data and show that default effects are much more powerful than subsidies in how they impact consumers' savings portfolios. Active choosers are more likely to be wealthy, college educated, and have an economics- or finance-oriented degree. Our study differs in several key ways including (i) our analysis of several dimensions of peer effects and (ii) the

[7] There are also quite a few papers studying peer effects in health behaviors more broadly (see, e.g., Fadlon and Nielsen (2019), Chen, Persson and Polyakova (2019))

fact that we study a context where we can, to a large extent, determine whether consumers are allocated to a good vs. bad option for themselves. Our paper makes similar contributions relative to Andersen et al. (2020) study mortgage refinancing decisions with granular data on demographics, income, and education from Denmark. They find, somewhat in contrast to our results, that wealthy consumers act as if they have relatively high psychological costs of refinancing though, consistent with our results, they find that poorer and less educated households refinance with lower probabilities regardless of the underlying incentives.[8]

The rest of the paper proceeds as follows. Section II describes health insurance in the Netherlands and describes our data. Section III presents our choice framework and consumer cost risk prediction model. Section IV presents our empirical analysis of deductible choice and its social determinants. Section V quantifies the resulting heterogeneity in choice quality and presents our analysis of counterfactual policies. Section VI concludes.

II Institutional Context and Data

We exploit a unique consumer choice setting in the health insurance market in the Netherlands and link data on health insurance choices to data from various administrative registers. We present the institutional context and data here.[9]

II.A Health Insurance in the Netherlands

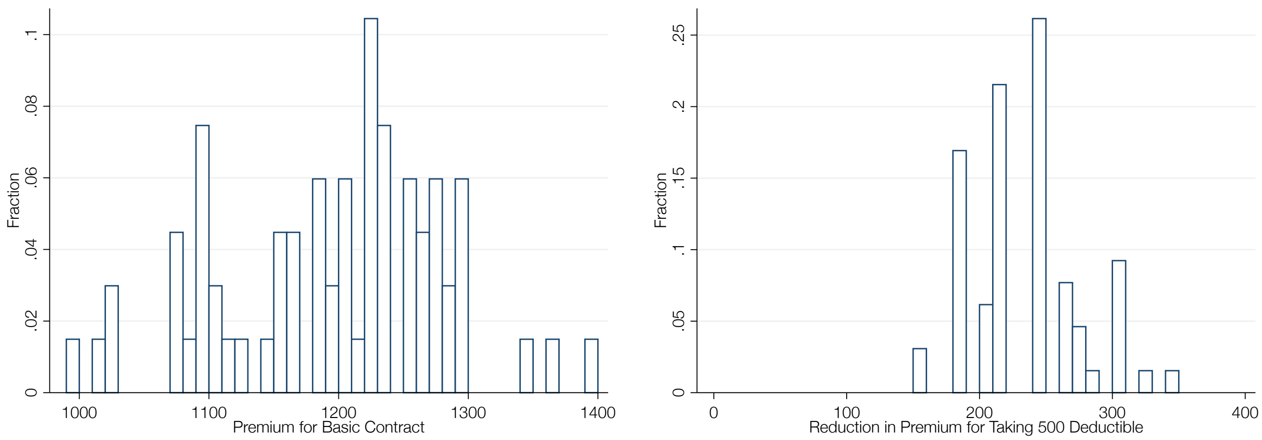
All individuals in the Netherlands are obligated to directly buy health insurance from a private health insurance market.[10] The Health Insurance Act of 2006 introduced a managed competition model in which the government strictly regulates the contents of the basic package of health insurance. The regulation also (i) prohibits price discrimination, (ii) prohibits the rejection of individuals from purchasing

[8] Cronqvist and Thaler (2004) discuss the privatization of social security in Sweden with a focus on how subtle design factors, such as default effects, can have important implications for consumer choices. Beshears et al. (2016) show that, in a retirement setting, consumers are more likely to switch away from the default option if they will benefit more from doing soon, which is similar to what we find in our environment. Madrian and Shea (2001), Carroll et al. (2009), and Beshears et al. (2008) are additional examples of papers of default effects for consumers when engaging with financial products, a literature that is nicely summarized in Beshears et al. (2018).

[9] A more comprehensive overview of the health system and changes to the health insurance model in the Netherlands can be found in Kroneman et al. (2016).

[10] Every adult individual is required to choose a plan. Legally, individuals can only make the insurance policy choice for other adults if they have been provided with written consent from that adult. Children aged below 18 years are also required to have an individual plan. In practice, most parents register their children with the same insurer.

Figure 2.1: Distribution of Premia and Premium Reductions



Notes: Histograms of yearly premiums in 2015 for basic coverage (left-hand side) and premium reductions for those contracts when electing a maximal voluntary deductible of 500 for a total deductible of 875 EUR (right-hand side). Data on prices are obtained from homefinance.nl.

insurance and (iii) mandates that all individuals purchase coverage.[11] Insurers compete for consumers on premiums, provider choice, and supplementary insurance.[12] In 2015, there were 25 health insurers that together offered 53 separate insurance contracts. As shown in the left panel of Figure 2.1, yearly premiums for the mandatory health insurance with the smallest possible deductible have a mean of 1195 EUR and a fairly compact distribution around this mean.

Consumers enroll between mid November and the end of December for the following year.[13] During that period, health insurers advertise their insurance packages through various media. If no action is taken by the consumer, she will automatically extend her current contract. Relatively few consumers switch insurers each year (only 6.8% of individuals in 2015).

Each individual faces a compulsory deductible (375 EUR in 2015), but can opt for an extra voluntary deductible of 100, 200, 300, 400 or 500 EUR on top of this compulsory deductible (maximum total

[11]To limit incentives for selection of consumers based on their health, the government has installed a sophisticated risk adjustment system. Yet, van Kleef, Eijkenaar and van Vliet (2019) show it is still profitable for insurers to attract healthy consumers.

[12]The basic package covers drugs, doctor and hospital expenditures. Supplementary insurance covers dental care, additional physical therapy, alternative medicine, and other care. In 2015, approximately 90% of insurees bought supplementary insurance. The average premium for the supplementary insurance averaged 233 EUR in 2015.

[13]Generally, there are no brokers involved in the choice of health insurance in the Netherlands.

deductible of 875 EUR in 2015).[14,15] The compulsory deductible, introduced in 2008, has gradually increased from 150 EUR in 2008 to 385 EUR in 2017[16], while the options for the extra voluntary deductible have remained the same. By opting for a higher deductible, consumers receive a premium reduction. Figure 2.1 shows the (unweighted) histogram of premium reductions consumers can get by electing the additional 500 EUR deductible across health plans offered in 2015. The distribution has a mean of 233 EUR and most of the mass lies between 200 and 300 EUR, making the deductible election a quite standardized decision across all insurance contracts.

Insurers can make agreements with employers, municipalities and various associations to offer group plans. These group plans are selected packages of basic and supplemental insurance on which the insurers offer premium reductions (*collectiviteitskorting*) of up to 10%. This feature in the insurance market leaves the choice of voluntary deductible unaltered for a given insurance contract. An exception to this are collective agreements between some municipalities and insurers for low-income individuals (*gemeentepolissen*), with income thresholds below 130% of the minimum wage. These policies are subsidized by municipalities, sometimes by covering the mandatory deductible amount, in which case they would not involve a deductible choice.[17]

The design of the compulsory deductible combined with a voluntary deductible has been a central topic of the policy debate. The desirability of consumer deductible choice has repeatedly been discussed in the Dutch parliament. In 2016, the Minister of Health Affairs, Schippers (2016) argued that having the option of a voluntary deductible increases general support for the health care system by the healthy, and makes individuals more aware of their health costs. Similar arguments have been put forward in recent exchanges in the Parliament in 2018 and 2019.

---

[14]The government does not mandate insurers to provide the choice of voluntary deductible. However, in practice, almost all insurers provide the option take the voluntary deductible. In 2015, of the 68 insurance contracts that have price information available on `https://www.homefinance.nl`, only one insurer (the new entrant ANNO12) does not provide a voluntary deductible option.

[15]Preventive, maternal and GP care is covered at zero cost by all insurers by law, and the deductible does not apply to the corresponding expenses. We exclude these preventive expenses from our cost prediction model.

[16]The size of the compulsory deductible was 350, 360 and 375 EUR in 2013, 2014 and 2015 respectively, then 385 EUR from 2016 onwards.

[17]In 2015, just over half a million individuals, about 3% of the population, were covered by this type of contract. As these contracts are mostly tied to generous supplemental coverage, the premium remains high relative to basic plans with high-deductible option, which is still the better option for predictably health individuals (see Douven et al. (2019)).

Table 2.1: Distribution of Deductible Choices

Default Deductible	90.94%
Extra Deductible (+100 to +500EUR)	9.06%
Breakdown of Extra Deductible Choices	
+100EUR	10.64%
+200EUR	10.41%
+300EUR	6.02%
+400EUR	1.72%
+500EUR	71.21%

Notes: This table shows the breakdown of deductible choices in 2015. A large majority (90.94%) sticks to the default 375 EUR deductible. Of the 9.06% individuals that take an extra deductible, most individuals take the 500 EUR extra deductible.

II.B Data and Sample

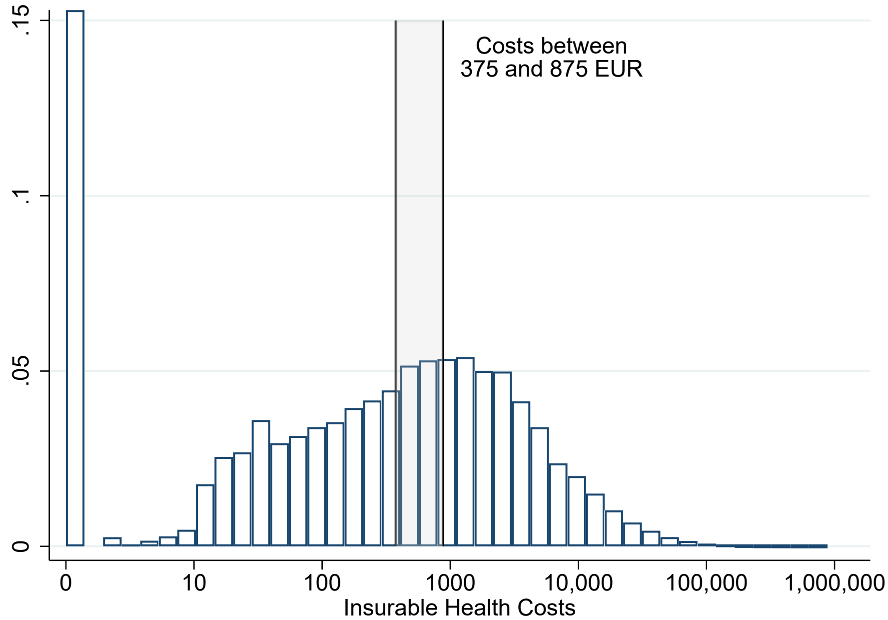
We use data on health insurance choices and health expenditures for all individuals in the Netherlands. The data is linked at Statistics Netherlands to other administrative registers, which provide information on their income, wealth, education, employment and other demographic variables.

We restrict attention to all individuals who are at least 18 years old in January of the year in which they decide on their health insurance contract and deductible. We exclude from the sample adults who have incomplete or unreliable health data records in the two previous years.[18] The remaining sample consists of about 13.25 million adults in each year. As explained in Section III.B, we use a random sample of 1.25 million of these individuals to estimate and calibrate a cost prediction model, leaving approximately 12 million adults each year for the analyses, which we call our baseline sample.

Health Insurance Deductible Data on health insurance contract choices in the years between 2013 and 2017 are obtained from Vektis, an organization that is responsible for the collection of data from all health insurers. Our data include only information on an insurer and deductible choice. We do not observe the choice of provider network nor whether individual takes supplementary insurance, but these choice dimensions are orthogonal to the deductible choice except for minor price differences. Table 2.1 shows the take-up of different deductible amounts in 2015. The voluntary deductible take-up in our sample is 9.06% in 2015. More than 2 out 3 individuals opting for an extra deductible take the maximum extra deductible of 500 EUR.

[18] Insurers in the Netherlands are split in two categories: insurers who actually bear the risk and proxy insurers who only act as middleman. Vektis, the data provider, deems data from the about 10 proxy insurers over our sample period, covering approximately 4% of people ($\approx$ 500,000), to be unreliable. Hence, we do not use these observations in our analysis.

Figure 2.2: Distribution of Insurable Health Care Costs



Notes: This figure shows the distribution of the $\log_{10}$ of total yearly insurable health care costs in 2015, for all individuals in our baseline sample. 13.1% of individuals have health costs falling in the 375 to 875 EUR interval.

Health Care Costs Data on health care costs contain annual health care expenditures by category. The categories are medicines, hospital care, geriatric care, paramedical care and physiotherapy, mental health care, aids and tools for health, health care in foreign countries, health care transport, multidisciplinary care, sensory handicap care, and other care. In addition to these categories which are subject to the deductible, we also have data on neonatal and maternal care, care by GPs and home care, where cost sharing does not apply.[19]

Figure 2.2 presents the distribution of the (log) aggregate health care expenditures that are subject to cost-sharing in 2015. This aggregate distribution is skewed with about 19 percent of individuals making zero expenditures and more than 10 percent of individuals spending more than 5000 EUR. Table 2.2 presents the distributions of annual expenditures for the different categories of medical spending. These distributions are similarly skewed. Hospital expenditures (1,388 EUR), drugs expenditures (320 EUR) and mental health care (243 EUR) are the three categories with the highest mean spending.

[19] Some miscellaneous items are also exempt from cost sharing. These include preventative care such as breast cancer screening and flu shots, as well as costs made for organ donation. We cannot separately identify these costs from hospital care, so our measured insurable costs will be slightly overestimated.

Table 2.2: Distribution of Annual Health Care Costs

	Mean	p10	p50	p90	p99
All Care	2,695	86	495	6,032	35,974
Insurable Care	2,272	0	332	5,043	31,133
Hospital Care	1,388	0	85	2,829	21,575
Medicines	320	0	53	758	3,253
Mental Care	243	0	0	0	4,801
Tools and Medical Aid	107	0	0	145	2,284
Geriatric Care	53	0	0	0	0
Transport	45	0	0	0	1,081
Multidisciplinary Care	33	0	0	124	397
Physiotherapic Care	32	0	0	0	1,095
Dental Care	26	0	0	0	825
Other Care	7	0	0	0	151
Sensory Handicap Care	3	0	0	0	0
Always Insured Care	423	75	121	327	8,042
Nursing Care	228	0	0	0	7,587
GP Care	157	75	119	272	659
Maternal Care	37	0	0	0	1,796
Observations					11,991,629

Notes: This table shows the distribution of health expenditures by subcategory, for the full sample in 2015. Expenditures are divided into insurable expenditures, that are subject to cost sharing (and to which the deductible applies) versus always insured expenditures, that are not subject to cost sharing. All values are in EUR.

Table 2.3: Summary Statistics

	Mean		Mean
Demographics		**Household Financial Status**	
Male	48.8%	Gross Household Income	73,289
Age	50.3	*10th Percentile*	*20,077*
Has Children	69.2%	*Median*	*60,358*
Has a Partner	62.9%	*90th Percentile*	*135,981*
Education Level		Household Net Worth	166,890
Less than High School	13.2%	*10th Percentile*	*-28,918*
High School	24.1%	*Median*	*32,694*
College	16.8%	*90th Percentile*	*403,923*
Further Studies	0.6%	Mortgage Debt	54.1%
Unknown	45.4%	Other Debt	34.2%
Employment Status		Savings > 2000 EUR	80.4%
Employee	44.3%		
Self-Employed	9.9%		
Retired	24.2%		
Student	6.3%		
Other Not Working	15.3%		
Observations			11,991,628

Notes: This table shows summary statistics for the full sample in 2015.

Other Data We obtain information on other variables from a number of administrative registers and link these to the health and insurance data. Our data includes standard demographics like age, gender and household status. We use third-party reported information from tax registers on household income and household wealth. The former includes pre-tax income from labor, self-employment and capital and government transfers. The latter includes information on net worth, liquid and other financial assets, mortgage and other debt. We also observe data on the highest formal education level attained for more than half of the sample. These data also include information on the specific field of study for individuals who proceed past high school. Finally, we use employer-employee data to link individuals at the firm level and identify their sector of employment. We provide more detail about the different registers and variables in the Data Appendix VI. Table 2.3 provides some summary statistics for the year 2015.

III Deductible Choice and Health Risk

In this section we study the relationship between deductible choice and predicted risk, both theoretical and empirically. We first develop a stylized model of choice. We then predict individuals' health care cost, the central input into a frictionless, rational model of choice. We finally document large discrepancies between the model's predictions and the observed choices. This motivates our empirical analysis in the next section, relating the discrepancies to social factors that underlie the barriers to choice.

III.A Deductible Choice in a Model without Frictions

Each individual is subject to a compulsory deductible of 375 and can choose a voluntary deductible d at corresponding premium p from menu $\Omega = \{(d, p_d)\}$. An individual draws health cost x from an individual-specific distribution $F_i(x)$. Depending on her deductible choice d, health cost translates into an out-of-pocket expense $s = min\{d, x\}$. We denote by $G_{i,d}(s)$ the distribution of out-of-pocket spending, derived from $F_i(x)$ and the deductible choice d. Expected utility for a rational individual in a frictionless environment is defined, therefore, as:

$$U_{i,d} = \int u_i(W_i - p_d - s)G_{i,d}(s)ds. \tag{2.1}$$

Using this definition of expected utility, we can define an individual's certainty equivalent from choosing one contract as $CE_{i,d}$, where $U_{i,d} = u_i(W_i - CE_{i,d})$.

A central decision variable when considering to elect a deductible higher than the compulsory level of 375 EUR is the chance that expenditures stay below 375 EUR, which we denote by π_i. We simplify the decision to a binary choice between the baseline deductible of 375 EUR and adopting the full 875 EUR deductible while gaining the associated premium savings. In theory, the optimal decision depends on the probability distribution of expenditures between 375 EUR and 875 EUR too, but the share of expenditures that fall in this range is small. Empirically, most individuals who elect a deductible higher than the compulsory deductible choose the maximum possible deductible. As we discuss below, interior choices between the two levels are not easily rationalized under standard preferences.

Under this simplified environment, we approximate expected utility by:

$$U_{i,d} \approx \pi_i u_i(W_i - p_d) + (1 - \pi_i)\, u_i(W_i - p_d - d), \tag{2.2}$$

and the contract space, including the following two contracts:

$$\Omega = \{(0,0)\,, (500, -250)\}\,.$$

This setup demonstrates the relative simplicity of the environment we study. In expected payoff terms, $\bar{\pi} = 0.5$ is the (approximate) threshold between optimally choosing the additional 500 EUR deductible and saving 250 EUR in premium.[20]

There are a couple of different ways that frictionless preferences choices could differ from those in the simple model specified here. While these differences do not impact our positive empirical results in Section IV they could impact our normative discussion in Section V so they are important to consider.

First, consumers could have classical risk aversion that pushes them towards choosing the low deductible option. For a standard but lower value of absolute risk aversion of 10^{-5} (e.g., Cohen and Einav (2007)), this threshold increases very slightly to 0.5006.[21] For a very high level of absolute risk aversion of 10^{-3}, this threshold is still only 0.56 (see discussion in Barseghyan et al. (2018) for typical risk preference estimates in different contexts). A model with constant relative risk aversion parameters typical of past work yields similarly small threshold changes.[22] After we discuss our cost model predictions, we show in Figure 2.4 that variation in the choice threshold as a result of risk aversion is small relative to the dispersion in predicted cost distributions.

Second, consumers could have liquidity constraints that lead them to act in a risk averse manner

[20]The threshold of 0.5 is exact if consumers always spend more than 875 EUR if they pass the baseline deductible level of 375 EUR. This conforms well to a binary model where someone is either sick or healthy and sick implies high spending. Figure 2.2 shows that there is minimal mass of total spending between 375 and 875 EUR, implying that 0.5 is a close approximation to the optimal threshold using a fully-specified cost distribution. To the extent that this threshold is an approximation, it is an upper bound on the exact threshold.

[21]For a CARA utility function of the form $u(z) = -\sigma e^{-\sigma z}$, the cutoff value π^* for switching to the high deductible being optimal is given by $(1 - e^{\sigma 250})/(e^{-\sigma 250} - e^{\sigma 250})$.

[22]An alternative model that could yield larger changes in the threshold for choosing the high deductible is a model of background risk where spending risk incurred in health insurance is correlated with other sources of financial risk, e.g. income risk. See Campbell and Viceira (2002) for a discussion. In our context, high risk aversion (10^{-3})combined with background risk of 1000 EUR income loss with a health event leads to a threshold of $Pr(Spend < 375) = 0.78$. Note that this level of risk aversion is likely implausibly high when integrating large scale background risk, due to the Rabin critique (Koszegi and Rabin (2006)). In our data it is possible to test whether negative health shocks are also correlated with negative income shocks.

when choosing a deductible (see Ericson and Sydnor (2018)). Note that in theory, liquidity and debt constraints could either increase the demand for insurance (to avoid large expenditures) or reduce the demand for insurance (to avoid paying the premium). As shown in Chetty and Szeidl (2007), under some assumptions one can characterize liquidity constraints as increased risk aversion, which relates to the discussion of changing threshold discussed above for different absolute risk aversion parameters. In our empirical analysis we include variables on (i) liquid savings (ii) income and (iii) net worth and show that the lack of liquid savings explains only a very small portion of why consumers under-adopt the high deductible when healthy.

Third, moral hazard could cause consumers to reduce care consumption in response to greater cost sharing (e.g., Newhouse (1993), Einav, Finkelstein and Schrimpf (2015), Brot-Goldberg et al. (2017)). Under a classical model of moral hazard, our framework under-predicts value from the high deductible plan since it rules out reductions in care that are lower in value than the associated cost savings. Since our empirical results focus on significant under-adoption of higher deductibles, having the lower bound interpretation does not impact the main import of our results. Moreover, we will show that deductible choice has a small impact on realized spending, holding all else equal, suggesting that the combined impact of selection on private information and moral hazard is small relative to the value embedded in the deductible choice.[23]

Given this discussion and our simple framework, the central factor that is important for assessing deductible choice value is an estimate of individuals' risk of spending more than 375 EUR (π). To do so, we develop an in-depth cost prediction model, described in the next section.

III.B Cost Prediction Model

For every individual, we generate yearly health risk predictions, with the explicit goal of evaluating the choice of the voluntary deductible. We set up our prediction model as a binary classification algorithm that predicts the probability (π_i) of having health expenditures below the compulsory deductible level

[23] An alternative explanation related to moral hazard is that consumers are subject to 'behavioral hazard' whereby they forego needed / valuable care when faced with higher cost sharing (see, e.g., Baicker, Mullainathan and Schwartzstein (2015)). If consumers were rational ex ante about their ex post behavioral hazard, they might optimally choose a low deductible despite having great financial value from not doing so. We investigate the impact of deductible choice on specific, potentially high value, categories of health care (e.g. always insured preventive care, basic primary care, drug use, mental health care) and show small impacts on consumption across these areas, suggesting that it is quite unlikely that foresight about behavioral hazard underlies low adoption of high deductibles (see Table A2.1 in Appendix VII.B). We discuss other potential behavioral micro-foundations in Appendix VII.D.

of 375 EUR. Thus, our prediction model accords with the underlying behavioral model.[24]

The yearly predictions of π_i are made using an ensemble learning model consisting of a random forest model, a boosted regression trees model and a LASSO model. Using such an ensemble learner is a standard technique to maximize prediction accuracy of a classification problem (Einav et al. (2018)). We only include predictors that are known at the time of choice of deductible (at the end of year $t-1$). The predictors that we include are: gender, year of birth, pre-tax household income in deciles ($t-2$, $t-1$), working status, education level, education field, and past health spending per category ($t-2, t-1$). In each year, there are approximately 20 variables for per-category health spending, so we have a fine level of detail with which to predict future medical spending. On average, we have approximately 50 predictors in our model every year.[25]

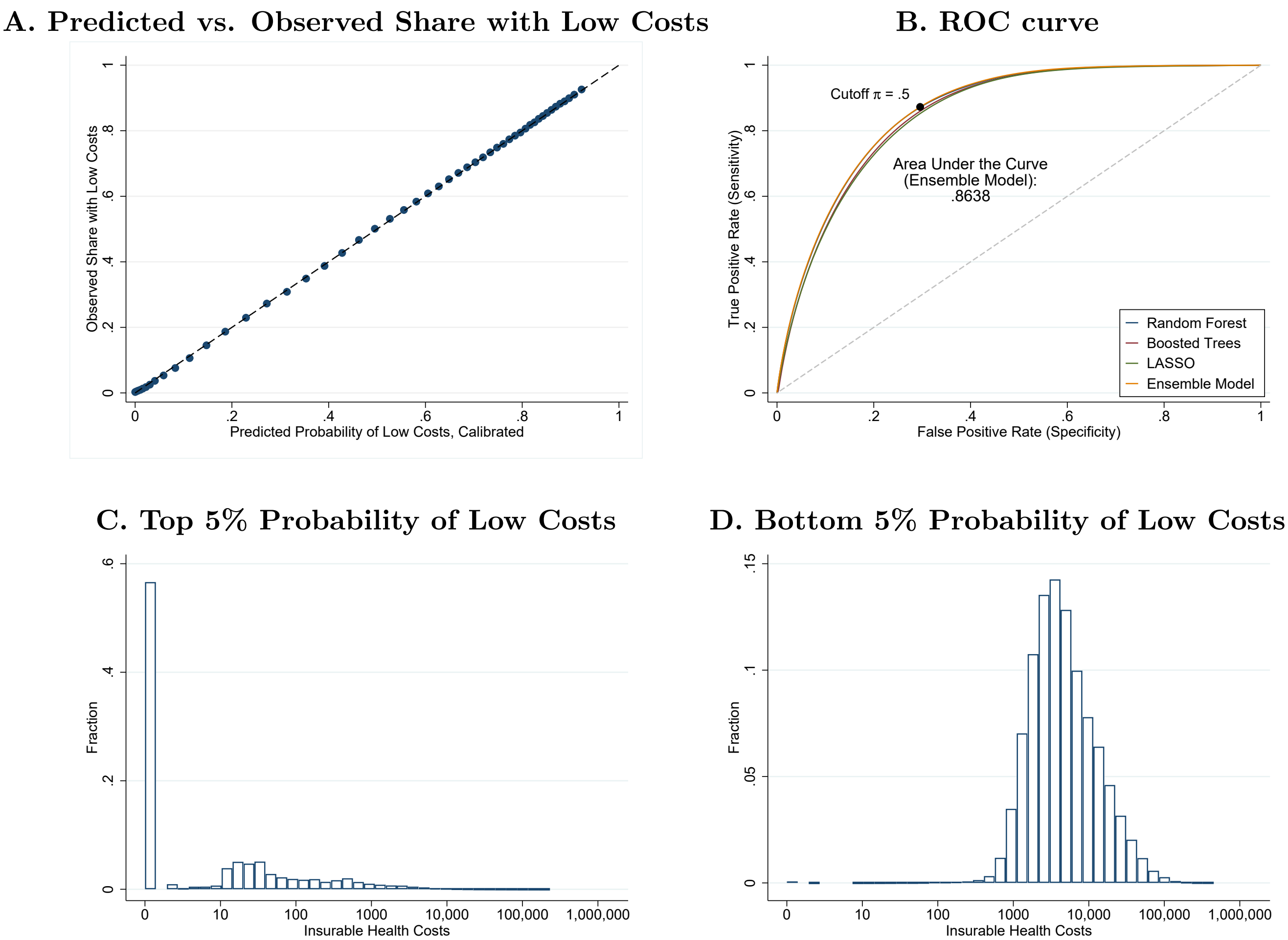
Our prediction algorithm follows four steps, similar to the prediction analysis in Einav et al. (2018). First, several key parameters of the random forest, boosted regression trees, and LASSO models are tuned. Second, these three separate prediction models are trained using a training sample. Third, the obtained predictions are combined into an ensemble predictor. Finally, the ensemble prediction is calibrated. We train the ensemble learner algorithm on a random sub-sample of 800,000 individuals. The training sample contains an additional 450,000 observations to combine the predictors and calibrate the ensemble predictor to observed data. All the results and plots in the analyses in this paper are then using only the hold-out sample of about 12 million observations each year. In Appendix VII.B, we provide more information on the detail of each step of the prediction.

Figure 2.3 describes the precision and outcomes of the prediction model. Panel A shows a bin scatter plot of the share of low-cost realizations by the predicted low-cost probability. The relationship between *ex ante* probabilities and *ex post* realizations is very strong as all observed shares are close to the 45 degree line. The ROC curve in Panel B shows that the ensemble model performs best and improves on the individual models. Panels C and D illustrate the predictive value of the model, comparing the distribution of realized cost for the top 5% and bottom 5% in terms of predicted low-cost probabilities.

[24] The empirical prediction model also underscores why intermediate ranges between 375 and 875 EUR are not useful choices. The distribution of health spending makes falling in that range of expenditures extremely unlikely. Therefore, predicting risk in this range is extremely difficult and choosing such a choice is almost never *ex ante* optimal. Figure 2.2 shows that the share of *ex post* realized expenditures that fall between 375 and 875 is 13.1%. When *ex ante* predicting which bracket individuals' costs would fall into, the sum of the raw predicted probabilities for the intermediary brackets is smaller than 1%. We provide further detail on this in Appendix VII.B.

[25] We have a different number of predictors in some years, as the categorization of health costs changes slightly in our study period. Every year, we include all health cost categories in our data set as predictors.

Figure 2.3: Predicted vs. Realized Costs

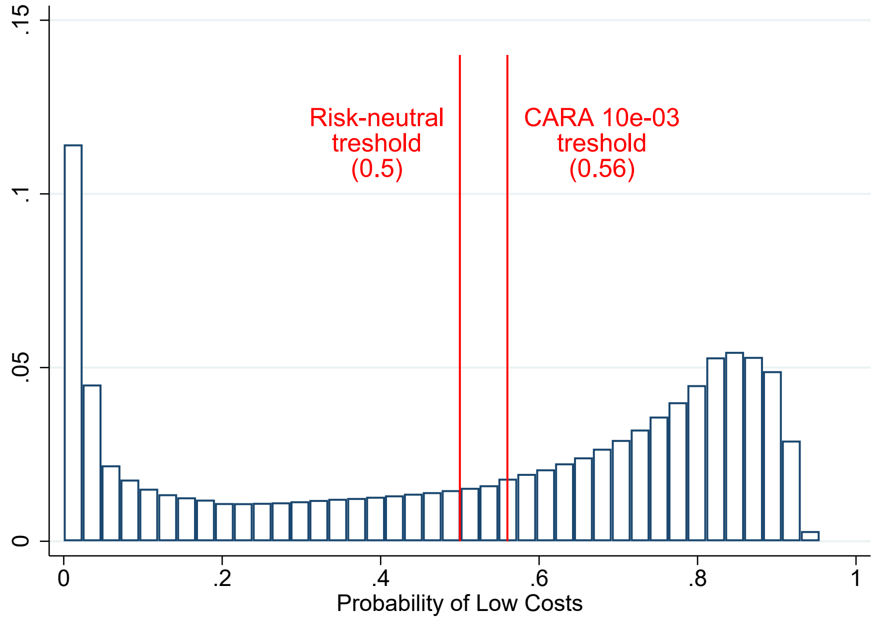


Notes: Panel A presents a binned scatter plot of our predicted probability of having low costs against the realized share of individuals with low costs. Panel B plots the ROC curve of the different prediction methods used. The bottom figures present ex-post cost realizations of individuals with predicted low (Panel C) and predicted high (Panel D) costs. The year is 2015 for all Figures.

The ex-post spending for the group that is predicted to be healthiest is much more skewed towards the low end of the distribution than the same distribution for the consumers predicted to be sickest, which is skewed towards the high end of the cost distribution.

One potential concern is that the cost model fits well "on average" but not for specific sub-groups that we study. Appendix Figure A2.1 shows that the prediction model is similarly well-calibrated for subgroups of individuals with different ages, education levels and income quartiles, showing that our empirical results finding different deductible take-up as a function of these variables (holding all else equal), is not due to cost mis-prediction. In addition, one might be concerned about the impact of

Figure 2.4: Distribution of Cost Probability Predictions



Notes: This figure shows the distribution of the predicted probabilities of having health costs below 375 EUR. These probabilities are obtained when predicting the binary variable (having insurable health costs below 375) with the ensemble machine learner described in Section III.B, and further in Appendix VII.B. The figure presents the risk-neutral threshold for someone to choose the 500 EUR incremental deductible if the incremental premium reduction is the modal incremental premium reduction of 250 EUR. It then presents the same threshold for extreme risk-aversion (CARA coefficient $1 * 10^{-3}$).

private information and/or moral hazard on cost prediction. In Panel A of Appendix Figure A2.1, the cost model prediction accuracy is plotted for individuals who take the 500 EUR deductible, and individuals who do not. The model fit is extremely strong conditional on take up of the low deductible. While individuals who take up an extra 500 EUR deductible do have an *ex post* higher chance to be low cost relative to our model predictions, the figure illustrates how this gap is small, suggesting a minor role for the combined effects of private information about health risk or moral hazard conditional on the predictors. These effects are certainly not big enough to have a meaningful impact on the positive results in Section IV and, as discussed in Section III.A and Appendix VII.B, quite unlikely to have a meaningful impact on our normative results in Section V.

Having established the predictive performance of the model, Figure 2.4 presents the histogram of the predictions for the *ex ante* probability of being in the low spending group. There is substantial dispersion in predicted risks over the full range of potential probabilities. The distribution is bi-modal, with a substantial share of individuals having either a very low probability or a very high probability of being low spenders. We include threshold measures for choosing the 500 EUR deductible to demonstrate that the distribution of risk places a significant share of the population well above and below the cutoffs respectively.

Taken together, these figures show that health expenses are, to a large extent, predictable and bimodal in our population. These features allow us to assess, with a good degree of robustness, whether a given individual is better off electing a high or low deductible.

III.C Barriers to Choice

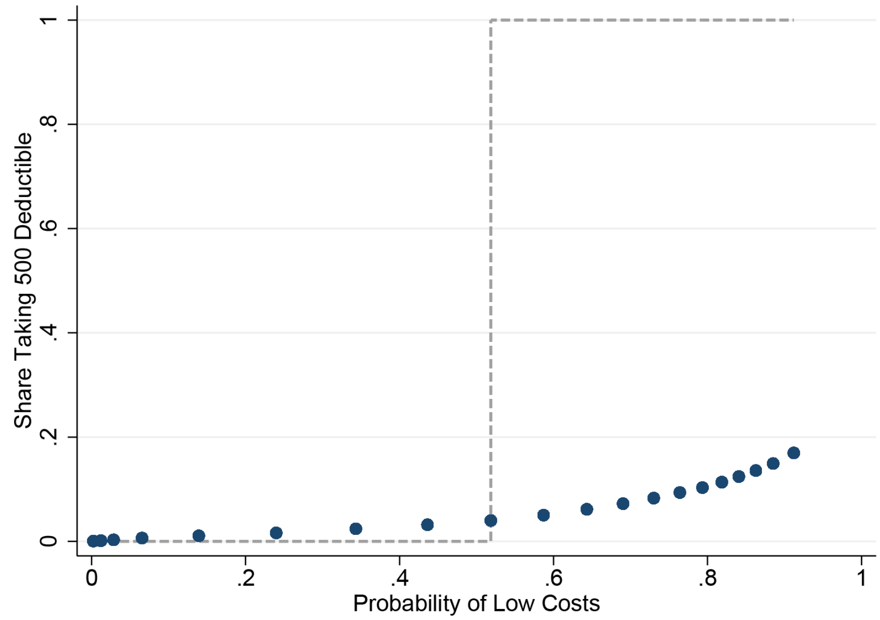
We can now study how deductible choices relate to predicted health risk, the primary component of deductible choice in a frictionless, rational model. Figure 2.5 plots the empirical relationship between predicted health risk and deductible choice and shows the optimal choice in the frictionless, rational model for comparison. Two key facts emerge. First, as expected, people who are healthier are more likely to elect the higher incremental 500 EUR deductible. Second, the relationship between risk and deductible choice is substantially weaker than one would expect if consumers were making utility-maximizing choices in the frictionless model. For example, the share of consumers in the healthiest predicted health bin electing the high deductible is only 17%, despite the fact that 100% would gain *ex ante* from taking the high deductible.[26] The same two key facts are confirmed when using only within-individual variation in predicted health risk. Appendix Table A2.1 reports the estimated coefficient on predicted health risk in a regression of 500 EUR deductible take-up when using also cross-sectional variation or only within-individual variation. In both cases increases in health risk lead to statistically significant decreases in extra deductible take up, but the effect size of the response is small.[27]

The empirical relationship in Figure 2.5 is in sharp contrast to the predictions of the frictionless model. This model suggests that, assuming individuals know their predicted health risk, the take-up rate should jump from 0 to 100% around a low-cost probability of .5. We recall that risk aversion,

[26] We note that moral hazard could underlie some of the positive correlation between deductible choice and health risk making, if anything, the selection on risk is *smaller* than what we observe. The confounding effect is arguably small, though. We use *ex ante* predicted health risks rather than *ex post* cost realizations limiting scope for actual spending to impact the relationship. We also find that the difference in predicted and realized risk for individuals who do take the extra deductible is very small (see Panel A in Appendix Figure A2.1). In the literature, the moral hazard effects of health coverage are generally estimated to be small relative to the effects we find. This is confirmed for the specific context by Remmerswaal, Boone and Douven (2019) using an age-discontinuity in the deductible choice at 18 years old.

[27] The estimated coefficient decreases from .115 to .0570 when using within-individual variation in a linear regression. More flexible specifications indicate that the response rate is proportional to the size of the change, but individuals are more responsive to negative (relative to positive) changes in predictable health risk. Appendix Figure A2.2 also shows the baseline plot disaggregated for the different years in our sample period.

Figure 2.5: Take-up of Voluntary Deductible as Function of Predicted Health Costs



Notes: This figure shows a binned scatterplot of the relationship between the predicted probability of having costs below 375 EUR (the compulsory baseline deductible) and the take-up of the voluntary 500 EUR extra deductible. The optimal choice in the frictionless, rational model is also shown for comparison.

liquidity effects and moral hazard have little impact on optimal choices in our setting, as discussed in Section III.A. However, there are a plethora of models with choice barriers one could write down that could help rationalizing the data (e.g., inertia, limited attention, misperceptions). For example, a model with default effects in combination with imperfect information about health risks can fit the data very well, as we illustrate in Appendix VII.D together with a number of alternative models.[28] Regardless of the nature of the choice barriers, the evidence shows that these barriers need to be large. Some of the healthiest individuals face a 90% chance of making costs below the lowest deductible, exposing themselves to an expected cost of only about 50 EUR when taking the highest deductible. Still, more than 80% of them forego on the 250 EUR savings in premium. Our goal is not to micro-found the discrepancy between the observed and what seem to be more desirable choices. We simply note

[28] Appendix VII.D simulates the choices for a set of alternative models of decision making that are proposed in the literature. This analysis also illustrates that some choice barriers do not help fitting the data. This includes rational inattention and random mistakes. Moreover, we already showed that even extreme risk aversion does not sufficiently depress the deductible take-up. We find the same /for example for loss aversion a la Kőszegi and Rabin (2007), which does not discourage individuals who are predictably healthy from taking the higher deductible. In contrast, in a model with switching costs and imperfect information, we find that sufficiently large switching costs can explain the depressed willingness to opt for a voluntary deductible, even when individuals are predictably healthy, and imperfect information about health risk can explain why even predictably unhealthy individuals sometimes choose to opt for a higher deductible.

that the role for standard consumer preferences in explaining the gap seems limited and we focus on uncovering the social factors that are related to this gap instead. As we will show, the nature of the relevant social determinants further indicates the importance of choice barriers.

IV Empirical Analysis of Deductible Choice

In this section we study the social determinants of deductible choice. We continue to build on our stylized model, relating choice to health risk, but we incorporate individual and environmental factors that may capture choice elements outside of the standard model and drive a wedge between the observed and optimal choices.

IV.A Socio-Economic Correlates of Deductible Choice

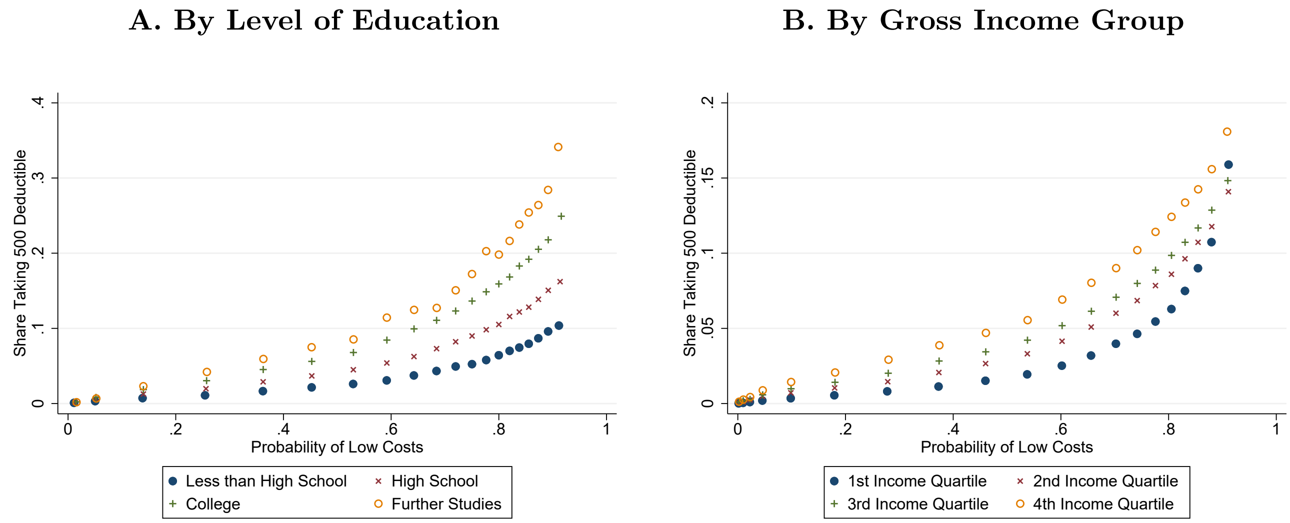
We first turn to understanding how different individual factors, in particular demographic and socio-economic variables, change deductible choice with respect to health risk. We do so both by presenting non-parametric graphical evidence, following Figure 2.5 but dividing the population by observable characteristics, and by formalizing these results in a simple regression framework, using variation across and within individuals. We rely on a simple OLS regression in a linear probability model:[29]

$$Y = \alpha + \gamma X + [\beta + \nu X] P(costs < 375) + \epsilon \tag{2.3}$$

where Y is an indicator variable taking the value of 1 when an individual takes the 500 voluntary deductible and 0 otherwise, $P(costs < 375)$ is the predicted probability of having costs lower than 375 EUR (π_i in our theoretical model), and X includes all variables of interest. The primary coefficients of interest are γ and ν. The former captures how different observables affect the intercept, i.e., the average take-up of the 500EUR deductible by individuals who are the sickest (with $\pi_i = 0$). The latter measures how different factors affect the relationship between risk and deductible choice. $\gamma + \nu$ captures the impact on average take-up by individuals who are the healthiest (with $\pi_i = 1$). Each regression also includes year and insurer fixed effects. The insurer fixed effects control for potential differences in insurer marketing / steering and/or differences in insurer incremental deductible premium, though as

[29] As alternatives, we relax the linearity assumption: $Y = \alpha + \gamma X + [\beta + \nu X] \times 1[P(costs < 375) \geq .5] + \epsilon$, and also consider a probit model: $Pr(Y = 1) = \Phi\left(\alpha + \gamma X + [\beta + \nu X] \times P(costs < 375) + \epsilon\right)$. The findings are unchanged, as shown in Appendix Table A2.2.

Figure 2.6: Deductible Take-up by Education and by Income



Notes: These figures show binned scatter plots of the relationship between the predicted probability of having costs below 375 EUR (staying under the voluntary deductible range) and the take-up of the voluntary 500 EUR extra deductible, by education level in Panel A and by household gross income quartile in Panel B. In Panel B, we have excluded the group of individuals with gross income below minimum social assistance, which mostly consists of students, self-employed and households with negative capital income.

we showed earlier there is limited dispersion in the latter.

Socio-Economic and Demographic Factors Figure 2.6 plots the relationship between health and deductible take up by education level and income. Panel A shows a large difference in the relationship by education level. Those in the healthiest predicted risk decile with a college degree (i.e., bachelor or master) elect the higher deductible about 23% of the time and those with an advanced degree choose the highest deductible 30% of the time. In contrast, those with less than high school education in the healthiest predicted decile elect the higher deductible only 10% of the time and those with high school education only approximately 15% of the time. For all of these education levels, when people are predicted to be sick they almost never elect the higher deductible. Panel B of Figure 2.6 shows the same relationship by quartiles of gross income (including capital income and government transfers). The relations are similar to those we see for education. Higher levels of income are associated with higher take-up of high deductible among the healthiest.[30]

Table 2.4 presents results from the regression model in equation 2.3 focusing on income and education

[30]We note that income effects may directly increase the take-up of deductibles, but their interaction with the predicted low-cost probability is ambiguous.

Table 2.4: Deductible Take-Up: Baseline Regression Estimates

	Take-up of 500 Deductible	
	intercept	*slope*
High School	-0.011***	0.057***
College Degree	-0.034***	0.165***
Further Studies	-0.047***	0.226***
2nd Income Quartile	0.004***	-0.007***
3rd Income Quartile	0.004***	0.007***
4th Income Quartile	0.002***	0.039***
36 to 50 years old	0.020***	-0.045***
51 to 65 years old	0.029***	-0.047***
65+ years old	0.034***	-0.082***
Male	-0.004***	0.025***
Has Partner	-0.002***	0.013***
Has Children	0.004***	-0.028***
Self-employed	-0.006***	0.026***
Constant	-0.041***	
Prob. Low Costs		0.098***
Year and Insurer FE	YES	
Observations	57,100,388	

Notes: This table plots coefficients from our regressions studying deductible choice, as explained in Section IV. Each variable is interacted with the probability of having low health expenses; the impact on the intercept is reported in the first column, and the impact on the slope in the second column. The dependent variable in all specifications is a dummy that takes value of 1 when the individual takes up the voluntary 500 EUR extra deductible. The prob. costs < 375 EUR variable is obtained from our prediction algorithm. The reference groups for the different demographic categories are: 1st income quartile, education lower than high school or unknown, and age between 18 and 35. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

level. The estimated intercept coefficients are α and γ and slope coefficients β and ν.[31]

There is significant and economically meaningful variation in slopes, as expected based on the graphical evidence. The effects, however, are mostly driven by differences in education. The interaction with the predicted health risk is indeed substantially larger for those with higher education reflecting the fact that individuals are more responsive to their health status in selecting the higher deductible with higher education levels. An individual in good health — *ex ante* very high probability of being low cost — who has completed graduate studies beyond college is 23% more likely to take up the high

[31]For completeness, Appendix Table A2.3 shows the estimates when not including interaction terms between controls and predicted risk.

deductible that an equivalent person with less than a high school education.

The interaction of income and the gradient of take-up has the same pattern though the effects are modest once we include controls; the highest income quartile is slightly less than 4% more likely to take up the high deductible if they are in good health compared to the lowest income quartile.[32] Because the regression includes both income and education these results suggest a stronger role for education itself rather than income levels

In comparison to the variation in slopes, there is relatively little variation in the intercepts. For those in worst health — *ex ante* probability of zero of having cost less than 375 EUR — higher education is associated with a lower rate of take-up of the higher deductible. The effect of income, however, is the opposite. As can be expected from the graphical evidence, some of these differences change when relaxing the linearity assumption on the relation between take-up and risk, but they are consistently small (see Appendix Table A2.2).

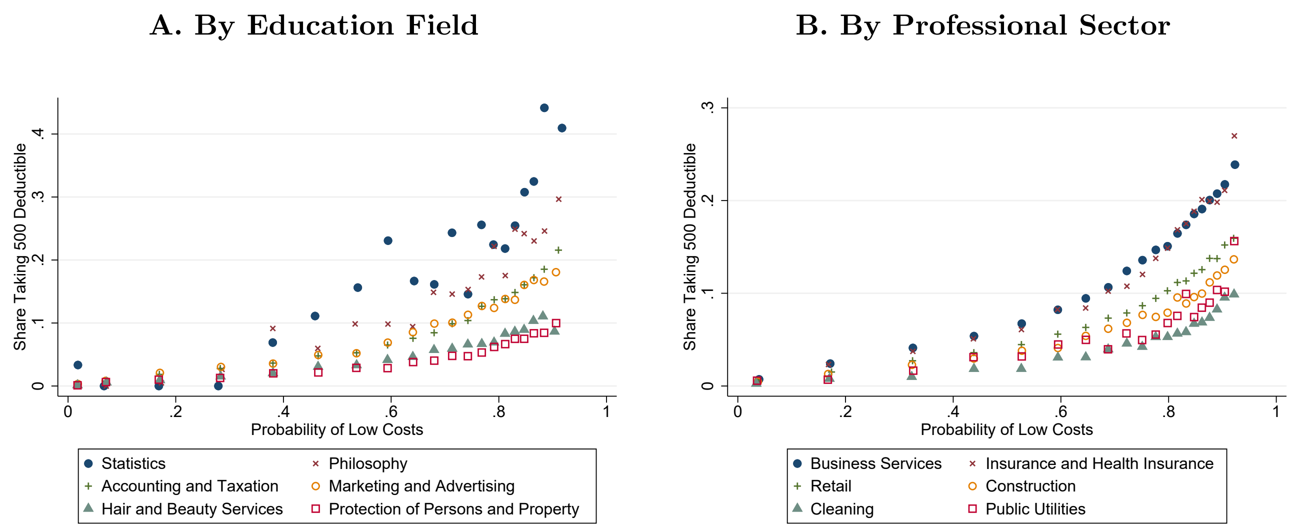
Table 2.4 also presents the effects of age, gender and household composition on deductible choice, controlling for health risk, income and education level. There are statistically significant differences in responsiveness to underlying health risks, though the magnitude of the effects are relatively small. We also note that, despite the relative simplicity of the models we estimate, these effects are very robust to alternative specifications. For brevity, we present those results in Tables A2.2.

Human Capital Overall, Table 2.4 demonstrates that the strongest relationship between deductible take-up and observable characteristics is for education level. This is indicative of the potential role of expertise, cognitive ability or information frictions in insurance choices. To shed more light on the role these effects may play we perform the same analysis as above but use richer data on the specific field of education and professional sector of employment.

Figure 2.7 plots the relationship between deductible choice and predicted health risk by education field and professional sector. Since there are many education fields and professional sectors, we present only 6 specific fields and sectors that are indicative of the broader patterns. Statistics majors are the most responsive to predicted health risk: they choose the additional deductible approximately 43% of the time when they are in the healthiest predicted health bin and choose the additional deductible

[32] To illustrate this further, Appendix Figure A2.1 plots the relation between take-up and household income. The income gradient is as important in magnitude as for predicted health risks, but once we control for predicted health and other variables capturing socio-economic status, the gradient becomes nearly flat.

Figure 2.7: Deductible Take-up by Education Field and by Professional Sector



Notes: This figure shows for 6 fields of study and 6 professional sectors a binned scatterplot of the relationship between the predicted probability of having costs below 375 EUR (the compulsory baseline deductible) and the take-up of the voluntary 500 EUR extra deductible. Refer to Tables A2.5 and A2.6 for an overview of the deductible take-up in all fields and sectors, respectively.

almost never when they are in the sickest predicted bin. The effect stands in stark contrast to those with training in "Protection of Persons and Property" or "Hair and Beauty Services." Even for the healthiest group in those fields, take-up of the higher deductible is only approximately 10%. Similarly, for professions that are more analytical in nature and professions that require more advanced schooling, deductible choice is also higher for those with low risk — the prediction of the standard, rational model. Despite that, however, we also see that even for those in the insurance industry, take-up is only around 30% for those in the best health.

Table 2.5 presents the corresponding regression analysis, including baseline controls for predicted health risk, income, education level, age, gender and household structure. Even controlling for these other factors, more quantitative / analytic professions (e.g., statistics) are more responsive to predicted health when making deductible choices (column 1). For example, among the predictably healthy, someone with statistics training is 28.2% more likely to choose a higher deductible, controlling for age, income, gender, and education level than someone with hair and beauty training. Column 2 of Table 2.5 also confirms that more analytic professions (e.g., business services and insurance) are more responsive to predicted health when making deductible choices. For example, someone in the insurance sector who is predictably healthy is approximately 8% more likely to choose a higher deductible, controlling for

Table 2.5: Deductible Take-Up Regression by Subgroup

	(1) Education Field		(2) Professional Sector		(3) Liquidity and Financials		(4) Environment	
	intercept	*slope*	*intercept*	*slope*	*intercept*	*slope*	*intercept*	*slope*
Statistics	-0.042**	0.247***						
Philosophy	-0.003	0.046***						
Accounting and Taxation	-0.003***	0.024***						
Marketing and Advertising	-0.000	-0.004						
Hair and Beauty	0.007***	-0.035***						
Protection of Persons	0.008***	-0.068***						
Business Services			-0.012***	0.045***				
Insurance			-0.025***	0.078***				
Retail			-0.002***	-0.002*				
Construction			-0.001	-0.018***				
Cleaning			0.003***	-0.033***				
Public Utilities			0.006***	-0.008*				
2nd Net Worth Quartile					0.003***	-0.004***		
3rd Net Worth Quartile					0.000*	0.021***		
4th Net Worth Quartile					-0.002***	0.061***		
Has Savings > 2000EUR					-0.006***	0.028***		
Has Mortgage Debt					-0.000	0.005***		
Has Other Debt					0.005***	-0.023***		
Share of Colleagues with 500 Ded.							-0.105***	0.459***
Share in Postcode with 500 Ded.							-0.329***	1.055***
Father With 500 Deductible							-0.029***	0.288***
Mother With 500 Deductible							0.015***	0.294***
Constant	-0.043***		-0.050***		-0.042***		0.010***	
Prob. Low Costs		0.101***		0.117***		0.094***		-0.060***
Baseline Controls	YES		YES		YES		YES	
Year and Insurer FE	YES		YES		YES		YES	
Observations	30,799,129		32,299,835		57,013,765		16,938,401	

Notes: The regressions follows our baseline specification (see Table 2.4). Additional controls are: in Column (1), dummies for six selected educational fields of study, as well as their interactions with health risk. The reference category for field of study is all other fields of study; in Column (2) dummies for six selected professional sectors, as well as their interactions with health risk. The reference category is all other sectors; in Column (3), a dummy for liquidity (household savings>2000EUR), a dummy for having household mortgage debt and other household debt, household net worth quartiles, as well as their interactions with predicted health risk; and in Column (4), the fraction of individuals taking up an extra 500 EUR deductible in firm and neighborhood, and dummies for whether the father or mother is taking up an extra 500 EUR deductible. Note that the shares are calculated excluding the individual for which the share is calculated (i.e. the person's take-up is excluded from both numerator and denominator), and shares are calculated only if there are more than 10 individuals that firm or neighborhood. ***p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

age, income, gender, and education level, than someone in the public utilities sector.

To shed further light on the relationship between the specific field of study and deductible choice we report key take-up measures for a selection of fields in Table 2.6. Columns 1 and 2 present the share taking up the high deductible and the predicted low-cost probability respectively. The primary results of interest are presented in column 3, which shows the rate of take-up of the high deductible among those with a high probability of having low cost — the group for which we expect high adoption under the standard model. The table shows that quantitative fields are grouped at the top of the table, exhibiting greater responsiveness to predicted health risk when making deductible choices, while those in less quantitative fields are grouped at the bottom of the table, exhibiting lower responsiveness. An exhaustive list of education fields is presented in Appendix Table A2.5. We present a similar analysis of professions in Table A2.6 in the appendix, showing a very similar gradient by professional sector. More analytical sectors exhibit greater responsiveness to predicted health risk when making deductible choices while those in less analytical sectors exhibit lower responsiveness.

Financial Capital In addition to an individual's human capital and income, we observe a range of additional variables related to a household's financial capital. We study the relationship between deductible choices and wealth, measured by the household's net worth, debt (mortgage or any other debt) and a measure of liquidity that takes on a value of 1 if a household has more than 2000 EUR in liquid savings and 0 otherwise.

Table 2.5 presents the results of a regression examining the association between these financial variables and incremental deductible take up, controlling for predicted health spending and our baseline controls. We find that household liquid savings are positively correlated with deductible take up: having liquid savings of greater than 2000 EUR is associated with a 0.8 percentage point increase in deductible take up. Note that in theory, liquidity and debt constraints could either increase the demand for insurance (to avoid large expenditures) or reduce the demand for insurance (to avoid paying the premium) (see Ericson and Sydnor (2018)). The sign of the effect we find is consistent with the former explanation. In line with this, we also find that households who are in debt (excluding mortgage debt) are also less likely to take-up the deductible. The effects, however, are small in both cases. Finally, we find that take-up rate for wealthier individuals is higher and this effect is fully driven by wealthier individuals with better health. That is, wealthier individuals are more responsive to taking the incremental deductible as they become healthier. Hence, rather than capturing wealth effects on insurance choices, this result is indicative of choice barriers for people with fewer financial resources.

Table 2.6: Deductible Take-up and Field of Study

	Education Field	(1) Take-up of 500 Deductible	(2) Probability Low Costs	(3) Take-up of 500 Ded. | Being Predictably Healthy
1	**Statistics**	29%	87%	34%
2	Mathematics	21%	85%	27%
3	Physics	21%	91%	26%
4	Architecture and town planning	18%	88%	21%
5	Physical science	18%	82%	22%
6	Earth science	18%	88%	21%
7	**Philosophy and ethics**	17%	82%	21%
8	Medicine	17%	83%	20%
16	Sociology and cultural studies	14%	82%	18%
17	Mining and extraction	14%	91%	17%
18	Economics	14%	84%	17%
19	Humanities and Arts	14%	84%	18%
41	**Accounting and taxation**	11%	78%	14%
42	Agriculture, forestry and fishery	10%	81%	13%
43	**Marketing and advertising**	10%	80%	13%
83	Secretarial and office work	5%	65%	7%
84	**Protection of persons and property**	4%	78%	6%
85	Child care and youth services	4%	66%	6%
86	Computer use	4%	65%	6%
87	**Hair and beauty services**	4%	65%	5%
90	Literacy and numeracy	2%	62%	4%

Notes: For a selection of fields of study, this table shows: in Column (1), the fraction of individuals who take-up the 500 EUR extra deductible, in Column (2), the fraction of individuals with a probability of low costs $<$ 375 EUR, and in Column (3), the fraction of individuals who take-up the 500 EUR extra deductible, conditional on having predicted health costs $<$ 375 EUR. The full list of fields is provided in Appendix Table A2.5.

IV.B Peer Effects on Deductible Choice

Thus far we have studied socio-economic characteristics and documented the important role of long-run human capital on deductible choice. We now turn to the role of environmental factors, measured by exposure to peers' choices. Specifically, we investigate the impacts of the deductible choices by (i) co-workers, (ii) neighbors and (iii) parents. We analyze these effects in two ways. First, we follow the same cross sectional approach from equation 2.3 including peers as observable characteristics. Second, to get causal estimates we exploit within-individual variation in peers due to moves across firms or geography and in parents' choices.

Cross-sectional Estimates Table 2.5 presents the cross-sectional regression estimates for the association between individuals' take-up of the deductible and the take-up rates by their respective peers, closely following our main regression equation and controlling for health risk, baseline demographics, education level and income. For the firm take-up rate, we calculate the proportion of individuals taking the 500 EUR deductible in an individual's firm, defined at the establishment level, excluding herself. For the location take-up rate, we calculate the proportion of individuals in an individual's 6-digit postcode taking the 500 EUR deductible, excluding herself. For parental deductible choice, we use a variable that is one if a given parent elects the 500 EUR deductible.

The cross-sectional associations between these environmental factors and deductible choice quality are very strong. For example, these regressions find that when the share of colleagues choosing a high deductible in a firm is 10% higher then the probability a given individual chooses an extra deductible is 1.0% lower when predictably healthy, but 3.5% higher when predictably healthy. For location, an increase in the local take-up rate by 10% increases the take-up probability by 7.0% for predictably healthy individuals. For intra-family deductible choices we find that if an individual's father (mother) chose the 500 EUR deductible, and that individual has good predicted health, then that individual is 25.7% (31.0%) more likely to elect the high deductible themselves.

While these cross-sectional correlations are instructive, there is a long literature discussing the reflection problem in analysis of peer effects, where it is easy to confound underlying correlated unobservables for a peer group (see, e.g., Manski (1993)). We now turn to panel analyses that aim to quantify the causal implications of these peer effects for deductible choice.

Co-workers and Neighbors: Movers Design

We use the deductible choices by firm switchers and location movers in a two-part framework to quantify the causal impact of place of work or home on deductible choice. Note that this causal impact could be a combination of both (i) peer effects and (ii) firm or location-specific unobserved heterogeneity (e.g. the firm promotes a certain kind of deductible choice).

The first part of our framework obtains individual fixed effects and firm or location fixed effects from an linear OLS framework, similar in spirit to Abowd, Kramarz and Margolis (1999):

$$y_{i,x,t} = \alpha_i + \gamma_t + \theta_x + \beta_1 w_{i,t} + \beta_2 \pi_{i,t} + \epsilon_{i,t}$$

Here, α_i is an individual fixed effect, γ_t is a time period fixed effect and θ_x is a firm fixed effect. $w_{i,t}$ and $\pi_{i,t}$ are the individual household's gross income and health level (in deciles).

In the second step, we regress the obtained fixed effects θ_x on the average share of the of high-deductible take-up in firm or location x over time:

$$\theta_x = \beta \bar{h}_x + \epsilon_x$$

Crucially, when we implement this two-step framework we must decide whether to include switchers themselves in the second step. The argument against including switchers in the second step is that these are the same individuals identifying the fixed effects in step one, so that if there are a lot of switchers at a firm/location regressing the fixed effect on $\bar{h}_x$ becomes closer to regressing a variable on itself. The argument for including switchers is that they are likely more dynamic than others in the firm/location, since they may have re-optimized things about their lives, including deductible choice, more recently. If we exclude these switchers in step two, but they are more influential on the choices of peers, then we may mis-estimate the impact of $\bar{h}_x$ on θ_x.

To deal with this issue, we split the sample in half, and run step one on half the sample and run step two on the other half of the sample, using step one fixed effect estimates on the left-hand side of the step two regression. This approach allows for switchers to be included in both step one and step two without having step two regress fixed effects on observations that directly identified those fixed effects. Our large sample size allows us to have strong statistical power despite only using half the sample in each regression, though we do focus on larger firms to mitigate issues related to estimating noisy fixed effects for smaller firms. While we use this split-sample approach as our primary approach, we also show results for this two step framework where we don't split the sample and we either (i) exclude all switchers from step two or (ii) include all switchers in step two.

Table 2.7 presents our results for firms and postcodes. We show results for all firms with more than 100 employees and for all firms with greater than 500 employees. The split-sample results show that, when the proportion of individuals in a firm choosing the high deductible is 10% higher, the firm fixed effect is around 1.4% higher. Panel A of Figure 2.8 shows the relation between the firm take-up rates and the fixed effects to highlight the strong fit. Thus, there is a meaningful causal effect: someone who switches to a firm is more likely to choose a high deductible if others in the firm are

Table 2.7: Switchers Design: Firm and Postcode Effects

	Firms		Postcodes	
	> *100 employees*	> *500 employees*	> *500 inhabitants*	> *2000 inhabitants*
Baseline case: split sample	0.135***	0.145***	0.101***	0.151***
	(0.010)	(0.015)	(0.009)	(0.017)
Including switchers	0.208***	0.169***	0.120***	0.166***
	(0.008)	(0.012)	(0.008)	(0.016)
Excluding switchers	0.099***	0.149***	0.057***	0.088***
	(0.008)	(0.014)	(0.009)	(0.018)

Notes: This table displays the results of an AKM-style regression capturing peer effects at the firm and the postcode level. In a first step, firm and postcode fixed effects are obtained from regressing individual take-up of the 500 deductible on household gross income and probability of low costs in deciles, with individual and time fixed effects. In a second step, firm and postcode fixed effects are regressed on the share of take-up in an individual's firm or postcode. The results of this regression are displayed here for different minimum sizes of firms and postcodes, and different identification methods for the fixed effect. In the first row, fixed effects are computed off one randomly selected half of the sample, and the second step regression is computed off the other half. In the second row, both the first and the second step are performed on the entire sample. In the third row, the first step is performed on the entire sample, but the second one excludes firm or postcode switchers. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

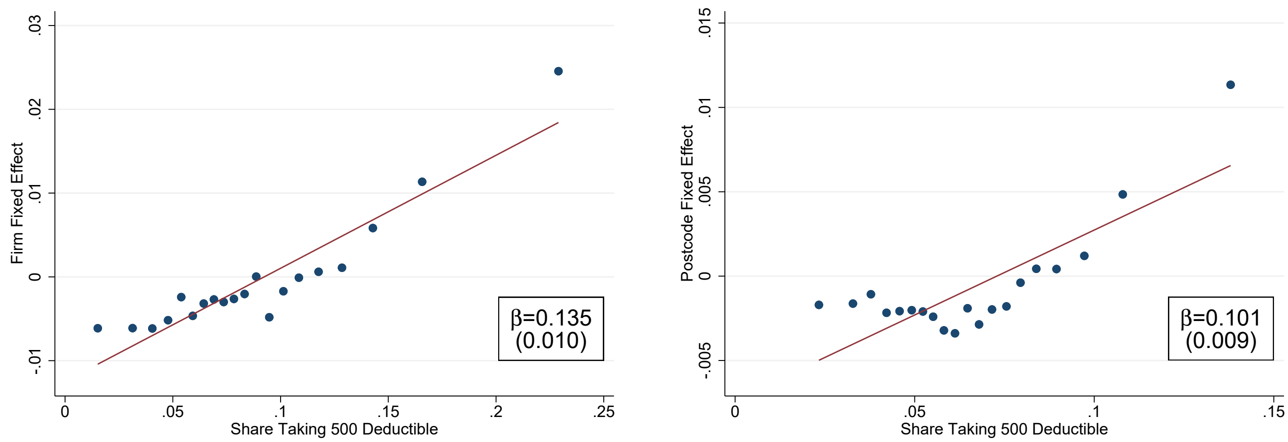
doing so, controlling for health and income. The causal estimate also explains more than half of the cross-sectional relationship between firm and individual choices (see Appendix Table A2.3).

When we include switchers in step two, and don't use the split-sample approach, these estimated coefficients are higher, implying an effect of 2.1% for firms with more than 100 individuals. Thus, not surprisingly, including the same individuals in step two who we used to identify fixed effects in step one biases our coefficients upwards. Conversely, when we don't use the split-sample approach but exclude switchers from step two, our coefficients are biased downward (1.0% for firms with more than 100 people). This suggests that these switchers may ultimately be more influenced by the choices of peers than other, more static, employees at the firm.

We use the same two-step approach to investigate the impact of neighbors / postcode on deductible choice. Table 2.7 presents the results for postcodes with more than 500 individuals and postcodes with more than 2000 individuals. Our primary split-sample approach shows that for a 10% increase in postcode high-deductible take-up, 1.0 % more individuals causally take-up the high deductible in neighborhoods with more than 500 people, and 1.5 % more do so in neighborhoods with more than 2000 people. Interestingly, for neighborhoods, these numbers are very similar when we implement the full sample specification including movers, perhaps because movers are a lower proportion of people in the postcode relative to firms. When movers are excluded, the estimates are much lower (.57%)

Figure 2.8: AKM Results: Firms and Postcodes

A. Firm Effects **B. Postcode Effects**

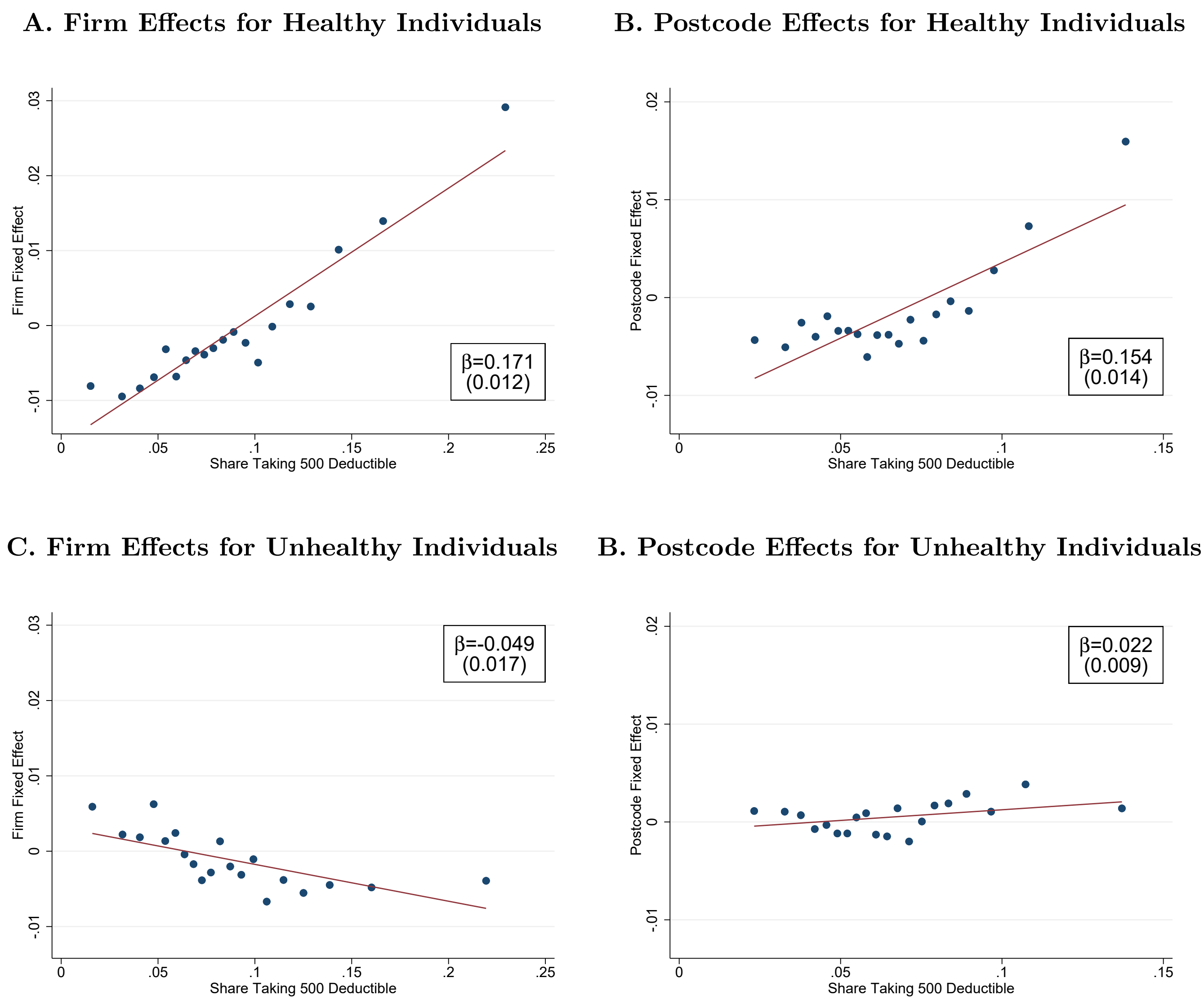


Notes: This figure shows the relationship between firm (Panel A) and postcode (Panel B) fixed effects, and the share of take-up of the 500 deductible in the firm or postcode. Fixed effects are obtained from regressing individual take-up of the 500 deductible on household gross income and probability of low costs in deciles, with individual and time fixed effects. The share of take-up is then computed for each individual as the share of colleagues or neighbors who chose the high deductible (i.e., excluding herself), averaged over employees and over the five years in our sample. In Panel A, we include all firms employing 100 people or more; in Panel B, all postcodes with a population of 500 or more.

for postcodes with more than 500 people) suggesting perhaps that the movers do have an out-sized influence on neighborhood peer effects. Figure 2.8 plots the regression fit for our primary estimates (split-sample), highlighting the strong fit and the differences in results across these two specifications.

While these results are informative about the causal effects of firms and locations on deductible choice, we also want to gauge their potential impact on deductible choice quality. To shed further light on this, we also re-run our AKM approach on each of two samples: (i) individuals who are predictably healthy (with predicted low cost probability above 50% - such that the high deductible is the right choice - in all 5 years) and (ii) individuals who are predictably unhealthy (with predicted low cost probability below 50% in all 5 years). The left panels in Figure 2.9 present the results for the firm fixed effects using our primary split-sample approach. The results are clear: when an individual is predictably healthy, the firm effect is strong and positive, with a 10% increase in the number of healthy people taking up a high-deductible causing a 1.7 % increase in high-deductible take-up for healthy people switching into the firm, holding all else equal. Conversely, an individual who is predictably sick is *less* likely to take up a high deductible if more people in the firm do take up that deductible, though this relationship is relatively flat. The right panels in Figure 2.9 present the results for the location fixed effects. The

FIGURE 2.9: AKM RESULTS: SPLITTING BY HEALTH STATUS



Notes: Notes from Figure 2.8 apply; but here the relationship between the fixed effects and the share of take-up is plotted separately for individuals who are predictably healthy (i.e., with a probability of low costs greater than .5 in all five years in our sample), in Panel A and B, and predictably unhealthy in Panel C and D (for whom the probability of low costs is below .5 for all five years).

results are very similar: when an individual is predictably healthy, the postcode effect is strong and positive, with a 10% increase in the number of healthy people taking up a high-deductible causing a 1.5% increase in high-deductible take-up for healthy people switching into the postcode, holding all else equal. Conversely, an individual who is predictably sick is not more likely to take up a high-deductible. This relationship is now basically flat.

In the same spirit, Panels A and B of Appendix Figure A2.4 plot the relationship in the data between predicted health and deductible choice for individuals grouped in quartiles of the firm and location fixed effects. The Figure shows that the difference in take-up rates across firms and locations are again larger for individuals who are predictably healthy, while the take-up rates by individuals who are predictably unhealthy are consistently low. As the dispersion in firm and location take-up rates is relatively small, the overall differences in take-up rates are less pronounced than our earlier results, for example comparing individuals with different education and income in Figure 2.6.

Taken together, these results suggest that both firm and location effects are strong and positive, but only when an individual is predictably healthy and *should* take up a higher deductible and not when they are predictably sick and should not take up the higher deductible.

Parents: Event-study Design

For the effects of parents' choices on their children, we obviously cannot use the AKM design. Instead, we rely on an event-study design to investigate the causal linkage between parents' and adult children's decisions. In particular, we study the deductible choice of adult children when one parent switches from not taking any voluntary deductible to the 500 EUR deductible. We estimate the following specification:

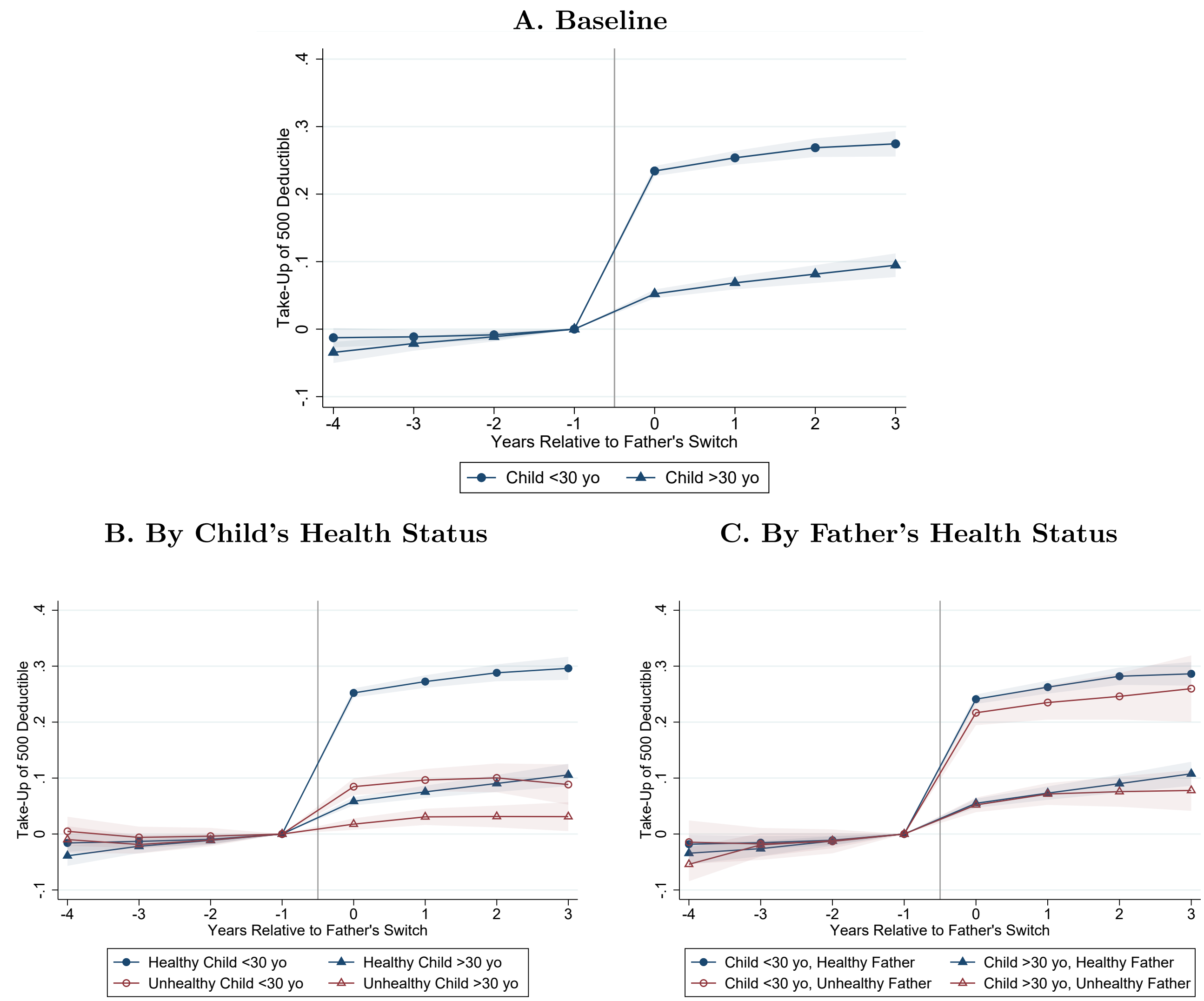
$$d_{it} = \gamma_t + \sum_{j=-N_0}^{N_1} \beta_j \cdot \mathbf{1}[J_{it} = j] + X_{it}\beta + \epsilon_{it}.$$

Here, γ_t is a time fixed effect, $J_{it} = t - E_i$ denotes event time, that is the time in years relative to the moment that the parent switched, and $[-N_0; N_1]$ is the window of dynamic effects around the event. We restrict our sample to "stable" changes, i.e., we exclude individuals whose parent's deductible was not always zero before the switch and is not always 500 after the switch, during the five year window we consider. The causal impact could be a combination of peer effects - either from the parent on the child or vice versa - and some unobserved heterogeneity in the family. In particular, the parents may make the actual deductible decision for their adult children. To mitigate the latter, our main specification presented here excludes families where the parents and adult children are still living together and we report the estimates for children who are younger and older than thirty.

Figure 2.10 shows the dynamic impact of a father's deductible switch on his children's decisions.[33] The estimates show a clear discontinuous increase in the take-up of the deductible in the year the father

[33] We focus on one parent here to abstract away from number of parents who switch. Results for mothers are similar to results for fathers.

Figure 2.10: Parent Effect on Deductible Choice



Notes: The Figure shows the estimates of the dynamic effects using an event-study design of the impact of a parent (here, father) switch from a 0 to a 500 deductible on a child's take-up, excluding all children who still live with their parents. The baseline regression displays the estimates, split between children who are younger or older than 30 years old. The two bottom figures split the impact between predicably healthy/unhealthy children (left) and between predictably healthy/unhealthy fathers (right). Years considered are 2013 to 2017.

switches. Children over 30 are, not surprisingly, less likely to follow their father's lead, though their is still a meaningful effect. The increase is 23 percentage points for children under 30 and 6 percentage points for children above 30. These causal estimates are respectively above and below the cross-sectional estimate of 18 percentage points reported in Table 2.5. In both cases, there is little anticipation in the take-up rate in the years before and the effect persists in the years after.

We also investigate the heterogeneous event impacts as a function of children's health status and also as a function of parent's health status. We abstract away from health status changes that occur in the five-year window and assign individuals to healthy or unhealthy based on their average predicted health over this time period. The impact is significantly larger for children who are in good health, as shown in Panel B of Figure 2.10. For the children under 30, there is a 30-40 % higher chance that they also switch to a high-deductible when in good health. This increase is only about 15% when they are in bad health. Panel C of Figure 2.10 shows the same analysis, but as a function of the father's health status instead of the child's health status. Interestingly, effect heterogeneity as a function of father's health is much lower than heterogeneity as a function of child's health, as children are similarly likely to switch regardless of whether their father took the 'right' decision by switching to the high deductible or not. The overall relation between childrens' predicted health and deductible choice grouped by the take-up of their parents is shown in Panel C of Appendix Figure A2.4. Unlike for firms and locations, we cannot rank individuals by the causal effect their parents take-up may have.

Taken together, these results suggest that parental effects are strong and positive, but only when a child is predictably healthy and *should* take up a higher deductible.

V Inequality in Choice Quality

The analysis thus far has covered a variety of observable characteristics that affect deductible choices and how their choices compare to choices we expect from rational consumers in a frictionless environment. The evidence on the key factors explaining these gaps - in particular the role of human capital and peers in particular - corroborates the earlier conjecture in Section III that barriers to choice are important in practice. This section provide a quantification of the potential welfare loss under this interpretation of choice barriers, but ignores any direct welfare effects of the specific underlying choice frictions beyond the misallocation to plans (e.g., search or switching costs). This quantification allows us (i) to underline the key dimensions of inequality in choice quality, (ii) to highlight some important interactions and (iii) to evaluate the welfare impact of choice-based government interventions, accounting for both efficiency and equity considerations.

V.A Heterogeneity in Choice Quality

We begin by defining an empirical measure of choice quality. We follow our stylized model in Section III.A, where an individual can opt for the extra 500 EUR deductible at a premium of 250 EUR and her expected utility depends on her predicted probability of achieving low costs (less than 375 EUR). Using our earlier benchmark of frictionless decision making, we can define the welfare loss due to barriers to choice expressed as a money-metric as:

$$\Delta w_i^* = CE_i^* - CE_i,$$

denoting the certainty equivalent for individual i's observed choice by CE_i and for the utility-maximizing choice by CE_i^*. For risk-neutral preferences, the difference in certainty equivalents corresponds to the potential cost savings from choosing the deductible that minimize one's expected out-of-pocket expenditures:

$$\Delta w_i^{*,\sigma=0} = CE_i^{*,\sigma=0} - CE_i^{\sigma=0}. \tag{2.4}$$

As discussed before, allowing for risk aversion makes only small differences to the value of different choices.[34]

Using the expected cost savings as measure of consumer welfare, we find that approximately 52% of consumers would have been better off with the 500 EUR voluntary deductible in 2015, but less than 7% of consumers took it. Of the population of the Netherlands, only 54.4% of individuals chose the cost-minimizing deductible. The average amount of money left on the table per individual is 66.2 EUR. While small in absolute value, these savings are roughly half of the total surplus at stake in the decision, which is 145 EUR on average.[35]

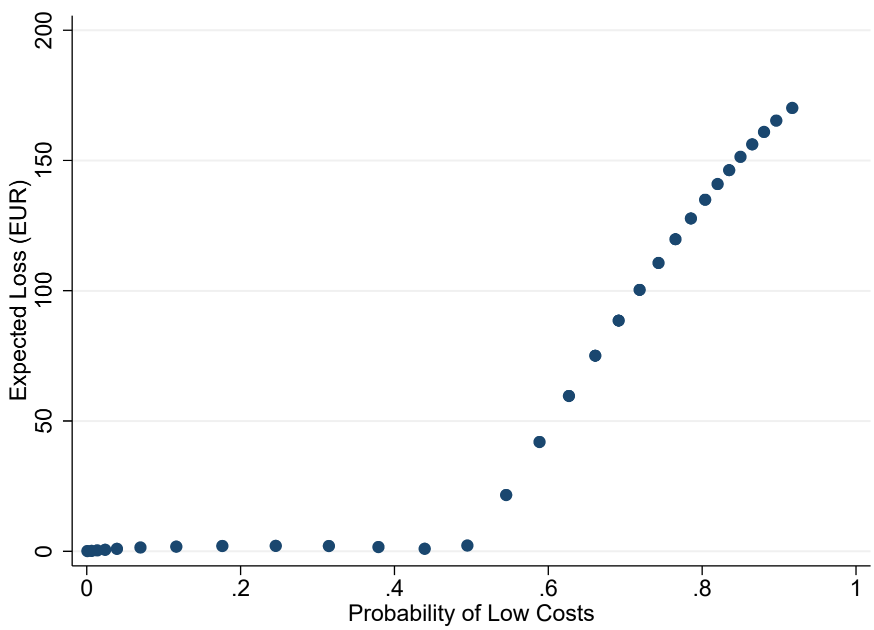
Choice Quality by Health Our individual measure of choice quality specifically conditions on an individual's predicted health. A first key dimension of heterogeneity to consider is thus how choice quality varies across individuals with different health. Figure 2.11 shows how the average cost savings vary with the predicted probability π in a bin-scatter plot. The overall costs savings combine

[34] Note also that we over-estimate the cost savings for those who do not take the 500 EUR deductible, but do take an intermediate deductible. However, we under-estimate the gain for those who do not take a voluntary deductible with predicted probability just below 50%.

[35] We define the stake as $|250 - (1 - \hat{\pi})500|$ EUR, which is at most 250 EUR and equal to 0 for individuals with $\pi = .5$.

the expected loss from over-insurance for low-cost individuals ($\pi \geq .5$) who do not take the extra deductible and from under-insurance for high-cost individuals ($\pi < .5$) who do. This graph is the result of the combination of the mechanical relationship between π and the potential cost savings, which are V-shaped around $\pi = 0$, as well as the actual distribution of choices made conditional on π. Very few individuals under-insure: most individuals with high predicted risk stick to coverage without extra deductible, as they should. On the other hand, relatively few individuals opt for the deductible when they should and the expected loss from over-insurance increases as the predicted risk is lower.

Figure 2.11: Expected Loss and Health Cost Probability



Notes: This figure is a binned scatterplot of the relationship between the predicted probability of health costs below 375 EUR and the expected loss due to over- or under-insurance. For individuals with a predicted probability of low costs below 0.5, the expected losses due to under-insurance are very small (on average close to zero), as a very low fraction of people under-insures by taking the 500 EUR extra deductible. For individuals with a predicted probability of low costs above 0.5, expected losses due to over-insurance increase with this probability, and reach almost 170 EUR for people with a very high chance (0.9+) of low costs, as most people leave money on the table by over-insuring for costs that happen with a very low probability.

Figure 2.11 thus demonstrates an important feature of our setting. Choice error is strongly correlated with health risk. In particular, those in the best health tend to leave the most money on the table. For the purposes of measuring welfare and equity this correlation will affect our results if an individual's health is also correlated with socio-economic background. Our earlier findings, however, demonstrate that socio-economic factors affect choice directly, even conditional on health, so there is ample opportunity for policy options that are mediated through choice to change welfare.

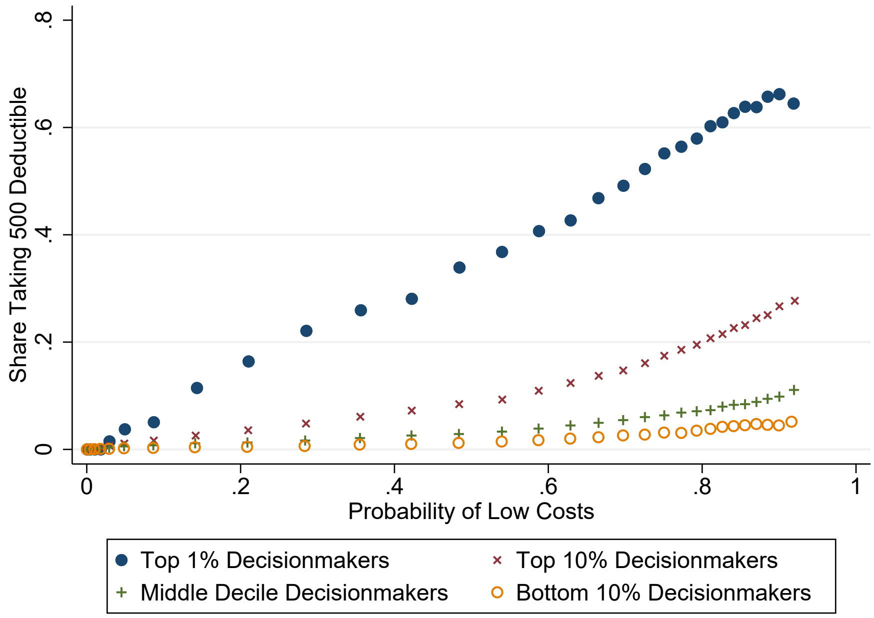
Heterogeneity Conditional on Health We now consider the heterogeneity in choice quality and which kinds of consumers are the best and worst choosers, conditional on health. To control for differences in health, we predict consumers' choices as a function of their underlying health risk π_{it} and observable characteristics X_{it}, allowing for interactions between the two, in a first step. We thus get predicted deductible choice probabilities $d(X_{it}, \pi_{it})$, which we then translate into consumer welfare $\Delta w^{*,\sigma=0}(X_{it}, \pi_{it})$ based on equation 2.4 in a second step. In a final step, we average the cost savings over the different health risks using the population distribution of predicted health risks, $\Delta w^{*,\sigma=0}_{\pi_{pop}}(X_{it})$. We then rank individuals from worst to best decision makers based on how much value they are predicted to leave on the table on average across a representative distribution of population health. We provide more detail on this procedure in Appendix VII.F.

We find significant heterogeneity in choice quality, even when controlling for differences in health risk. The very best decision makers (the top .1%) choose the cost-minimizing deductible 73% of times, conditional on some health risk drawn from the population distribution (see Panel A of Appendix Figure A2.2). The top 5% decision makers have a probability of 55% to make the right choice. All other decision makers are predicted to make worse choices than an individual choosing randomly. Figure 2.12 shows the responsiveness of deductible choices to health risk for different quantiles of choice quality. The performance of the very best decision makers is striking relative to the others. The take-up rate of the top 1% of decision makers is much steeper, coming close to the 45-degree line. The median quality decision-maker, on the other hand, essentially sticks to the compulsory deductible regardless of the underlying health risk.

Table 2.8 compares the observable characteristics for the best and worst decision makers and paints a telling picture of who is making the best choices in our context. The best decision-makers have an average gross income of 105K EUR and net worth of about 250K EUR. The worst decision makers, though, only have an average income of 40k EUR and net worth of 5K EUR. The massive difference in income and wealth are complemented with substantial differences in education. For example, those with college education are 3.48 times more likely to be in the best decision making group and with further education are even 15.57 more likely.[36] Individuals with quantitative degrees or occupations are

[36]Note that a zero value in the right panel of Table 2.8 does not mean that no single individual with the respective characteristic can be in that group. Instead, it means that given the predicted choices based on observable characteristics

Figure 2.12: Heterogeneity in Choice Quality



Notes: This figure illustrates dispersion in choice quality, by showing a binned scatter plot of the relationship between the predicted probability of having costs below 375 EUR (staying under the voluntary deductible range) and the take-up of the voluntary 500 EUR deductible for four selected subgroups that differ in their expected loss. The bottom 1% expected loss group comes close to a rational consumer, with high take-up of the deductible for low expected costs. The top 10% expected loss group has losses that are due almost entirely to over-insurance.

similarly over-represented in this top group. We also find that better decision-makers are significantly younger on average (36 year old vs. 63 in bottom 5 %), more likely to be male and more likely to have children. Finally, we also see that high quality decision makers are in peer settings where decision making quality is higher, both in terms of where they work and where they live. The average firm and postcode fixed effects decile for the top 5% decision makers is 6.41 and 6.07 respectively. The differences in parental take-up across the different groups are striking too.

V.B Peer Effects and Inequality Acceleration

Overall, our results show a strong socio-economic gradient in choice quality, with poorer and less-educated individuals being far more likely to make worse decisions when predictably healthy. As discussed in the introduction, our analysis confirms and deepens the findings in prior work documenting similar patterns of choice frictions linked to specific socio-economic characteristics. Furthermore, our

for all individuals, no individual with that specific characteristic (and his/her other respective characteristics) is predicted to end up in that group.

Table 2.8: Best and Worst Decision Makers

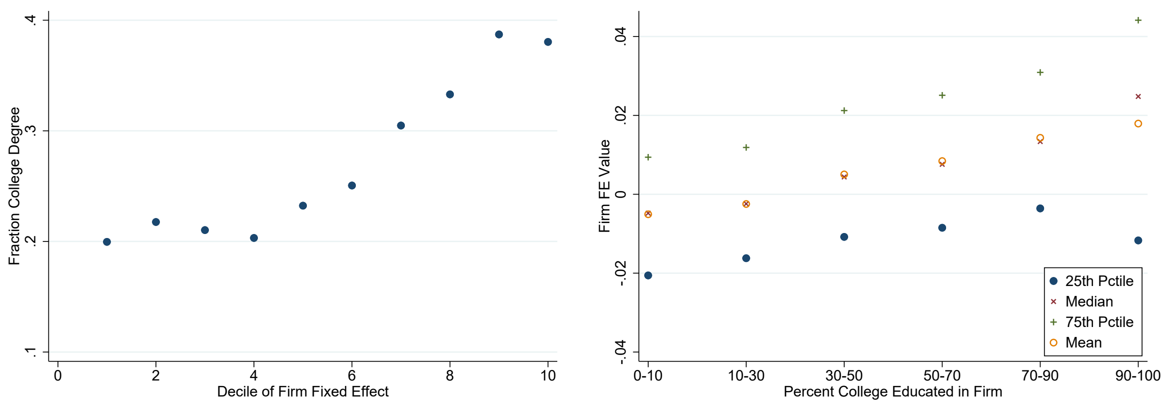
	Mean			Over/underrepresentation	
	Top 5% decisionmakers	*Bottom 5% decisionmakers*		*Top 5% decisionmakers*	*Bottom 5% decisionmakers*
Demographics			**Education level**		
Gender (male)	62%	28%	Less than high school	0.30	2.99
Age	36	63	High school	0.82	0.33
Has children	59%	34%	College	3.48	0.00
Has a partner	46%	90%	Further Studies	15.57	0.00
Financials			Unknown	0.08	1.05
Gross income	105,801	39,347	**Education field**		
Net worth	250,632	4,969	Statistics	19.66	0.00
Has Mortgage Debt	64%	19%	Philosophy	13.14	0.00
Has Other Debt	27%	53%	Economics	6.95	0.01
Has Savings >2000EUR	91%	38%	Tax and administration	3.30	0.01
Peer Effects			Marketing and advertising	1.91	0.06
Firm FE decile	6.41	4.09	Hair and beauty services	0.64	1.79
Postcode FE decile	6.07	5.47	Protection of persons	0.38	2.24
Mother With 500 Deductible	37%	0%	**Work Status**		
Father With 500 Deductible	45%	0%	Student	2.80	0.16
			Retired	0.07	2.47
			Self-employed	2.07	0.05
			Employee	1.16	0.31
			On Benefits	0.32	1.94
			Professional sector		
			Business services	2.77	0.09
			Insurance	2.13	0.07
			Retail	1.10	0.34
			Construction	0.75	0.24
			Cleaning	0.26	1.40
			Public utilities	1.51	0.11
Observations					11,369,800

Notes: This table presents observable characteristics for the groups that our model considers to be the top 5% and the bottom 5% decision makers. The entries in the left panel give the average value of the variable in each group. The entries in the right panel give the ratio of the proportion of consumers with that characteristic in each group relative to the proportion of consumers with that characteristic in the population overall. For example, the group of best decision makers has 6.95 time more economics majors, proportionally, than the population overall.

empirical analysis uncovered other dimensions of heterogeneity underlying choice quality and identified the importance of peer effects in particular. Individuals with different socio-economic background, however, may be exposed to different peers. We therefore turn to the question: can peer effects accelerate the socio-economic inequality?

Figure 2.13 relates the firm peer fixed effects estimated in the prior section to the education level of employees. The left panel presents the fraction of employees who are college educated as a function of the decile of the firm peer fixed effect. Firms across the five lowest fixed effect deciles have a similar percentage of college educated (approximately 20%). After that, there is a strong positive relationship between estimated firm peer effects and education level. The percentage of those who are college educated rises steadily from about 25% in the sixth decile to 40% in the top two deciles.

FIGURE 2.13: FIRM PEER FIXED EFFECTS AND EDUCATION

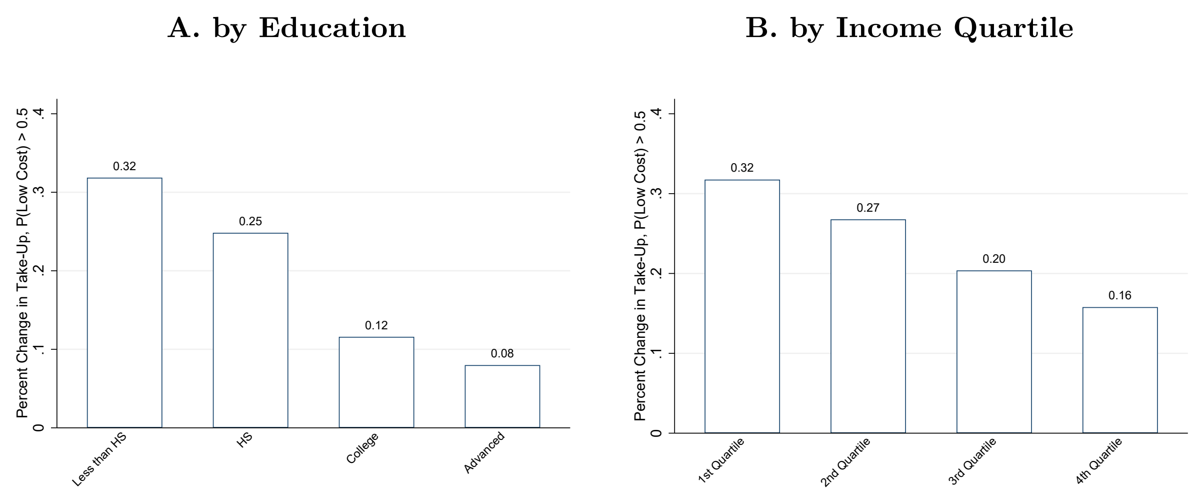


Notes: The left panel in this figure shows the fraction of employees with a college degree as a function of the firm peer fixed effect decile. The right panel shows distributional statistics for firm peer fixed effects as a function of the percent of employees at a firm who have a college degree.

The right panel of the figure looks at the same association from a different angle: what does the distribution of firm fixed effects look like for a given level of percent college educated within a firm? The figure shows that the mean, median, and 75th percentile of firm peer fixed effects increase monotonically as the percent college educated in a firm increases. For example, the 75th percentile of the firm peer fixed effect jump from .01 for firms with a low proportion of college educated ($< 20\%$) to .04 for firms with a high proportion of college education ($> 90\%$). This gap of .03 equates to roughly 33% of the proportion of consumers who choose a high deductible overall, suggesting a meaningful impact of the firm peer effect / education gradient relative to baseline choices. There are similar effects for the median and mean of the firm peer fixed effect distribution conditional on percent college educated. Also, though we focus this discussion on firms, results are similar in spirit for estimated neighborhood peer effects.

These results suggest that peer effects accentuate inequality in opting out of the default low deductible when an individual is predictably healthy. We quantify this impact in a counterfactual exercise where we take our firm and neighborhood peer effects estimates and, holding all else equal for an individual, assume that every individual experiences firm and neighborhood peer fixed effects equal to the average of the top decile in each domain. Figure 2.14 shows the results of this exercise on high deductible take-up rates for predictably healthy consumers who have at least a 75% chance of very low spending (< 375 EUR). It shows this impact as a function of (i) education level and (ii) income. Equating peer effects across individuals increases % take-up of high-deductibles for the healthy high-school drop outs by 35% (2.2pp) and increases that take-up for those with high-school, college, and advanced degrees by 27% (2.2pp), 13% (2.0pp), and 9% (1.8pp) respectively. There is a similar, though more muted, gradient with respect to income ranging from a 32% (2.3pp) increase in take-up for those in the lowest income quartile to a 18%(2.1pp) increase for those in the highest income quartile.[37]

Figure 2.14: Peer Effects and Inequality Acceleration



Notes: This figure presents the % effects of a counterfactual exercise that equalizes firm and neighbor peer fixed effects for the entire population at a value equal to the mean of the top decile for effects in each domain. The figure shows results for predictably health consumers, defined as those with a greater than 75% of spending lower than 375 EUR as estimated in ML-based cost prediction model. It shows the % effects on consumers as a function of education level in Panel A and income quartile in Panel B.

V.C Counterfactual Policies

This last part studies the welfare impacts of counterfactual choice policies and aims to gauge the potential to improve consumer welfare in our specific context. First, we consider how much better off consumers would be if everyone were allocated to the best option for them *ex ante* (according to our estimates and welfare model). This is useful as a first-best benchmark given the current choice architec-

[37]We also perform this analysis for (i) a more inclusive sample of those with at least a 50% chance of having low spending and (ii) moving everyone to the average of the 9th decile of peer effects, rather than the average of the 10th. The former leads to very similar results to those presented here while the latter leads to similar directional results, though, of course, the effects are somewhat muted due to the lower magnitude of peer effects achieved.

ture. It is also a measure of the impact of policy interventions that improve consumer decision-making or use predictive models to establish "smart defaults" (Handel and Kolstad (2015*a*), Gruber et al. (2020)). Next, we consider the impact of two alternative policies that limit choice; one that offers only the high deductible option and one where only the low deductible option is offered. These policies are clearly feasible and also reflect the underlying trade-off between offering greater choice and exacerbating choice errors. By accounting for how the incidence of choice frictions falls on individuals with different observable characteristics, we explore not only the efficiency, but also the equity implications of the different policy options.

In assessing the efficiency implications for each policy — the surplus generated by the plans chosen — we allow for four different values of risk aversion (assumed to be homogeneous in each implementation) including (i) risk neutral (ii) CARA = 10^{-5} (iii) CARA = 10^{-4} and (iv) CARA = 10^{-3}. To assess the equity implications we rely on income as the measure of inequality and consider alternative welfare weights for deciles of the income distribution. Following Atkinson (1970), the welfare of an individual in income decile y_δ is weighted by $y_\delta^{-\epsilon}/\left(\Sigma y_\delta^{-\epsilon}/10\right)$ for $\epsilon = .5$ and $\epsilon = 1.5$.[38] In our primary analysis, we rely on the observed correlations between health and socio-demographic status in the data. In the appendix, we also perform an analysis that assumes identical health distributions conditional on non-health X_{it}, using the predicted choice probabilities $d(X_{it}, \pi_{it})$ as in subsection V.A.

Table 2.9 presents the average welfare impact per person (in EUR) for the three different policies we consider. Column 1 presents the results for the scenario where individuals are allocated to their *ex ante* optimal deductible choice in the current environment. The average consumer welfare gain, when not weighted for inequality, is 68.8 EUR for risk neutral individuals. This gain decreases only slightly when introducing reasonable levels of risk aversion and is still 58 EUR for individuals assuming our highest level of risk aversion. When we weight for equity as a function of income the gain of the *ex ante* optimal allocation is reduced. With high inequality aversion the average benefit of this policy is 37.4 EUR for a risk neutral consumer. The decline results from the fact that lower income individuals are less likely to be healthy and, thus, more likely to have the default option of a low deductible be the correct choice for them. Because most choice errors result from not actively choosing the higher

[38]The Atkinson index of inequality uses a social welfare function of the form $y_i^{1-\epsilon}$ with $\epsilon \geq 0$ a measure of inequality aversion. Here, we weigh the welfare gain for each individual depending on income decile they are in by $y_\delta^{-\epsilon}/\left(\Sigma y_\delta^{-\epsilon}\right)/10\big)$, which ensures comparability with the unweighted case. We could model equity concerns more broadly by differentially weighting outcomes for individuals as a function of their predicted health π_i and characteristics X_i.

Table 2.9: Welfare Impact of Alternative Policies

	Optimal Deductible	High Deductible Only (875 EUR)	Low Deductible Only (375 EUR)
Risk Neutral			
Unweighted	68.8	-26.2	-8.3
Low Inequality Aversion	56.9	-64.4	-6.3
High Inequality Aversion	37.4	-133.6	-3.4
σ=.0001			
Unweighted	67.8	-28.1	-8.2
Low Inequality Aversion	56.0	-66.1	-6.2
High Inequality Aversion	36.8	-135.1	-3.3
σ=.001			
Unweighted	58.0	-44.6	-7.0
Low Inequality Aversion	47.7	-81.6	-5.3
High Inequality Aversion	30.9	-148.7	-2.7

Notes: This table shows the average welfare impact (in EUR per person) of three alternative policies concerning the extra deductible: optimal deductible (all individuals taking the optimal deductible given their health risk), high deductible only (only the 500 EUR extra deductible is available), and low deductible only (the low deductible is the only option). The welfare impact is calculated with equal weights for all income deciles, low inequality aversion or high inequality aversion. Weights y_δ are computed as $y_\delta^{-\epsilon}/\left(\Sigma y_\delta^{-\epsilon}/10\right)$ for $\epsilon = .5$ and $\epsilon = 1.5$. The welfare impact is calculated not controlling for health. The corresponding welfare impact when assigning each individual the population health distribution is in Appendix Table A2.1. Our sample contains the choices of $9,415,666$ individuals in 2015 (out of $11,991,629$ individuals for which the probability of low costs and the deductible choice are both non missing), excluding students, self-employed people, individuals with a gross income below the social assistance level and individuals with missing observables.

deductible, there is less to be gained if many low income enrollees are better off in the low deductible plan. Appendix Table A2.1 shows how this relationship is reversed when controlling for differences in health, reflecting the higher incidence of choice frictions among low-income individuals.

Columns 2 and 3 show the consumer welfare impacts when consumers are offered only the high deductible (with the corresponding premium reduction) or the low deductible, respectively. Neither policy that limits the choice offerings is welfare-increasing, even relative to the status quo where consumers are making poor choices in general. Mandating the extra 500 EUR deductible leads to a welfare losses with no inequality aversion of 22.6 EUR when risk neutral and 44.6 EUR with high risk aversion. With high inequality aversion, however, this policy is much worse, with welfare losses of 133.6 EUR when risk neutral and 148.7 EUR with high risk aversion. This policy is especially bad because it is forcing sick, lower income consumers into what would have been the wrong choice for them. Mandating a low deductible, on the other hand, has a much smaller impact due to the fact that, in practice, most people already choose that deductible. The small impact ranges between 0 and 10 EUR on average across the range of risk aversion and inequality aversion parameters we investigate.

Discussion Our counterfactual analysis allows to draw some important conclusions for choice-based policies more generally and for the specific implementation in the Netherlands, using a low baseline deductible with the option to take a higher deductible. While a policy that is able to move people to plans based on *ex ante* risk could substantially increase welfare, the welfare gain from the offered deductible choice is small. Moreover, due to both the correlation between income and health and the correlation between income and choice quality, accounting for higher inequality aversion actually reduces the welfare loss of this policy. The option to select a higher deductible increases welfare mostly for the high-income individuals, who are healthier and make better choices. The value of this option is very limited for low-income individuals and may well become negative when factoring in equilibrium price changes.

Importantly, our analysis has ignored any direct welfare effects of choice frictions beyond the misallocation to plans. In our setting we do not have good measures of potential costs associated with decision making. If making a decision imposes a cost on enrollees — as has been shown in a number of other settings (see e.g. Handel and Kolstad (2015*b*) in health insurance) — these costs may exceed the relatively small gains we find from offering the option to take a higher deductible. Our analysis has also been limited to consumer welfare without accounting for the potential implications of moral hazard and adverse selection. In the presence of moral hazard, the reduction in health expenditures in response to an extra deductible could benefit the insurer as discussed in Section III, but we also presented evidence of limited moral hazard with respect to the deductible policy. In the presence of adverse selection, we also expect equilibrium prices to respond to the regulation of choice, which would further affect sorting and consumer welfare. In particular, the option to buy less comprehensive coverage allows individuals with good health to contribute less to the health insurance system. We have ignored the pricing repercussions this may have.[39]

[39] By removing choice frictions, we may expect adverse selection to become worse (e.g., Handel, Kolstad and Spinnewijn (2019)). Interestingly, comparing the average predicted low-cost probability for workers taking the extra deductible (.763) and for those who should take the extra deductible (.760) suggests that in this context the pricing repercussions from reducing choice frictions may be limited.

VI Conclusion

Many policy makers rely on market-based solutions to supply products, from health insurance managed competition to private retirement benefits and beyond. The rationale for these approaches is that regulated, market-based provision of impure public goods can deliver greater product variety and improved efficiency, getting the returns we expect from a market while still accounting for the public nature of these goods and services. An important potential limitation to the effectiveness of these policies is the ability of consumers to choose between the available options and maximize their surplus. Ineffective decision-making and/or ineffective choice architecture undermines the gains from such policy approaches and can, in principle, be large enough to entirely undercut reliance on market-based provision for these products.

Using granular data from the Netherlands, we characterized nationwide quality in deductible choices and found that (i) these choices were poor on average, in line with prior work on default options, and (ii) higher SES consumers make better choices than lower SES consumers, with a meaningful impact on realized surplus. Most notably, highly educated individuals who have more quantitative training make better choices than their counterparts, holding constant other key factors like income, net worth, and health risk. In addition, we use a causal movers-design, we find that peer and environment effects from the workplace, neighborhood, and family are important determinants of choice quality. A variety of other socio-economic factors have more limited impacts on choice quality, including household income, household net worth and household liquidity. We show that peer effects accelerate inequality in the sense that more positively influential peer effects are correlated with higher education and income levels. Finally, we investigate the efficiency and equity implications of several counterfactual regulatory scenarios related to (i) smart defaults and (ii) menu design. While smart defaults have the potential to unlock significant surplus, simple menu design scenarios like choice-set simplification generally reduce surplus and impact equity negatively if more generous deductible options are the ones removed from the choice set.

Given the policy importance of our results, both for choice quality overall and for the choice quality - SES gradient, we believe that there are several fruitful directions for future research. At a micro level, it will be valuable to assess how different policy and technology solutions can improve choices in different market and regulatory environments, both overall and for lower SES consumers specifically.

For example, a field experiment at scale (e.g., Banerjee et al. (2019)) distinguishing between distinct behavioral foundations and/or distinct behaviorally-motivated policies (e.g., Bhargava, Loewenstein and Sydnor (2017)) could provide valuable additional insights, especially if linked to data similar to what we use in this study. For example, while Brot-Goldberg et al. (2021) show that default effects for Medicare Part D low-income enrollees are primarily due to inattention rather than switching costs, it is unclear whether the better choices we document for higher-SES consumers are due to increased attention, relative to lower-SES consumers, or due to better active decisions once paying attention. If higher SES consumers are more attentive but not much more sophisticated otherwise, this has important implications for the welfare impacts of policies and on our understanding of the potential for insurance markets to deliver value.

In addition to understanding these underlying mechanisms, it is important to explore policy options that account for the distributional consequences of decision-making issues. For example, one could design the choice menu to combat the regressive nature of choice quality by matching the default option closer to the typical low SES consumer than to the typical high SES consumer. Targeted defaults as a function of key consumer characteristics, as discussed in Handel and Kolstad (2015*a*) and Abaluck and Gruber (2016*a*), are another interesting path forward from a policy design standpoint. Finally, while the evidence for the importance of choice frictions and their unequal incidence in the population seems strong in our context, it will be valuable in future work to study the trade-offs between potentially regressive choice quality and the efficiency gains from competing insurers, e.g. via improved products, lower premiums, or improved health outcomes.

VII Appendix

VII.A Data Appendix

This Data Appendix provides information on the additional datasets we linked to our health cost and insurance data at Statistics Netherlands. Datasets are linked at the individual level based on anonymized individual identifiers.

Age and gender Dataset *Gbapersoontab* provides an overview of all people registered living in the Netherlands at any point since 1995. These registers form a basis for the administrative records of all individuals in the Netherlands. For our purposes, *Gbapersoontab* is used to obtain age and gender, and we use this person registry as the primary dataset to match all other datasets with.

Family and household links Family links come from the dataset *Kindoudertab*, which contains all known legal child-parent links. Household identifiers as well as family status variables in *Ipi* and *Inpatab* allow us to identify partners and other household links. Partnerships consist of all partners who are living in the same household.[40]

Education *Hoogsteopltab* is a dataset that includes the highest attained educational course for each individual, and originates from several educational registers and survey data. We link each educational course to its relevant International Standard Classification of Education (ISCED) level and field of education. There is almost universal coverage for the youngest cohorts, but educational information is missing for many individuals aged over 40. Overall, we observe highest education obtained for 54.6% of our full sample.

Income and Employment Status Datasets *Ihi* and *Inhatab* contain information on households' income, and originates from tax authorities. Our main definition of income used in the analysis is household gross income (called *bruto inkomen* by Statistics Netherlands). Gross income includes all labor income and capital income, as well as government transfers (e.g., UI, DI, pensions), and other transfers and income. We also use a socio-economic classification variable *seccoal1*, which classifies each individual based on where the majority of his or her personal income comes from. This variable is obtained from datasets *Ipi* and *Inpatab*.

Wealth Dataset *Vehtab* contains information from tax authorities on households' assets and debts.

[40] This includes married partners, registered partners, but also partners who have not registered their partnership but are living in the same household.

This information is partly self-reported (on tax forms) and third-party reported. Assets include financial assets (savings, stocks, bonds, and other participations), real estate and other assets (such as cash and movable assets). Debts include mortgages, study debt and other debt. The net wealth variable in the main text equals household assets minus household debts.

Employee-Employer links We use the dataset *Spolisbus* to link individuals to their firms, colleagues and sector. *Spolisbus* is a highly detailed dataset with monthly information on all employment contracts in the Netherlands, collected by the tax authorities based on third-party reported data. We adopt the same definition of a firm as in the firm registry (*Algemeen Bedrijfsregister*) of Statistics Netherlands. We sum each individual's total hours worked by year by firm. For each individual, we then select the firm at which that individual has worked the most hours in each year. The colleagues that we identify are thus all individuals who work the majority of their hours at the same firm. The sector categorization that we adopt is made by the authorities based on the collective labor agreements.

Location We match every individual with their yearly 6-digit postcode based on their registered residence. For this, we use datasets *Gbaadresobjectbus* and *Vslgwbtab*. Postcodes are obtained for each year on 1 October, as this is close to the period of deciding on their health insurance contract. 6-digit postcode information is at a neighbourhood level, and there are 12'116 distinct postcodes in 2015.

VII.B Health Cost Predictions

In this Appendix, we describe the binary prediction algorithm that we use to obtain risk probabilities, and discuss its accuracy across different subgroups, and the most important predictors. We also discuss an alternative non-binary prediction algorithm and argue why the binary predictions are preferable for the analysis in this paper.

Prediction Algorithm

We use an ensemble machine learning algorithm to predict the probability that an individual's health costs will not exceed the mandatory deductible of 375 EUR in any given year. The prediction algorithm we use is a standard machine learning method for binary classification, an ensemble learner that consists in our case of a random forest model, gradient boosted regression trees and LASSO model. To avoid overfitting, we train and calibrate the prediction algorithm on a training sample of 1.25 million individuals. We then use this trained prediction algorithm to obtain predictions for a hold-out sample of about 12 million individuals. All the analyses and statistics in the paper are developed use only this hold-out sample.

The prediction method we use follows four steps, which closely resemble the steps used in Einav et al. (2018). First, we follow standard practice in machine learning by tuning key parameters that govern the prediction models by 3-fold cross-validation. Second, we train the three resulting prediction models separately. Third, we combine the three obtained predictions into one using a linear combination that we calibrate in the data. Finally, we calibrate the resulting final ensemble predictions using a linear spline. As there is some variation in the number and definition of predictors that we have across time, we repeat these four steps for all years of study (2013-2017). We describe each of the four steps in more detail here.

Parameter Tuning As the three machine learning models that we use have parameters that are choosable by the researcher, we follow standard practice and tune these parameters using 3-fold cross validation. More specifically, we tune the following parameters using 100,000 observations: minimal node size (mid.node.size), number of variables used at each node (mtry) for the random forest model, learning rate (eta) for the boosted regression trees, and the shrinkage parameter (lambda) for the LASSO.[41] For each of these parameters, we optimize among 5 alternatives. We tune these parameters using 3-fold cross validation, where we are optimizing the area under the receiver operating characteristic curve (AUC).[42] Thus, for each of the parameter values we want to test, the model is trained on 2 folds (subsets of the training sample), and then the performance is measured in the 3rd fold. The parameter values for which the AUC in the hold-out sample is highest for each prediction algorithm are: mtry = 10, min.node.size = 10, eta = 0.2, lambda = 0.0001.

Estimating the Models Using these tuned parameter values, all models are estimated using a training sample of 800,000 individuals.

Obtaining Ensemble Predictor We combine the predictions from the random forest, gradient boosting regression trees, and LASSO into one ensemble prediction. Following Einav et al. (2018), we

[41] We use the package CARET in R that provides a standardized way to tune parameters. The prediction models we use are RANGER (random forest), XGBLINEAR (boosted regression trees), and GLMNET (LASSO).

[42] This is a common metric used in the machine learning literature to measure the performance of a prediction model.

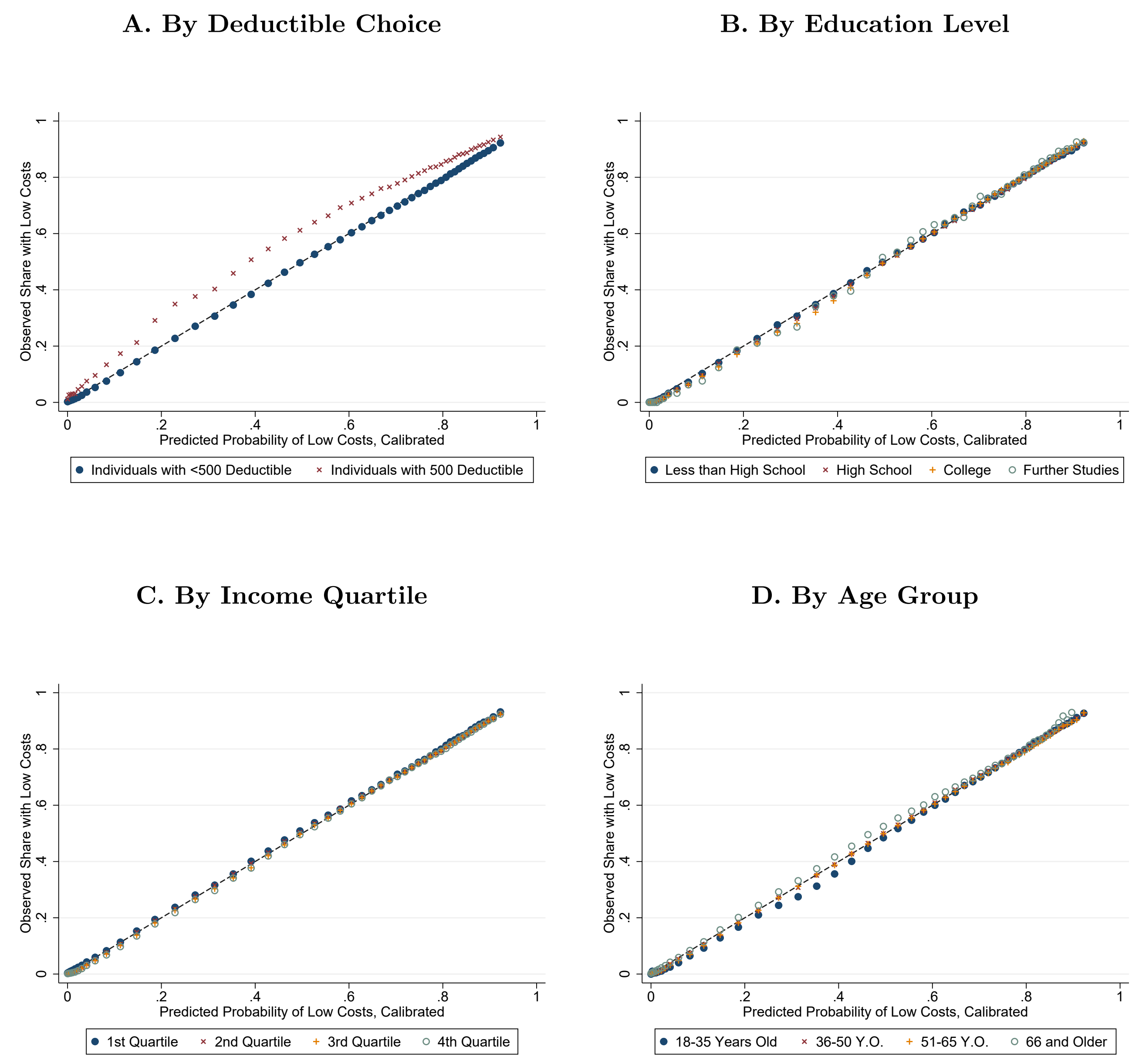
construct the ensemble prediction to be the linear combination $p_{ensemble} = \hat{\beta}_{rf}\hat{p}_{rf} + \hat{\beta}_{gb}\hat{p}_{gb} + \hat{\beta}_{lasso}\hat{p}_{lasso}$, where $\hat{p}_x$ is the prediction from algorithm x and $\hat{\beta}_x$ is the associated weight.

We obtain estimates for the weights from a constrained linear regression (with no constant and the weights summing to one) of the dummy for having costs below 375 EUR on the three individual predicted probabilities. For this step, we use 100,000 observations that we did not use in either step of parameter tuning nor the estimation of the models. We find associated weights in 2015 that are $\hat{\beta}_{rf} = 0.67$, $\hat{\beta}_{gb} = 0.08$ and $\hat{\beta}_{lasso} = 0.25$.

Calibrating Probabilities Finally, the raw probability predictions we get from the ensemble step are calibrated to the actual observed probabilities by estimating a linear spline. This calibration is done using 350,000 observations that are used in none of the previous steps. 10 equal sized bins are created based on the ranked predicted probability. In every bin the mean probability is calibrated to the observed mean probability for these observations. The piece-wise linear spline that follows from linearly interpolating all intermediary points serves as the last step in the prediction mechanism.

Additional Discussion of Prediction Model

Figure A2.1: Predicted vs. Observed Share of Low Costs, by Subgroups



Notes: This figure shows the calibration plot of the predicted probability of low costs for various subgroups of the sample. Panel A plots our prediction against the observed share of people with health costs below 375 EUR, separately for people having chosen the 500 deductible and people who have not. Panel B does the same exercise splitting the sample by education level. In Panel C, the sample in split by income quartile, and in Panel D, by age group.

While Figure 2.3 shows a calibration plot for the entire sample, Figure A2.1 shows a calibration plot for certain subgroups of the sample. We see from Panel B, C and D that probabilities are well calibrated for distinct groups of education level, income quartile and age group. This makes us comfortable that the observed differences in choice quality across these different groups are not due to differential prediction accuracy of our ensemble predictor.

Moreover, panel A of Figure A2.1 shows that individuals who choose a 500 EUR deductible are more likely to have low costs than individuals who choose no extra deductible, conditional on the prediction of our model. However, the difference in *ex post* realized low cost fraction is small, leading us to conclude that the private information and moral hazard, conditional on our predictors, is small. More specifically, the average gap across probability bins between individuals who choose and who do not choose an extra deductible is 6.667%. Taking into account that across probability bins, the average share with low costs among people without extra deductible is 51.215%, we find that individuals who take a deductible are on average are 13.017% more likely to have low costs than our model predicts. Importantly, as discussed in Section III.A, to the extent that consumers spend less under a high deductible plan because of classical moral hazard, our model threshold for choosing the high deductible ($\pi = 0.5$ for risk neutral, $\pi = 0.56$ for very risk averse) is slightly high (i.e. more people should choose the high deductible) and the normative benefits from doing so in Section V are too low, working against our main results. Relatedly, Table A2.1 supports the discussion of behavioral hazard in Section III.A, suggesting that up front rational avoidance of ex post behavioral hazard is not a major concern.

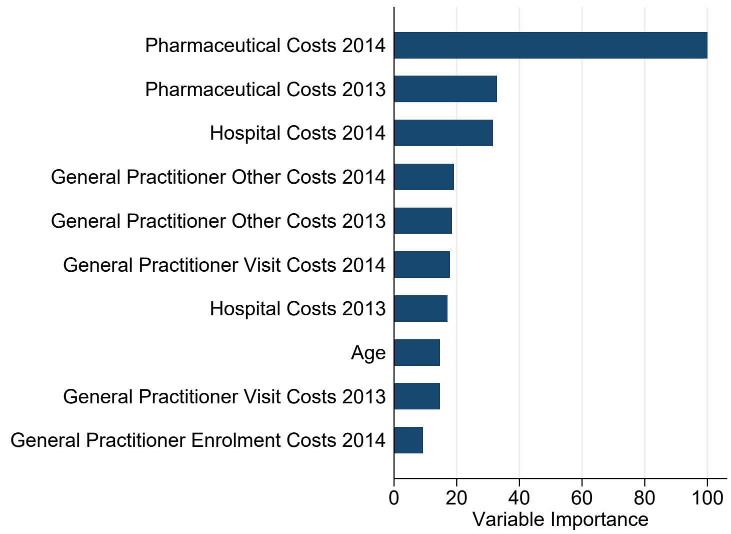
Table A2.1: Ex Post Health Expenses, by Subgroups

P(Low Costs)	Low Deductible	Any Incremental Deductible
N (Sample Size)		
0.6-0.7	1,156,446	91,263
0.7-0.8	1,514,402	171,016
0.8-0.9	1,850,417	298,369
0.9-1	471,746	96,877
Preventative Care (Always Insured)		
0.6-0.7	184.6	171.7
0.7-0.8	154.3	142.3
0.8-0.9	122.9	113.5
0.9-1	97.3	90.7
Drugs		
0.6-0.7	68.7	55.5
0.7-0.8	45.6	35.7
0.8-0.9	25.6	19.1
0.9-1	13.0	9.6
Maternity Care		
0.6-0.7	41.8	42.1
0.7-0.8	27.8	26.0
0.8-0.9	14.4	11.2
0.9-1	4.6	2.7
Mental Health		
0.6-0.7	234.3	173.2
0.7-0.8	155.5	117.0
0.8-0.9	98.0	66.1
0.9-1	64.9	38.0

Notes: This table presents statistics related to actual ex post spending on certain types of health care as a function of our ex ante prediction of the probability an individual has low costs. The top section gives the sample size for each group and subsequent sections give the mean EUR spent on each kind of care by individuals in each group. This table supports the discussion of behavioral hazard in Section III.A, suggesting that up front rational avoidance of ex post behavioral hazard is not a major concern.

Figure A2.2 presents the importance of different predictors in the random forest model, which is the model with the highest weight in our ensemble prediction. Not surprisingly, the most important predictors are different categories of past pharmaceutical spending, with $t-1$ values being more important than $t-2$ values. Hospital costs, costs to primary care visits and age are other important variables in the random forest prediction.

Figure A2.2: Variable Importance in Prediction with Random Forest



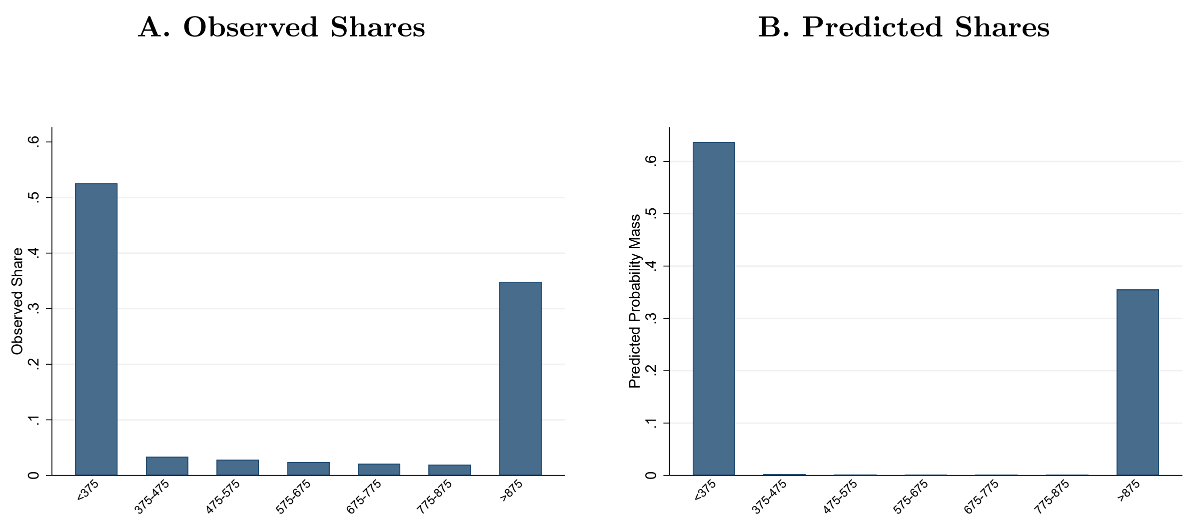
Notes: This figure shows the importance of selected variables in the prediction of health cost risk using only a random forest model. Variable importance is measured by the mean decrease in gini, ie. the average of a variable's total decrease in node impurity, weighted by the proportion of samples reaching that node in each individual decision tree in the random forest.

Non-Binary Prediction

In Section III.B, we simplified the deductible choice problem in the Netherlands into a binary choice between selecting a 875 EUR deductible, or the mandatory 375 EUR deductible. This is a simplification, as in fact there are 6 different deductible choices possible, which apply to different brackets with cutoffs at 375, 475, 575, 675, 775 and 875 EUR. However, two pieces of evidence show that reducing the problem to a binary one is appropriate for our context.

First, Panel A of Figure A2.3 shows that the ex-post observed shares within each intermediary deductible bracket are small. This means that only a small fraction of individuals fall into the intermediary deductible ranges, which decreases the likelihood that the intermediary deductibles are optimal choices. Second, we find that when using a machine learning classifier to predict which individuals are going to fall into the intermediary brackets, the predicted mass in these intermediary brackets is small. Panel B of Figure A2.3 shows that ex-ante, a random forest model trained on an unbalanced sample will give less than 1% probability mass to the intermediate categories. This is largely due to the unbalanced classes, where the majority of individuals fall into the lowest or highest bracket. However, insofar as we cannot expect individuals to predict their future costs more accurately, the low probability with which most individuals are predicted to be in the intermediary deductible brackets further strengthens the case for a binary decision rule.

Figure A2.3: Cost Predictions with Multiple Deductible Categories



Notes: Panel A plots the observed share of individuals with health costs in all the deductible health cost brackets in 2015. Panel B plots the predicted shares of individuals in all deductible health cost brackets, where the prediction is from a random forest with the same predictors as described in Section III.B.

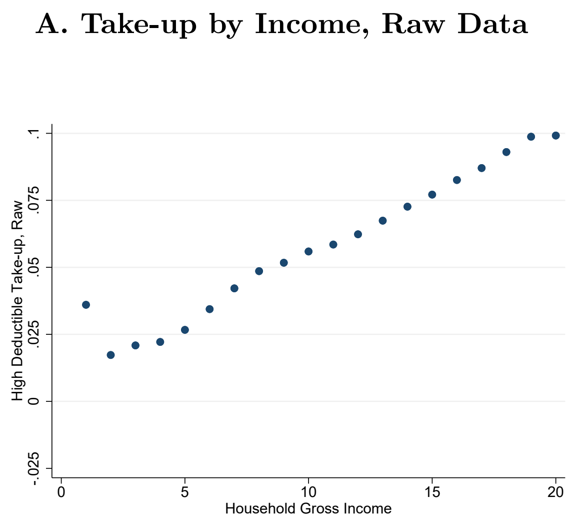
VII.C Deductible Choice: Appendix Figures and Tables

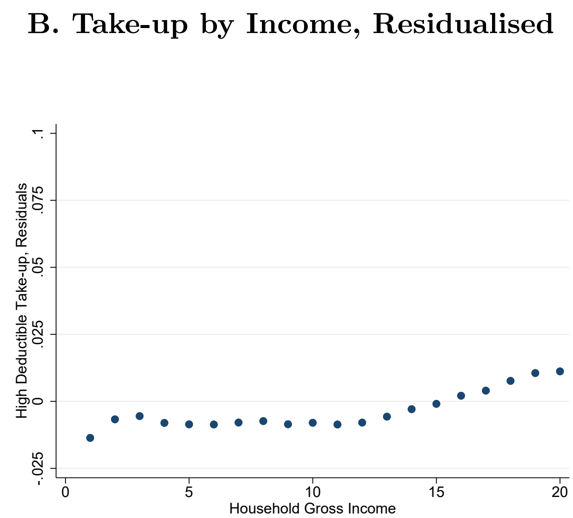
Table A2.1: Deductible Take-up: Impact of Health and Income Changes

	(1) No FE	(2) Individual FE	(3) First difference	(4) First difference	(5) First difference
Probability of Low Costs	0.115***	0.0570***	0.0422***		
Prob. Low Costs, Positive Δ				0.00691***	
Prob. Low Costs, Negative Δ				-0.0670***	
Δ Prob. Low Costs > +2 Deciles					0.0102***
Δ Prob. Low Costs = +2 Deciles					0.00685***
Δ Prob. Low Costs = +1 Decile					0.00342***
Δ Prob. Low Costs = -1 Decile					-0.00277***
Δ Prob. Low Costs = -2 Deciles					-0.00636***
Δ Prob. Low Costs < -2 Deciles					-0.0202***
Income ('000 EUR)	6.06e-05***	1.57e-05***	6.63e-06***	6.65e-06***	6.85e-06***
Number of Individuals	12,317,248	12,317,248	12,074,444	12,058,624	12,074,444
Observations	47,685,794	47,685,794	35,368,540	35,216,196	35,368,540

Notes: This table presents the result of an OLS regression of take-up of the 500 EUR extra deductible on probability of low costs, changes in probability of low costs, income, and changes in income. In column (1), take-up of the high deductible is regressed on the probability to have health costs lower than 375 EUR, and on income in thousands of EUR. Column (2) adds individual fixed effects. Column (3) regresses the first difference of deductible take-up on the first difference of the probability of low costs and the first difference of income. Column (4) splits the first difference in two distinct variables, one containing only positive shocks, the other only negative shocks. Column (5) creates six dummies capturing shocks of various magnitudes: positive and negative shocks of one, two, and strictly more than two deciles. In Columns (4) and (5), income first difference remains unchanged compared to Column (3). All regressions include year fixed effects. *** p<0.01, ** p<0.05, * p<0.1 with robust standard errors.

Figure A2.1: Deductible Take-up as a Function of Income





Notes: These figures plot the relationship between household gross income and the take-up of the 500 EUR extra deductible. Panel A plots take-up of 500 deductible by household income percentile. Panel B plots the residuals of an OLS regression of take-up of 500 EUR extra deductible on risk probability, four levels of education dummies, four age dummies, and indicators for gender, having a partner, and having children.

Table A2.2: Robustness Check

	Baseline			500 vs. 0 Deductible			0 vs. >0 Deductible		
	Without	With Interaction		Without	With Interaction		Without	With Interaction	
	Interaction	*intercept*	*slope*	Interaction	*intercept*	*slope*	Interaction	*intercept*	*slope*
High School	0.017***	-0.011***	0.057***	0.018***	-0.012***	0.061***	0.025***	-0.014***	0.077***
College Degree	0.065***	-0.034***	0.165***	0.071***	-0.038***	0.181***	0.089***	-0.037***	0.210***
Further Studies	0.091***	-0.047***	0.226***	0.099***	-0.052***	0.250***	0.123***	-0.044***	0.275***
2nd Income Quartile	-0.003***	0.004***	-0.007***	-0.003***	0.004***	-0.007***	0.002***	0.009***	-0.005***
3rd Income Quartile	0.004***	0.004***	0.007***	0.005***	0.003***	0.009***	0.014***	0.011***	0.013***
4th Income Quartile	0.024***	0.002***	0.039***	0.026***	0.001***	0.045***	0.041***	0.015***	0.048***
36 to 50 years old	-0.011***	0.020***	-0.045***	-0.010***	0.022***	-0.046***	-0.006***	0.024***	-0.042***
51 to 65 years old	-0.004***	0.029***	-0.047***	-0.004***	0.030***	-0.048***	0.003***	0.036***	-0.045***
65+ years old	-0.001***	0.034***	-0.082***	0.000**	0.036***	-0.085***	0.007***	0.043***	-0.092***
Male	0.011***	-0.004***	0.025***	0.012***	-0.004***	0.028***	0.017***	-0.001***	0.030***
Has Partner	0.003***	-0.002***	0.013***	0.003***	-0.002***	0.014***	0.002***	-0.005***	0.018***
Has Children	-0.010***	0.004***	-0.028***	-0.011***	0.004***	-0.031***	-0.014***	0.004***	-0.035***
Self-employed	0.009***	-0.006***	0.026***	0.009***	-0.007***	0.028***	0.013***	0.000	0.023***
Constant	-0.042***	-0.041***		-0.045***	-0.043***		-0.055***	-0.044***	
Prob. Low Costs	0.122***		0.098***	0.129***		0.100***	0.169***		0.124***
Year and Insurer FE	YES	YES		YES	YES		YES	YES	
Observations	57,100,388	57,100,388		55,335,880	55,335,880		57,100,388	57,100,388	

	Baseline			Probit			Binary Pred. Low Costs		
	Without	With Interaction		Without	With Interaction		Without	With Interaction	
	Interaction	*intercept*	*slope*	Interaction	*intercept*	*slope*	Interaction	*intercept*	*slope*
High School	0.017***	-0.011***	0.057***	0.022***	0.006***	0.023***	0.019***	0.002***	0.032***
College Degree	0.065***	-0.034***	0.165***	0.051***	0.014***	0.051***	0.068***	0.013***	0.081***
Further Studies	0.091***	-0.047***	0.226***	0.063***	0.005**	0.081***	0.093***	0.019***	0.105***
2nd Income Quartile	-0.003***	0.004***	-0.007***	0.003***	0.030***	-0.040***	-0.001***	0.003***	-0.005***
3rd Income Quartile	0.004***	0.004***	0.007***	0.010***	0.041***	-0.044***	0.007***	0.008***	0.002***
4th Income Quartile	0.024***	0.002***	0.039***	0.024***	0.057***	-0.048***	0.027***	0.017***	0.016***
36 to 50 years old	-0.011***	0.020***	-0.045***	-0.008***	-0.007***	-0.001	-0.013***	-0.005***	-0.009***
51 to 65 years old	-0.004***	0.029***	-0.047***	0.000	0.001**	-0.002**	-0.012***	-0.005***	-0.006***
65+ years old	-0.001***	0.034***	-0.082***	-0.011***	-0.008***	-0.004***	-0.017***	-0.008***	-0.025***
Male	0.011***	-0.004***	0.025***	0.006***	0.007***	-0.002***	0.015***	0.001***	0.023***
Has Partner	0.003***	-0.002***	0.013***	0.005***	0.007***	-0.003***	0.003***	0.000***	0.008***
Has Children	-0.010***	0.004***	-0.028***	-0.009***	-0.000	-0.011***	-0.011***	-0.000	-0.020***
Self-employed	0.009***	-0.006***	0.026***	0.008***	0.018***	-0.013***	0.011***	0.005***	0.009***
Constant	-0.042***	-0.041***					-0.014***	-0.003***	
Prob. Low Costs	0.122***		0.098***	0.169***		0.191***			
Pred. Costs <375							0.062***		0.034***
Year and Insurer FE	YES	YES		YES	YES		YES	YES	
Observations	57,100,388	57,100,388		57,100,388	57,100,388		57,100,388	57,100,388	

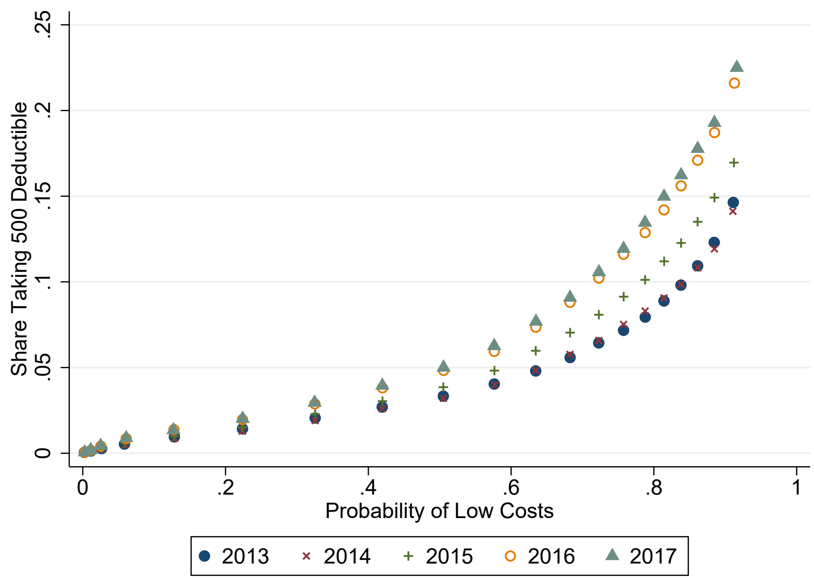
Notes: This table performs a range of robustness checks on our baseline results. In the top panel, we compare our baseline regression with alternative definition of take-up of the high deductible. In the baseline, we define take-up as choosing the 500 deductible, as opposed to choosing any other deductible. In the second top panel, we keep only choices that are the 500 or the 0 deductible, and drop intermediate choices. In the third top panel, we instead define take-up as choosing any deductible strictly greater than 0. In the second bottom panel, we compare our baseline OLS regression with a probit specification. Finally, in the third bottom panel, we replace our linear probability of low costs with a binary indicator taking value one if the individual is predicted to have health costs lower than 375 EUR. In each panel, we present a regression with and without interacting our regressors with the probability of low costs.

Table A2.3: Deductible Take-up Regression, Non Interacted

	(1) Baseline	(2) Education Field	(3) Professional Sector	(4) Liquidity and Financials	(5) Environment
High School	0.017***	0.016***	0.017***	0.015***	0.014***
College Degree	0.065***	0.062***	0.066***	0.062***	0.056***
Further Studies	0.091***	0.089***	0.097***	0.089***	0.088***
2nd Income Quartile	-0.003***	-0.008***	-0.011***	-0.009***	-0.006***
3rd Income Quartile	0.004***	0.000	-0.002***	-0.004***	-0.000
4th Income Quartile	0.024***	0.019***	0.017***	0.011***	0.014***
36 to 50 years old	-0.011***	-0.008***	-0.010***	-0.012***	0.007***
51 to 65 years old	-0.004***	-0.001***	-0.002***	-0.012***	0.027***
65+ years old	-0.001***	0.005***	0.003***	-0.016***	0.020***
Male	0.011***	0.015***	0.014***	0.012***	0.017***
Has Partner	0.003***	0.006***	0.004***	0.003***	0.008***
Has Children	-0.010***	-0.013***	-0.012***	-0.007***	-0.006***
Self-employed	0.009***	0.007***	0.008***	0.005***	0.007***
Statistics		0.139***			
Philosophy		0.024***			
Accounting and Taxation		0.012***			
Marketing and Advertising		-0.004***			
Hair and Beauty		-0.012***			
Protection of Persons		-0.033***			
Business Services			0.022***		
Insurance			0.027***		
Retail			-0.003***		
Construction			-0.013***		
Cleaning			-0.012***		
Public Utilities			0.001		
2nd Net Worth Quartile				0.004***	
3rd Net Worth Quartile				0.012***	
4th Net Worth Quartile				0.029***	
Has Savings > 2000EUR				0.008***	
Has Mortgage Debt				0.002***	
Has Other Debt				-0.009***	
Share of Colleagues with 500 Ded.					0.226***
Share in Postcode with 500 Ded.					0.404***
Father With 500 Deductible					0.181***
Mother With 500 Deductible					0.237***
Constant	-0.042***	-0.056***	-0.063***	-0.043***	-0.135***
Prob. Low Costs	0.122***	0.145***	0.148***	0.119***	0.160***
Year and Insurer FE	YES	YES	YES	YES	YES
Observations	57,100,388	30,799,129	32,299,835	57,013,765	16,938,401

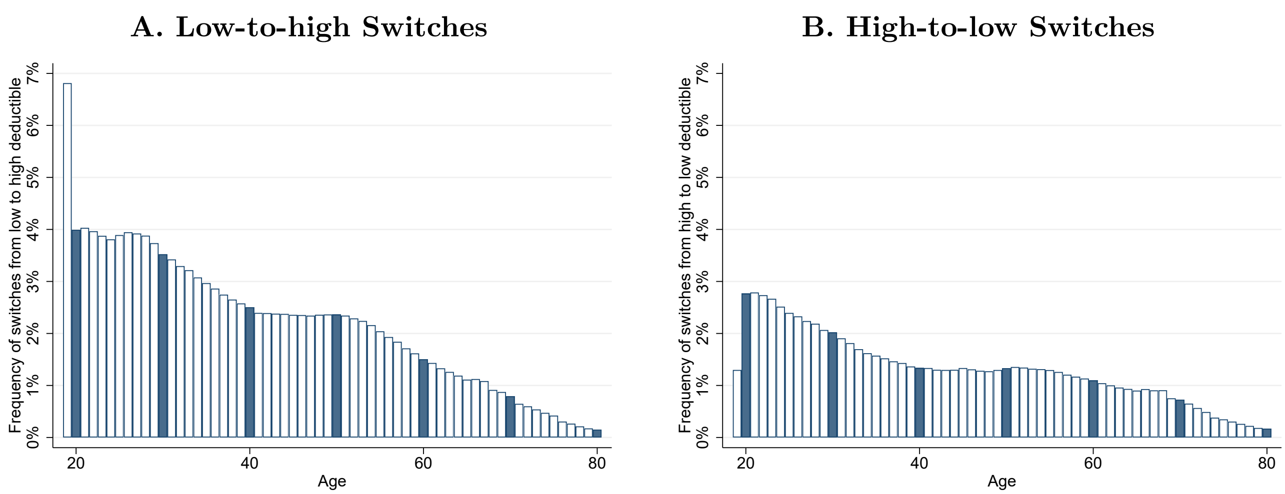
Notes: Notes from Table 2.4 and 2.5 apply; this table displays the same regressions without interacting the regressors with the probability of low costs.

FIGURE A2.2: DEDUCTIBLE CHOICE GRADIENT BY YEAR



Notes: This figure displays the relationship between take-up of the 500 deductible and the predicted probability of low costs, separately for the five years included in our final sample.

FIGURE A2.3: FREQUENCY OF DEDUCTIBLE SWITCHES BY AGE



Notes: This figure displays the frequency of deductible switches by age, in years 2014 to 2017. Panel A displays only switches to a higher deductible, and Panel B to a lower deductible.

Table A2.4: Predicted Health Risk by Observed and Optimal Deductible Choice

	2013	2014	2015	2016	2017
Probability of Low Costs	0.512	0.516	0.516	0.504	0.496
Healthy Individuals	0.752	0.758	0.760	0.759	0.762
Unhealthy Individuals	0.176	0.169	0.169	0.160	0.159
Individuals with 500 Deductible	0.748	0.760	0.763	0.762	0.763
Individuals with <500 Deductible	0.499	0.502	0.499	0.482	0.472
Share of Healthy Individuals	58.2%	58.9%	58.7%	57.4%	56.0%
Share of Individuals with the 500 Deductible	5.1%	5.3%	6.5%	8.0%	8.2%

Notes: This table displays, for the five years in our sample, the share of predictably healthy individuals and the share of individuals who took up the high deductible. It then shows the average probability of low costs for predictably healthy people (i.e., with a probability of low costs greater than .5), predictably unhealthy people, people who have taken up the 500 deductible and those who have not.

Table A2.5: Deductible Take-up and Predicted Health by Field

	Education Field	(1) Take-up of 500 Deductible	(2) Probability Low Costs	(3) Take-up of 500 Ded. | Being Predictably Healthy
1	**Statistics**	29%	87%	34%
2	Mathematics	21%	85%	27%
3	Physics	21%	91%	26%
4	Architecture and town planning	18%	88%	21%
5	Physical science	18%	82%	22%
6	Earth science	18%	88%	21%
7	**Philosophy and ethics**	17%	82%	21%
8	Medicine	17%	83%	20%
9	Chemistry	16%	87%	20%
10	Biology and biochemistry	16%	83%	20%
11	Science, Mathematics and Computing	16%	85%	19%
12	Computer science	15%	87%	18%
13	Environmental protection	15%	86%	18%
14	Political science and civics	15%	85%	18%
15	Design	15%	85%	18%
16	Sociology and cultural studies	14%	82%	18%
17	Mining and extraction	14%	91%	17%
18	Economics	14%	84%	17%
19	Humanities and Arts	14%	84%	18%
20	Dental studies	14%	76%	18%
21	History and archaeology	13%	82%	16%
22	Business and administration	13%	82%	16%
23	Pharmacy	13%	73%	17%
24	Health	13%	79%	16%
25	Environmental protection technology	13%	84%	15%
26	Medical diagnostic and treatment technology	13%	81%	16%
27	Religion	13%	80%	17%
28	Law	13%	80%	16%
29	Psychology	12%	77%	16%
30	Management and administration	12%	81%	16%
31	Engineering and engineering trades	12%	87%	15%
32	Forestry	12%	86%	14%
33	Therapy and rehabilitation	12%	78%	15%
34	Finance, banking, insurance	12%	80%	15%
35	Social and behavioural science	12%	79%	15%
36	Health and Welfare	12%	80%	15%
37	Fisheries	12%	94%	15%
38	Journalism and reporting	12%	80%	14%
39	Training for teachers w. subject specialisation	11%	79%	14%
40	Education science	11%	75%	14%
41	**Accounting and taxation**	11%	78%	14%
42	Agriculture, forestry and fishery	10%	81%	13%
43	**Marketing and advertising**	10%	80%	13%
44	Chemical and process	10%	85%	12%
45	Arts	10%	80%	13%
46	Electronics and automation	10%	86%	12%

Table A2.5: Deductible Take-up and Predicted Health by Field (cont'd)

47	Music and performing arts	10%	81%	12%
48	Training for teachers of vocational subjects	10%	81%	12%
49	Fine arts	10%	82%	12%
50	Humanities	10%	76%	12%
51	Library, information, archive	9%	78%	12%
52	Travel, tourism and leisure	9%	77%	12%
53	Electricity and energy	9%	88%	11%
54	Veterinary	9%	75%	12%
55	Mother tongue	9%	74%	12%
56	Audio-visual techniques and media production	9%	83%	10%
57	Building and civil engineering	9%	86%	10%
58	Life science	9%	79%	11%
59	Crop and livestock production	9%	79%	11%
60	Mechanics and metal work	9%	85%	10%
61	Wholesale and retail sales	8%	79%	11%
62	Foreign languages	8%	74%	11%
63	Motor vehicles, ships and aircraft	8%	87%	10%
64	Training for teachers at basic levels	8%	75%	10%
65	Materials (wood, paper, plastic, glass)	8%	86%	9%
66	Sports	8%	83%	10%
67	Teacher training and education science	8%	74%	10%
68	Military and defence	7%	81%	9%
69	Transport services	7%	83%	9%
70	Food processing	7%	78%	9%
72	Natural environments and wildlife	6%	86%	7%
73	Hotel, restaurant and catering	6%	77%	8%
74	Basic / broad, general programmes	6%	72%	9%
75	Social work and counselling	6%	70%	8%
77	Personal skills	6%	68%	8%
78	Textiles, clothes, footwear, leather	5%	70%	7%
79	Horticulture	5%	80%	6%
80	General Programmes	5%	71%	7%
81	Nursing and caring	5%	66%	7%
82	Domestic services	5%	66%	7%
83	Secretarial and office work	5%	65%	7%
84	**Protection of persons and property**	4%	78%	6%
85	Child care and youth services	4%	66%	6%
86	Computer use	4%	65%	6%
87	**Hair and beauty services**	4%	65%	5%
88	Occupational health and safety	4%	75%	5%
89	Training for pre-school teachers	3%	62%	0%
90	Literacy and numeracy	2%	62%	4%

Notes: For each field of study, this table shows: in Column (1), the fraction of individuals who take-up the 500 EUR extra deductible, in Column (2), the fraction of individuals with a probability of low costs < 375 EUR, and in Column (3), the fraction of individuals who take-up the 500 EUR extra deductible, conditional on having predicted health costs < 375 EUR.

Table A2.6: Deductible Take-up and Predicted Health by Professional Sector

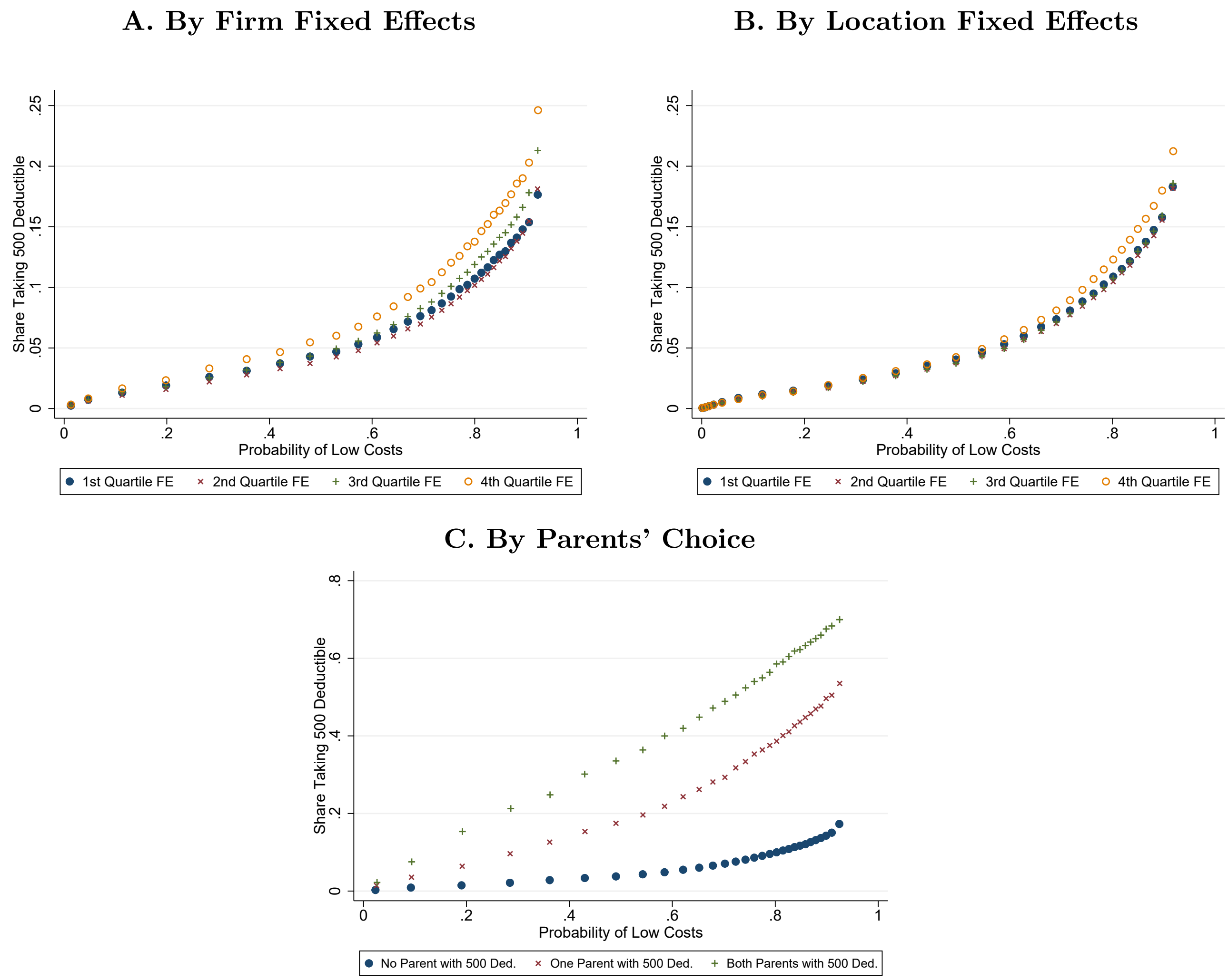
	Professional Sector	(1) Take-up of 500 Deductible	(2) Probability Low Costs	(3) Take-up of 500 Ded. | Being Predictably Healthy
1	**Business Services II**	13%	84%	16%
2	**Insurance and Health Insurance Firms**	12%	79%	15%
3	Business Services I	12%	82%	15%
4	Dairy Industry	12%	82%	14%
5	Banks	10%	81%	12%
6	Other Passenger Transport Land and Air	10%	79%	13%
7	Business Services III	10%	79%	13%
8	Agriculture	10%	85%	11%
9	Stoneware	9%	83%	11%
10	Publishers	9%	79%	11%
11	Cultural Institutions	9%	80%	11%
12	Telecommunications	9%	81%	12%
13	Government, Education and Science	9%	75%	12%
14	Food Industry	9%	80%	11%
15	Catering Industry I	9%	84%	10%
16	Tobacco Processing Industry	9%	76%	11%
17	Wholesale I	8%	82%	11%
18	Wholesale II	8%	81%	10%
20	Government, Police and Judiciary	8%	74%	11%
21	Wholesale of Wood	8%	82%	10%
22	Electronic Industry	8%	81%	13%
23	Carpentry	8%	83%	9%
24	Furniture and Organ Building	8%	83%	9%
25	Rail Construction	8%	78%	11%
26	NS Transport	8%	74%	10%
27	Sugar Processing Industry	7%	78%	10%
28	Chain Stores	7%	80%	9%
29	**Retail**	7%	79%	9%
30	Lending Industry	7%	81%	9%
31	Other Branches of Business	7%	79%	9%
32	Postal Transport	7%	72%	10%
33	Metal Industry	7%	80%	10%
34	**Construction**	7%	83%	9%
35	Merchant	7%	89%	8%
36	Mortar	7%	72%	9%
37	KLM Transport	7%	77%	9%
38	Bakeries	7%	79%	9%
39	Metal and Technical Industry	7%	82%	8%
40	Port Companies	7%	82%	9%
41	Chemical Industry	7%	79%	9%
42	General Industry	7%	81%	9%
43	Stone, Cement, Glass and Ceramic Industry	7%	77%	9%
44	Butchers Other	7%	80%	8%
45	Health, Mental and Social Industry	7%	71%	9%
46	Printing Industry	7%	80%	8%
47	Textiles Industry	7%	77%	9%
48	Inland Shipping	7%	83%	8%
49	Private Bus Transport	6%	70%	9%

Table A2.6: Deductible Take-up and Predicted Health by Professional Sector (cont'd)

50 Government, Local Government	6%	70%	9%
51 Butchers	6%	79%	8%
52 Wood, Brush and Packaging Industry	6%	82%	8%
53 Other Goods Transport Land and Air	6%	80%	8%
54 Government, Defense	6%	82%	11%
55 **Government, Public Utilities**	6%	77%	7%
56 Public Transport	5%	65%	8%
57 Security	5%	75%	7%
58 Plastering	5%	85%	6%
59 Taxi and Ambulance	5%	65%	8%
60 Catering Industry II	5%	70%	7%
61 Painting Industry	5%	81%	6%
62 Port Classifiers	5%	79%	6%
63 Fishing	4%	81%	6%
64 Work and Integration	4%	64%	6%
65 Dredging Industry	4%	85%	9%
66 Government, Other Institutions	4%	60%	7%
67 Roofing	4%	82%	5%
68 **Cleaning**	3%	70%	5%

Notes: For each professional sector, this table shows: in Column (1), the fraction of individuals who take-up the 500 EUR extra deductible, in Column (2), the fraction of individuals with a probability of low costs $<$ 375 EUR, and in Column (3), the fraction of individuals who take-up the 500 EUR extra deductible, conditional on having predicted health costs $<$ 375 EUR.

Figure A2.4: Take-up vs. Probability of Low Costs by Peer Effects



Notes: This figures shows the relationship between the probability of low costs and take-up of the high deductible for different subgroups. In Panel A, individuals are split in quartiles according to the fixed effect of the firm they are employed by. Those fixed effects are computed as detailed in Section IV.B. In Panel B, individuals are split in quartiles of postcode fixed effects, computed following the same method. In Panel C, individuals are split according to whether none of their parents, one of their parents, or both parents have taken up the 500 deductible.

VII.D Structural Choice Foundations

While it is not the focus of this paper to test different decision-making models, it is still useful to assess what kinds of micro-foundations can in principle rationalize the decision-making patterns that we document. This could also allow for a further refinement of the welfare analysis and policy recommendations. To shed some light on this, we simulate choice patterns under a range of distinct micro-foundations and compare the predictions of those simulations to our observed data. We consider a number of potential models of decision making that are proposed in the literature, including switching costs, loss aversion, imperfect information, rational inattention and mistakes.

Models of Choice Barriers

We first consider a model with default effects. Switching costs occur when consumers with a default plan option must pay some cost c_s to switch plans. This could be, e.g., a paperwork / transaction cost or reflect some reduced form of a multi-stage model with search and search costs. See a discussion of potential inputs into switching costs in Handel (2013). Specifically, setting the low deductible as the default plan option, a consumer chooses the high deductible if:

$$250 - (1 - \pi)500 - c_s > 0 \tag{2.5}$$

This assumes the model premium reduction of 250 EUR when taking the 500 EUR deductible. We consider heterogeneous population switching costs $c_s \sim U(0, 2 \times \bar{c}_s)$ for different average switching costs $\bar{c}_s$. As discussed, Brot-Goldberg et al. (2021) find strong default effects in Medicare Part D and show this is primarily due to inattention rather than switching costs. Note that we could alternatively model the default effects by for example allowing for a heterogeneous probability μ with which an individual is attentive and optimizes her deductible choice. Otherwise, she sticks to the default low deductible. The predicted choice patterns would be very similar.

Loss aversion occurs when losses loom larger than gains. In contrast with standard risk aversion, loss aversion can reduce the take-up of a deductible even when financial stakes are small. See Sydnor (2010) for a discussion of loss aversion as a potential driver of the over-insurance of modest risks. Following Kőszegi and Rabin (2007), we assume that realized payoffs are evaluated relative to expected payoffs, conditional on the deductible choice made, and losses receive a relative weight λ. In our setup, agents will then choose the high deductible if:

$$250 - (1 - \pi)500 - (\lambda - 1)\pi(1 - \pi)500 > 0. \tag{2.6}$$

Decisions could be made based on imperfect information. In our context, imperfect information enters by allowing consumers to receive a noisy signal $\hat{\pi}$ about their health, where $\hat{\pi} = \pi + \epsilon$ and $\epsilon \sim N(0, \sigma_\epsilon)$. They make a decision based on that noisy signal and choose the high deductible (for the model premium reduction of 250) if and only if

$$250 - (1 - \hat{\pi})500 > 0. \tag{2.7}$$

where the signal-to-noise ratio equals $\sigma_\pi / \sigma_\epsilon$.

Alternatively, individuals may decide rationally whether to pay attention and acquire information. In our context, rational inattention means that consumers, again, receive a noisy signal about their health, but then decide whether or not to pay a cost c_r to learn the true value of his/her health risk. Upon receiving the signal, agents face an expected choice value that integrates over the probability distribution of their potential true health statuses.[43] The value of acquiring the accurate information depends on whether the information would change her deductible choice and thus on the condition density $f(\pi|\hat{\pi})$ for $\pi > .5$ and $\hat{\pi} < .5$ and vice versa.[44] The result of our rational inattention setup is that, if a consumer starts with the low deductible, they will choose the high deductible if and only if one of the following conditions holds:

[43] Our model is similar in spirit to that laid out in Ho, Hogan and Scott Morton (2017), though there consumers obtain signals about plan characteristics while here they about signals about their own health status. We could recast our model as related to uncertainty about plan characteristics, likely with similar results.

[44] We simulate the conditional density by taking random draws from the empirical distribution of π and the normal distribution of ϵ. We then group the resulting π and $\hat{\pi}$ in ten bins of length 0.1, indexing them from 1 to 10. Then for each bin j of $\hat{\pi}$, we approximate the conditional density using:

$$p(\pi \in \pi_k | \hat{\pi} \in \hat{\pi}_j) = \frac{\#\text{individuals} \in \{\hat{\pi}_j \cap \pi_k\}}{\#\text{individuals} \in \hat{\pi}_j}$$

where π_k is bin k of π, and $\hat{\pi}_j$ is bin j of $\hat{\pi}$. To calculate the expected payoff, we use the middle value of each bin k of π.

$$\hat{\pi} > 0.5 \text{ and } \int_0^{0.5} [-250 + (1-\pi)500] f(\pi|\hat{\pi}) \, \mathrm{d}\pi < c_r \tag{2.8}$$

$$\hat{\pi} > 0.5 \text{ and } \int_0^{0.5} [-250 + (1-\pi)500] f(\pi|\hat{\pi}) \, \mathrm{d}\pi > c_r \text{ and } \pi > 0.5 \tag{2.9}$$

$$\hat{\pi} \leq 0.5 \text{ and } \int_{0.5}^{z} 1[250 - (1-\pi)500] f(\pi|\hat{\pi}) \, \mathrm{d}\pi > c_r \text{ and } \pi > 0.5 \tag{2.10}$$

The first condition results when consumers are so confident they are low that they don't find it worthwhile to pay the cost of precisely determining their health status, instead just electing to choose the high deductible right away. The second and third conditions occur when consumers decide to pay the cost to obtain a more precise signal, and are differentiated only by whether the initial signal value is bigger or smaller than the risk-neutral threshold of $\pi = 0.5$ for high deductible choice under the modal premium reduction.

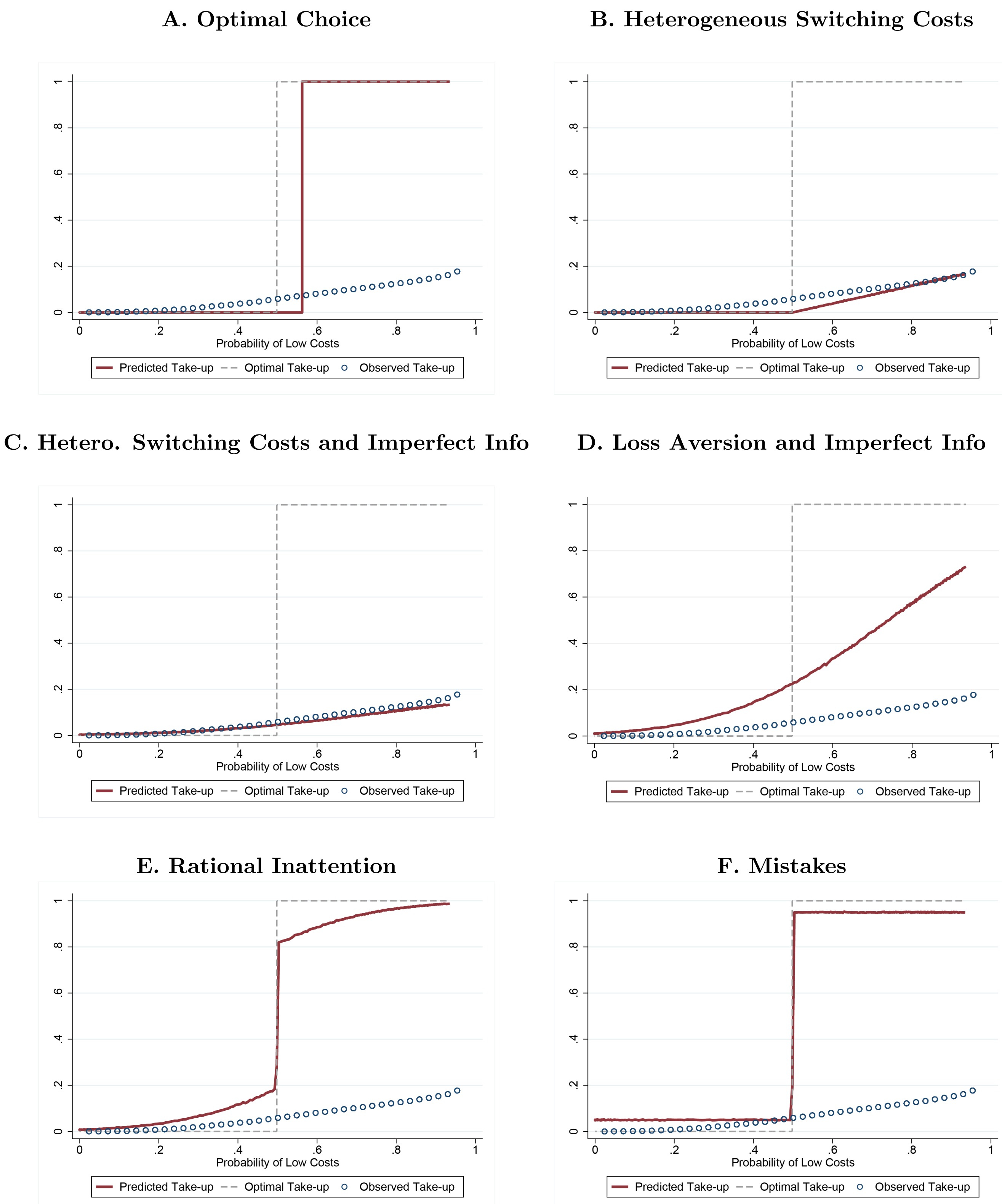
Finally, consumers may simply make mistakes. In our model, we assume a share $1 - \alpha$ of agents make rational, frictionless choices, while share α of agents make random choices.

Simulations

Figure A2.1 presents simulations of the deductible take-up rate as a function of health risk for the alternative decision models. For comparison, each panel plots the observed take-up rates and the deductible choice for the case where consumers are rational, frictionless, and risk-neutral, as in Figure 2.5. As discussed before, in a frictionless world, all consumers below a 50% probability of clearing the low deductible will elect the high-deductible, which looks starkly different from the observed low take-up rates. Risk-aversion only slightly alters this threshold, moving it to a marginally higher probability of low spending for the case where consumers are risk-averse with CARA coefficient of $1 * 10^{-4}$ (Panel A).

We then turn to the simulations for a decision models with switching costs. Note that with a homogeneous switching cost of 119 EUR, about 10 percent of the population would take up the high deductible, which corresponds to the observed take up rate. However, with heterogeneous switching costs uniformly distributed around the same mean of 119 EUR, we still predict meaningfully more high deductible purchases than we observe in the data, especially as consumers become predictably healthier

Figure A2.1: Deductible Take-up for Different Behavioral Models



Notes: This figure presents the results from decision-making simulations for the various models discussed in detail in the text. For each model, we contrast the predicted take-up rate with both the observed take-up rate and the take-up rate by a rational consumer in a frictionless world.

and healthier. Heterogeneous switching costs with a higher mean of 650 EUR (panel B) look much more similar to observed purchases as a function of health status. But this specification still predicts no purchasing of a high deductible for consumers with higher predicted probabilities of higher health spending. However, when we combine our model of high switching costs with our model of imperfect information about health status (with an assumed signal-to-noise ratio of 1), the simulated choices as a function of health status map very closely to observed choices (panel C).

Like switching costs for taking up the high deductible, loss-aversion helps to reduce the take-up rate of individuals around the 50% threshold. But similarly as for the case of risk aversion, the simulated take-up rates remain too high for reasonable loss-aversion parameters. Panel D simulates the deductible choices for a loss-aversion parameter of $\lambda = 2.25$ (i.e., when choosing the high deductible the payoff is reduced by $(2.25 - 1)\pi(1 - \pi)500$). Even with such strong loss aversion, individuals in very good health are predicted to always take up the deductible as the variance in financial payoffs they would get exposed to converges to zero.

Figure A2.1 also presents results for the rational inattention model (panel E) and the random mistakes model (panel F). The simulations for the rational inattention model use an information acquisition cost of $c_r = 25$ (for much higher values, no one pays this cost to learn about their true health status, making the model's predictions the same as the imperfect information model). We see that the take-up rate becomes more responsive to health risk around the threshold value, since individuals have to have probabilistic signals close to the marginal thresholds to acquire information, even with a reasonably small cost of 25 EUR. Furthermore, consumers with larger probabilities of being healthy are predicted to purchase the higher deductible much more than they actually do in practice. So we would need to combine the model of rational inattention with high switching costs to obtain predictions that are closer to observed choices. The simulations for the random mistakes model assume that a random 10% of consumers make mistakes. Clearly, the overall take-up rate is too high, so we again need an extra force to lower the take-up rate. Moreover, in the random mistakes model, the take-up rate is now also too high for individuals who are predicted to have high costs. This would not be resolved by combining the mistakes model with the imperfect information model.

This section illustrates how simulations based on different choice models compare with our data. Though there are a plethora of models one could write down that could help rationalizing the data (e.g., inertia, limited attention), a model of high switching costs combined with imperfect information fits the data very well. Importantly, high switching costs would further decrease the welfare gains from offering deductible choice. While we don't structurally estimate these models in our current context, these simulations give a sense of what models might make sense to estimate, and potentially test formally vs. one another, to implement a more detailed investigation of the mechanisms underlying the choice patterns we have documented.

VII.E Consumer Welfare and Policies: Further Details

This Appendix Section provides further details underlying our analysis of choice quality, the counterfactual analysis and the microfoundations of choice frictions.

VII.F Predicted Choice Model

For our analysis of choice quality in Section V, we start by predicting the deductible take-up rate $d(X_{it}, \pi_{it})$ as a function of their predicted health π_{it}, observables X_{it} and their interaction by running the regression:

$$Y = \alpha + \sum \beta_\delta 1[\pi = \delta] + \gamma X + \sum \nu_\delta 1[\pi = \delta] X + \epsilon$$

Here, Y is a binary variable that is 1 when an individual takes the 500 voluntary deductible and X is a rich set of controls, including demographics (gender, age, having children, living with a partner), financial variables (household gross income in deciles, net worth in quartiles, a dummy for having savings > 2000 EUR, for having a mortgage debt, for having another type of debt), education level and field, professional sector, and environment variables (firm and location fixed effect identified in Section IV.B in deciles, mother and father take-up of the high deductible).

We then define

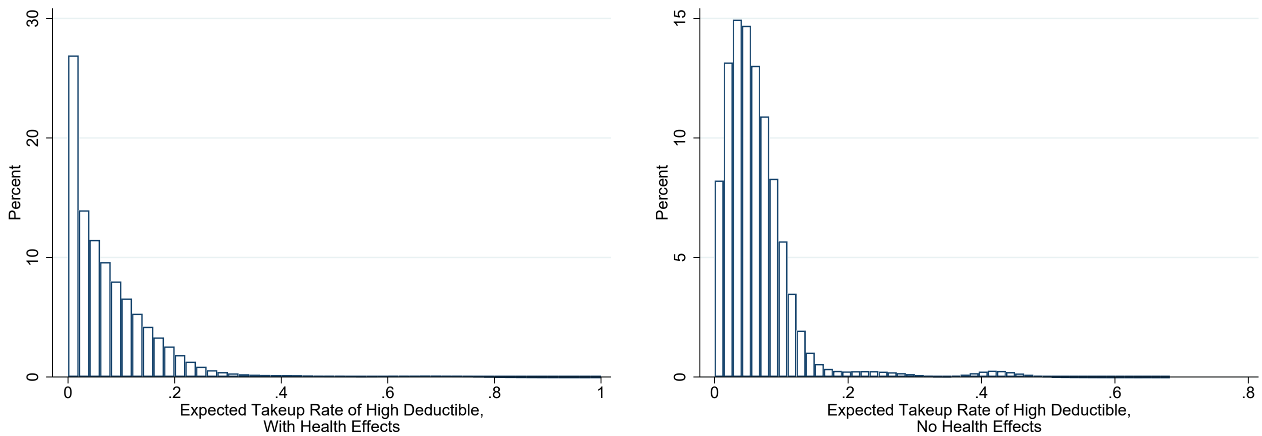
$$d_{\pi_{pop}}(X_{it}) = \sum_\delta d(X_{it}, \delta)\, dF_\delta,$$

which gives us the predicted deductible take-up rate for each observed X_{it} combination but as if there were a population of individuals with that X_{it} with the same health distribution as the overall population. In the same way, we predict the choice quality for individuals with demographic vector X_{it}, as captured by the probability to choose the contract that minimizes expected expenditures, $d^*_{\pi_{pop}}(X_{it})$, and the corresponding average financial loss $\Delta w^{*,\sigma}_{\pi_{pop}}(X_{it})$. That is,[45]

$$\begin{aligned} d^{*,\sigma}_{\pi_{pop}}(X_i) &= \sum_\delta \{1[\pi_\delta \leq .5][1 - d(X_{it}, \delta)] + 1[\pi_\delta > .5]\, d(X_{it}, \delta)\} dF_\delta, \\ \Delta w^*_{\pi_{pop}}(X_{it}) &= \sum \{1[\pi_\delta \leq .5]\, d(X_{it}, \delta)\left[CE^\sigma_{\pi_\delta,0} - CE^\sigma_{\pi_\delta,500}\right] + 1[\pi_\delta > .5][1 - d(X_{it}, \delta)]\left[CE^\sigma_{\pi_\delta,500} - CE^\sigma_{\pi_\delta,0}\right]\} dF_\delta. \end{aligned}$$

[45] Note that we use the average predicted risk for the different health deciles to calculate the certainty equivalents and to determine whether one should take up the deductible or not.

FIGURE A2.1: PREDICTED DEDUCTIBLE CHOICE

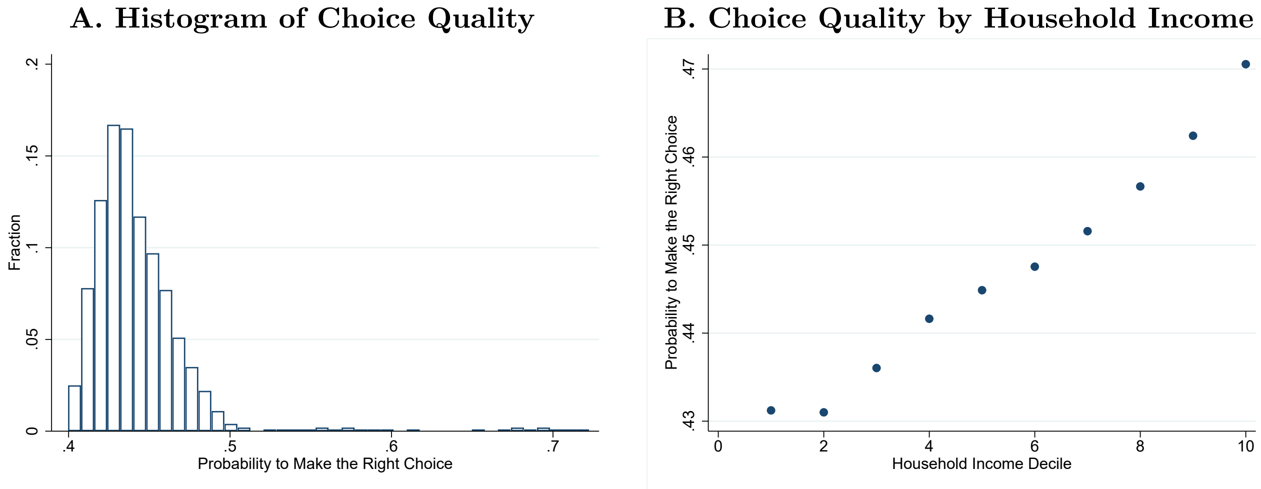


Notes: This figure shows the distribution of predicted 500 EUR extra deductible take-up rate. Panel A shows the predicted 500 EUR deductible take-up with health effects, while Panel B shows the take-up without the health effects.

The choice quality varies through the deductible choice predicted by the set of demographics X_i for different health risks, but again reflects the population distribution of health risks.

Figure A2.1 compares the distribution of predicted deductible choice, with and without the effect of healthcare cost risk. These are denoted in previous equations as $d\left(X_{it}, \pi_{it}\right)$ and $d_{\pi_{pop}}\left(X_{it}\right)$ respectively. As shown before, health has a meaningful impact on deductible choice, but there is substantial heterogeneity in likelihood of choosing a deductible just as a function of X_{it}, netting out health effects. While losses range up to 200 EUR when factoring health risk into choices, when assuming the population distribution of health for a given X_i the expected loss ranges between 50 and 80 as a function of X_i. Panel A of Figure A2.2 ranks individuals according to the quality of their choice first, as discussed in the text, and then shows the distribution of the probability to make the right decision for the different groups of quality choice. Panel B of Figure A2.2 shows the probability of making the right decisions for different income groups.

Figure A2.2: Heterogeneity in Choice Quality



Notes: Panel A shows the distribution of probabilities that consumers make the right deductible choice for a given set of socio-demographic characteristics $X_{i,t}$. The right choice is defined as the choice a rational consumer would make, as explained in Section III.A: to take the 500 EUR extra deductible if she expects her costs to be below 375 EUR with a probability larger than 0.5; to choose the low deductible otherwise. Individuals are binned in 1000 quantiles of choice quality; the variable displayed in this histogram is the binned average of the individual probability to make the right choice. Panel B shows the probability to make the right choice by income decile.

Counterfactual Policies

Table A2.1: Counterfactual Policies, Controlling for Health Effects

	Optimal Deductible	High Deductible Only (875 EUR)	Low Deductible Only (375 EUR)
Risk Neutral			
Unweighted	63.7	-11.1	-5.3
Low Inequality Aversion	64.2	-10.6	-4.8
High Inequality Aversion	65.0	-9.8	-4.0
σ=.0001			
Unweighted	62.8	-12.8	-5.2
Low Inequality Aversion	63.2	-12.3	-4.7
High Inequality Aversion	64.0	-11.6	-3.9
σ=.001			
Unweighted	53.6	-28.6	-4.3
Low Inequality Aversion	53.9	-28.2	-3.9
High Inequality Aversion	54.5	-27.7	-3.3

Notes: Notes from Table 2.9 apply. This table performs the same exercise, except that each individual is attributed the population's health distribution, such that the correlation between income and health is controlled for.

Chapter 3

The Income Gradient in Mortality during the Covid-19 Crisis: Evidence from Belgium

Co-authored by: André Decoster (KU Leuven), Thomas Minten (LSE) and Johannes Spinnewijn (LSE).[1]

Abstract: We use population-wide data from linked administrative registers to study the distributional pattern of mortality before and during the first wave of the Covid-19 pandemic in Belgium. Over the March-May 2020 study period, excess mortality is only found among those aged 65 and over. For this group, we find a significant negative income gradient in excess mortality, with excess deaths in the bottom income decile more than twice as high as in the top income decile for both men and women. However, given the high inequality in mortality in normal times, the income gradient in all-cause mortality is only marginally steeper during the peak of the health crisis when expressed in relative terms. Leveraging our individual-level data, we gauge the robustness of our results for other socioeconomic factors and decompose the role of individual vs. local effects. We provide direct evidence that geographic location effects on individual mortality are particularly strong during the first wave of the Covid-19 pandemic, channeling through the local number of Covid infections. This makes inference about the income gradient in excess mortality based on geographic variation misguided.

[1]We thank Statbel, and especially Patrick Lusyne and Lien Tam Co for help with the data. Chloé de Meulenaer, Sebastian Ernst and Stijn Van Houtven provided excellent research assistance. We gratefully acknowledge funding by ERC (grant #716485) and by the Belgian Ministry of Social Security (2020-DGSTRAT-Studie Covivat).

I Introduction

The Covid-19 pandemic affects everyone, but not everyone is affected equally. An important concern is that the burden of the Covid-19 crisis falls disproportionately on people with low income or socioeconomic status. A burgeoning literature studying the economic impact of the Covid-19 crisis and the associated policy measures on employment, earnings, and consumption (e.g., Adams-Prassl et al. (2020), Bachas et al. (2020), Chetty et al. (2020*a*)) documents substantial differences depending on socioeconomic status. In parallel, many research efforts have focused on the inequality of the health impact of the pandemic. While a rapidly growing literature suggests that socioeconomic factors are important determinants of Covid-19-related mortality (e.g., Chen, Waterman and Krieger (2020), Drefahl et al. (2020), Jung et al. (2020) and Williamson et al. (2020)), a strong income and/or socioeconomic gradient in health outcomes and in mortality in particular was present prior to the arrival of the coronavirus. Indeed, one seemingly perennial finding documented in many countries is that mortality rates are higher among individuals with lower socioeconomic status (e.g., Chetty et al. (2016), Mackenbach et al. (2019)). An outstanding question is how the Covid-19 crisis has affected this relationship between income or socioeconomic status and mortality in particular. Lower income households may have been more exposed to the virus, for example because of their living or work conditions, but may also have medical conditions that put them more at risk when infected.

To answer this question we use population-wide data drawing from several administrative registers in Belgium. Belgium has been hit particularly hard by the first wave of the pandemic, noting the highest per capita death toll of any country by 30 May 2020. We use the mortality registers updated until June 2020 and linked to income registers as well as other demographic and socioeconomic information. This allows us to measure the income and socioeconomic gradient in mortality at the individual level, which we compare during the height of the Covid-19 health crisis - from March until May 2020 - with the corresponding months from 2015 to 2019.

A first advantage of our data on all-cause mortality is that we can perform a counterfactual analysis comparing mortality during and before the crisis. This allows us to provide evidence of the unequal burden of mortality due to the Covid-19 pandemic and relate it to the "usual" inequality in mortality in Belgium. A large number of papers, as shown in the left column of Table 3.1, have used Covid-19-related deaths counted by the health authorities, mostly finding stark differences in mortality across different socioeconomic groups. But, importantly, not knowing the counterfactual mortality, these studies cannot infer how the Covid-19 pandemic has affected inequality in mortality.

A second advantage of our data is that we can measure income and mortality at the individual level and therefore separate individual income-related factors from location effects. This is important because the correlation between mortality and household income may be driven by many factors, not in the least the location one lives in, and the importance of these factors may have changed during the pandemic. At the local level, there have been clear differences in the inflows of infected individuals and the propagation of infections. Moreover, healthcare capacity varies across localities, leading to differences in access to healthcare during the pandemic. These local factors may translate into differential mortality at different income positions. But we can also expect differences in exposure to infections by income or socio-economic status at the individual level (e.g., due to differences in employment, housing, social contacts, etc). In addition, individuals at different income levels have different co-morbidities and hence a different case fatality risk once infected.

The rows of Table 3.1, however, show that all but one paper analyze the relationship between mortality and socioeconomic status measured at the municipality or another location-specific level in various countries. While most studies find a negative association, some indicate a more ambivalent relationship.[2] An important limitation of studies that use aggregate measures, however, is that they do not measure the direct link between individuals' socioeconomic status and mortality. By looking at area-level measures, these effects may confound various local factors like access to and quality of care, exposure risk and also local policy responses.

As listed in Table 3.1, only a few studies have looked at excess mortality, but using aggregate data, and only one study has used individual data, but looking at Covid-19-related mortality. This notable exception is the study by Drefahl et al. (2020), finding a negative association between individual income from Swedish registries and Covid-19-related deaths. There is thus a gap in the literature studying the relationship between individual-level measures of socioeconomic status and excess mortality during the

[2] Brandily et al. (2020), for instance, investigate excess mortality across municipalities in France, and find a negative income gradient, with excess mortality in the poorest municipalities twice as large as in other municipalities. In contrast, Jung et al. (2020) investigate the relationship between Covid-19 mortality and poverty across US counties and find that poverty and mortality are positively related in areas of low population density. In areas of high population density, however, they find a U-shaped relationship. Knittel and Ozaltun (2020) also analyze the county-level relationship between Covid-19 mortality and poverty in the US but find no correlation. They even find a *positive* relationship between mortality and median home value. Desmet and Wacziarg (forthcoming) find a positive correlation between Covid-19 cases or mortality and median household income in US counties in the first months of the pandemic, that has turned negative afterwards.

Table 3.1: Findings on the Association Between Socioeconomic Status (SES) and Mortality during the Covid-19 Crisis

	Covid-19 Mortality	Excess Mortality
Individual-level measure of SES	Drefahl et al. (2020) *Negative association* - Sweden	
Aggregate measure of SES	Abedi et al. (2020) *Negative association* - US Ashraf (2020) *Negative association* - World Brown and Ravallion (2020) *Negative association* - US Chen and Krieger (2020) *Negative association* - US Desmet and Wacziarg (forthcoming) *Mixed results*[a] - US Jung et al. (2020) *Mostly negative association*[b] - US Kim and Bostwick (2020) *Negative association* - US Knittel and Ozaltun (2020) *No/positive association*[c] - US Office for National Statistics (2020) *Negative association*[d] - UK Sá (2020) *Mixed results*[e] - UK Tubadji, Webber and Boy (2020) *Negative association* - UK Williamson et al. (2020) *Negative association* - UK	Brandily et al. (2020) *Negative association* - France Calderón-Larrañaga et al. (2020) *Negative association* - Sweden Chen, Waterman and Krieger (2020) *Negative association* - US

Notes: This table classifies the existing applied work on the relationship between SES and Covid-19-induced mortality into four quadrants, depending on the measure of mortality and SES used. Noted under each reference are the observed relationship between SES and Covid-19-induced mortality, as well as the country, in which the study was conducted. For papers that did not find a clear association, we provide further details below.
[a]Desmet and Wacziarg find that a higher level of Covid-19 mortality in a county was positively correlated with median household income in the first months after the onset of the pandemic, but the correlation turned negative afterwards. They also report the relationship with measures of poverty and educational attainment.
[b]The authors find a U-shaped relationship between Covid-19 mortality and SES in counties with high population density and a negative relationship in counties with low population density.
[c]Knittel and Ozaltun find no correlation between Covid-19 death rate and poverty rate but find a positive correlation between Covid-19 death rate and median home value.
[d]The authors look at both Covid-19 mortality and all-cause mortality, but do not examine excess mortality.
[e]Sá finds no simple correlation between deprivation and Covid-19 mortality. Regression results show Covid-19 mortality to be higher in more deprived areas, although the relationship disappears when controlling for self-reported health.

Covid-19 crisis. Our paper aims to fill this gap and provides three main sets of results:

First, when looking at the entire March-May 2020 period, we find only slight non-significant excess mortality for people under 65 in Belgium during the first wave of the Covid-19 crisis. While EuroMOMO (2020) do find significant excess mortality for 45-64 year olds in Belgium during some weeks in the Covid-19 crisis, when we look at the entire March-May 2020 period, we do not find significant excess mortality for this age group. Our findings for the 45-64 age group in Belgium contrast with findings for this age group in other countries studied in EuroMOMO (2020). We also do not find a meaningful change in the income gradient of all-cause mortality for this demographic group compared to the baseline years. The ratio between mortality among the bottom income and the top income decile stayed around 5 for men and 4 for women. In light of the earlier evidence on the unequal incidence along the income distribution in this age group of both Covid-19-related mortality (e.g., Drefahl et al. (2020)) and the underlying risk factors (e.g., Raifman and Raifman (2020), Wiemers et al. (2020)), this may come as an unexpected result.

Second, our results show that the Covid-19 pandemic significantly affected the mortality of individuals aged 65 and over, and that excess mortality for this age group declines significantly with income. For example for men, we estimate 326 excess deaths out of 100,000 in the bottom income decile compared to 131 in the top income decile. Importantly, the income gradient in mortality is strongly negative in normal times too. As a result, expressed in relative terms, the income gradient in all-cause mortality is only marginally steeper during the peak of the health crisis. We compare different measures for judging the inequality in all-cause mortality that the Covid-19 pandemic brought to this subpopulation, but can only reject lower-than-normal mortality inequality during its peak. Overall, our results for this age group are confirmed when looking into other socioeconomic factors. We find strong educational gradients in excess mortality, as elderly who did not complete primary school experienced higher increases in mortality rates (30.47%) than elderly with higher education (21.91%). The increase in mortality has also been higher among Italian-, Turkish- and Polish-born residents than among Belgian-, German- and Dutch-born residents. We study individuals living in nursing homes separately, as excess mortality during the Covid-19 pandemic has been particularly high for this subgroup, but we do not find any income gradient in mortality before or during the Covid-19 crisis for them.

Third, we try to separate the role of individual and local effects in determining the income gradient and to investigate whether their role has changed during the crisis. As mentioned, most prior work has been constrained by data availability and only considers differences in mortality by income aggregated at some local level. We find that our estimates of the income gradient using household income are robust to the inclusion of municipality fixed effects during the baseline years. This indicates that in the Belgian context, geographical differences in healthcare do not explain much of the pre-Covid income gradient in mortality. However, during the Covid-19 crisis, location becomes more important and explains about half of the increase in the income gradient at the household level. The relation between mortality and municipality income itself doubles during the crisis. This increase is much larger than measured at the individual level. Interestingly, this increase can be fully explained by differences in Covid-19 infections at the municipality level. Importantly, but not unexpectedly, inference relying on geographical variation about the individual socioeconomic factors of mortality during the pandemic would be misguided.

The paper proceeds as follows. Section II discusses the data and context. Section III present our main results, starting with the income gradients of all-cause and excess mortality and discussing the inequality implications, then studying other socioeconomic factors and the role of location effects. Section V concludes.

II Data and Setting

Our study focuses on Belgium, which has been faced with a high count of Covid-19-related deaths per capita. The introduction of the Covid-19 virus in Belgium has mostly been attributed to the return of ski tourists from Italy and Austria after the national holiday week from February 22 until March 1, 2020. In response to the quick surge of Covid-19 infections that followed, a nationwide lockdown was imposed from March 18. This was slowly phased out starting with the opening of garden stores and DIY stores on the 18th of April, followed by the staggered opening of selected sectors (May 4), retail stores (May 11), and cafes and restaurants (June 8). At the same time, there was a staggered loosening of the restrictions on the number of close social contacts citizens could maintain with individuals from other households, going from 2 (May 4), to 4 (May 11) and 10 (June 8). These policy measures during the first months of the Covid-19 crisis were set at the federal level with arguably limited variation at the local level.

To study mortality across the income distribution, we link administrative data on mortality from the national register with data on income from tax records. We also link this to data from other population-wide registers, including the 2011 census. Below, we discuss the different data sources, which have been linked and made available through the Belgian Statistical Institute (Statbel). We start by briefly sketching the Belgian health system against the background of a strong welfare state.

Health system and inequality in Belgium. Most inequality and poverty statistics for Belgium stand in sharp contrast with the conclusions for many other countries that inequality, poverty, material deprivation and insecurity are on the rise. OECD (2018) reports a minor change in the Gini from 0.257 in 1983 to 0.264 in 2011, and even a slight decline since 2004.[3] The risk-of-poverty rate has also remained stable during the last decades. Although a comprehensive explanation for these findings is still lacking, the strength of different components of the welfare state, such as labour market institutions, redistributive taxes and a high level of social protection financed by a high level of social insurance contributions, undoubtedly play their role. The Belgian healthcare system is another exponent of the strong welfare state with a combination of near universal coverage, regulated choice by both patients and healthcare providers, and no severe capacity constraints (hospital beds, doctors, nurses). Belgians enjoy a relatively high life expectancy and on average their self-reported health is high (5th place in the ranking of 28 EU-countries, see OECD/European Observatory on Health Systems and Policies (2017), Figure 4). But like in other developed countries, considerable socio-economic inequalities remain in health outcomes and behaviors, such as in self-reported health and in Covid-relevant unhealthy lifestyles such as smoking, alcohol use, diets and lack of physical activity. Despite the near universal coverage, Bouckaert, Maertens de Noordhout and Van de Voorde (2020) also point to an important socio-economic gradient in self-reported unmet care needs for financial reasons.

Mortality in Belgium. Using the mortality records from the national register, Appendix Figure A3.1 shows the dramatic increase in daily deaths in March to May 2020 following the onset of the Covid-19 pandemic. To investigate the effect of the Covid-19 pandemic and associated policy responses on mortality, we consider its impact on all-cause mortality and define excess mortality as the difference in mortality between 2020 and the average mortality in the corresponding period from 2015 to 2019. Positive excess mortality in 2020 primarily occured from March 16 to May 27, with a record number of 314 excess deaths recorded on April 10. Another period of significant excess mortality occurred between August 8 and August 20, yet is ascribed to a heat wave that lasted from August 5 to August 17. We

[3] Based on survey data, Van Rie and Marx (2014) conclude that Belgian income inequality remained fairly stable between 1985 and the late 2000s. Decoster, Dedobbeleer and Maes (2017) also do not find evidence that those at the top of the income distribution in Belgium have benefited disproportionately from the economic growth since the nineties.

therefore take only the March-May period as the relevant period with which to compare mortality during the Covid-19 crisis in 2020 to the baseline years.[4] We note that total excess mortality in Belgium in this period is 8,195, which is close to the official number of Covid-19 deaths of 9,467 counted by Belgian health authorities.[5]

Income. The income data originate from IPCAL, an administrative database that is drawn from personal income tax records. We use total net taxable income, which refers to income before tax, after social security contributions have been paid and costs deducted. It is a general definition of income, and includes labour income, unemployment benefits, sickness benefits and pensions.[6] Income data retrieved from tax declarations are contingent upon the tax legislation. Since capital income is subject to a liberating withholding tax, and some important benefits, such as child benefits, or the living wages (*leefloon*) are exempt from personal income tax, these income components are not included. We aggregate personal income over households to obtain household income. We do not aggregate income to the household level for individuals in nursing homes, as their household includes all other residents of the nursing home.[7]

Demographic and Socioeconomic Variables. Most of the demographic information (age, country of birth, gender, municipality) originates from the national registries in Demobel. We also have an indicator for whether an individual is residing in a nursing home (*woonzorgcentra*) from Statbel. Economic sector and education level originate from the 2011 census. Municipality-specific information on per capita income and density comes from Statbel.

[4] Mortality was significantly higher than in the previous five years continuously between March 21 and May 21, between May 22 and May 25, and between August 8 and August 20.

[5] We do find a 13% discrepancy between excess mortality and the official death count (see also Molenberghs et al. (2020)). Potential reasons for this discrepancy are the decrease in other-cause mortality in the study period, but also the over-counting of the Covid-19 death toll. Famously, all deaths with suspected involvement of Covid-19 were counted as Covid-19 deaths in Belgium. This has been actively portrayed as one of the reasons why the published death toll of Covid-19 in Belgium is one of the highest in the world.

[6] Pension income in Belgium is complex, and our data source based on taxable income captures annual pension income imperfectly. Pensions of the dominant 'first pillar' (the social security benefits) are a direct function of prior labor earnings and are mostly observed in the data. However, the treatment of the occupational pensions (the 'second pillar') and the personal private savings (the 'third pillar') is more problematic. Not only are these benefits only partly taxable in highly complex schedules, but tax payers can opt for the payment of this pension as a once-off lump sum amount. We find, nevertheless, that the correlation between our income measure when retired and earlier in life is quite strong, as evidenced by a high correlation of 0.63 between income decile at age 55 and income decile at age 65 for the same individual.

[7] The household indicators in our data come from the socioeconomic Demobel database. The income for individuals in nursing homes is dominated by pension income, as is the case for other 65+ year olds.

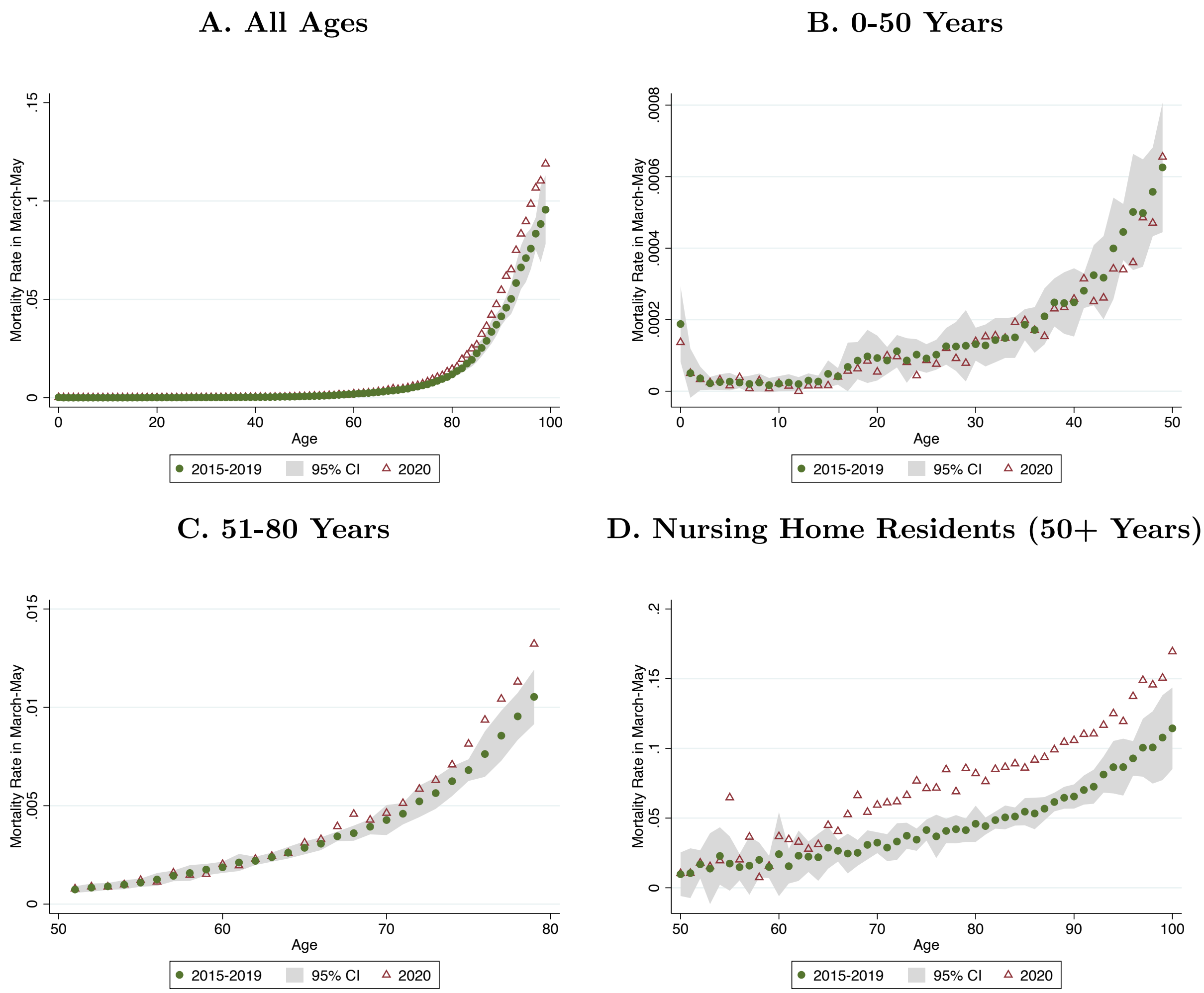
II.A Excess Mortality by Age

Figure 3.1 contrasts mortality rates by age during the months March-May in 2020 and the corresponding period in 2015-2019. Panel A provides a clear visual picture of excess mortality across different ages, indicating how concentrated it has been among the elderly. Panel B zooms in on individuals aged 0 to 50 and shows that there was no significant excess mortality for people of those ages. Panel C zooms in on individuals aged 51-80 and shows that significant excess mortality only shows up for individuals aged 65+. These findings may seem surprising, as the Belgian health authorities (Sciensano) counted several hundred Covid-related deaths in the 45-64 age bracket. Note that there were a few weeks with significant excess mortality for 45-64 year olds (weeks 13-17, EuroMOMO (2020)), but these differences have been too small to lead to significant excess mortality over the March to May 2020 study period. This pattern has been documented before in Belgium (Molenberghs et al. (2020)) as well as in other countries (EuroMOMO (2020)). Only a few European countries, such as Spain and the UK, experienced large and significant excess mortality for people under 65 over a longer period. Clearly, returning to panel A, excess mortality is highest for individuals aged 80+.

Panel D considers nursing home residents separately and shows a substantial increase in mortality for nursing home residents aged 70+. Interestingly, this increase seems rather uniform for all ages above 70, which might be due to the selection of individuals less able to care for themselves into nursing homes, so that health status does not vary as much between older and younger nursing home residents compared to the general population. Our calculations suggest an especially heavy toll on nursing homes, as we estimate that in March to May 3.6% of all residents of nursing homes in Belgium died due to the Covid-19 pandemic.

Overall, we find important differences in excess mortality in Belgium across the age distribution during the Covid-19 crisis. Based on the different patterns in excess mortality, our results in Section III distinguish between individuals aged 40-64, individuals aged 65+ not living in nursing homes (or other collective households), and individuals aged 65+ who are living in nursing homes. Appendix Table A3.1 provides summary statistics for the three samples. One way to aggregate the mortality effects throughout the age distribution is to calculate the *period* life expectancy, which is the life expectancy of an individual based on the age-specific mortality rates in a given period (e.g., Chetty et al. (2016)). While the mortality rates increased the most for the elderly, changes in the mortality rates of the elderly

Figure 3.1: Mortality Rates in March-May By Age



Notes: These figures show the average mortality rate by age in March-May of 2015-2019, with a 95% confidence interval, and in March-May of 2020. Panels A-C show mortality rates for all Belgian inhabitants, excluding people living in collective households, or households with more than 10 individuals. Panel D shows mortality rates for nursing home residents according to the classification of Statbel.

have a smaller impact on life expectancy measures than changes among younger age groups. Based on the mortality rates in the baseline years, the period life expectancy in 2020 was 79.09 for men and 83.40 for women. Using the mortality rates between March-May of 2020 instead, the period life expectancy would be 1.87 years shorter for men, and 1.83 years for women.[8]

[8]The period life expectancy for March-May 2020 is calculated in two steps. First, the 2015-2019 yearly mortality rates for each age-gender group are scaled with the P-score+1 obtained in March-May 2020, where the P-score is the estimated excess mortality divided by the baseline mortality within that group. Second, these scaled mortality rates are used to calculate life expectancy at birth for men and women separately.

III Income Gradient of Mortality Rates

We now turn our analysis to the socioeconomic correlates of mortality and how their relationship changed during the Covid-19 crisis. Our main focus is on the income gradient of mortality rates and in particular on the comparison of the income gradient during the Covid-19 crisis with that of the baseline years. Since income - and socioeconomic status more broadly - is central to equity considerations, a large literature has studied the importance of health inequality along this dimension. Importantly, income gradients by themselves do not allow one to draw any causal conclusions regarding the effect of income on health outcomes, either before or during the Covid-19 crisis. However, comparing the income gradients before and during the crisis sheds lights on how the crisis has affected health inequality along this dimension.

III.A Income Gradient Before vs. During the Covid-19 Crisis

To calculate the mortality-income gradient, we rank individuals based on their household income and calculate mortality rates for different income quantiles. In particular, for every year t, we rank individuals within their age-gender group based on their lagged household income in year $t-3$ and assign a decile based on these rankings. This means that the yearly deciles will be based on an individuals' lagged household income relative to all other individuals of the same gender and age in Belgium. We use a 3 year lag so that we observe lagged income for all years, including 2020, but by using lagged income we also reduce the potential response of income to health shocks (see Chetty et al. (2016)) and in particular the response of income itself to the Covid-19 crisis.[9]

Figure 3.2 shows mortality rates for men and women of different age groups, both in the control years (2015-2019) and in 2020, across deciles. The slope of the income gradient, either using a linear or loglinear regression specification, corresponds to two commonly used inequality measures in the literature (see Mackenbach and Kunst (1997) and Moreno-Betancur et al. (2015)): the SII or Slope Index of Inequality and the RII or Relative Index of Inequality respectively. Denoting mortality for decile d by $m(d)$, SII measures the difference $m(1)-m(10)$, and is often expressed in deaths per 100,000, whereas RII is defined as the ratio $m(1)/m(10)$ or as the percentage change in mortality across the

[9] Calculating household income deciles based on one year only is appropriate, as we find that individuals' household income deciles remain relatively stable over time, a finding corroborated in Chetty et al. (2016). Importantly, we find that the high correlation between individuals' income deciles continues after retirement.

income scale. Appendix Table A3.2 reports the slope estimates and the corresponding inequality indices for each of the income gradients.[10]

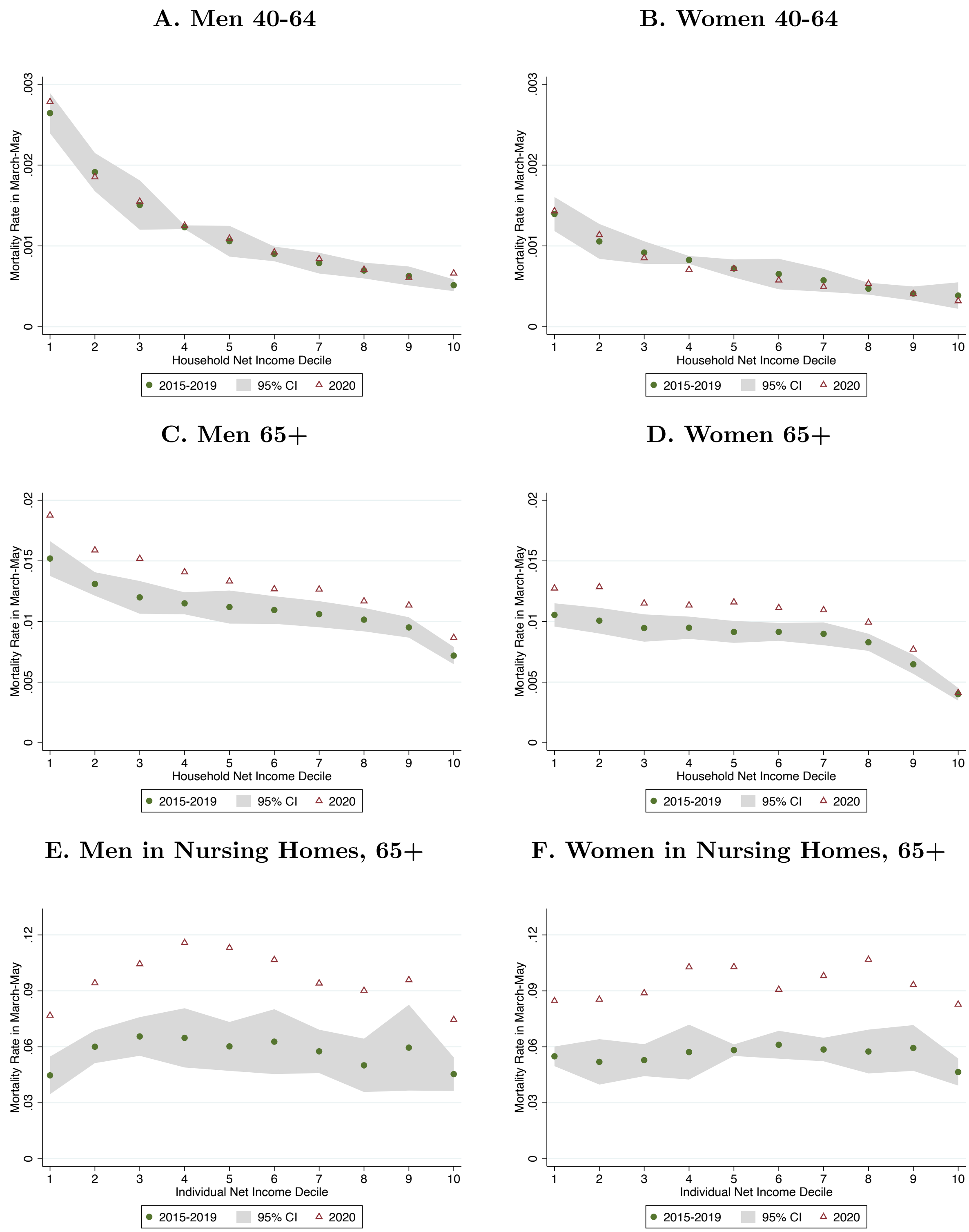
The top panels of Figure 3.2 focus on individuals between 40-64 years old. The income gradient is already strong and negative in the baseline years. For men, the mortality rate is estimated to be 5.3 times higher in the bottom income decile than in the top income decile. The same holds for women, be it somewhat less outspoken with a corresponding RII of 3.9. The negative income gradient in mortality rates is a persistent finding that underlies the substantial differences in life expectancy between low- and high income individuals (e.g., Chetty et al. (2016)). Importantly, the figure shows that for this age group the mortality rates during the Covid-19 months are indistinguishable from those of the control months. While we documented above that there is no average excess mortality in this age group, the income gradients confirm that this is also true for individuals in different income groups.

The middle panels of Figure 3.2 show a very different picture for the elderly. In the baseline years, the income gradient is again strongly negative. Compared to the younger age groups in panels A and B, the gradient is stronger when expressed in absolute terms, but smaller when expressed in relative terms (see Table A3.2 in Appendix). More importantly, the mortality rates jump significantly during the Covid-19 months and they do so in each of the income groups of this age group. The SII increases substantially for men and women. For example, for men, the estimated difference in deaths of 596 per 100,000 individuals between the bottom and top income deciles during the baseline years increases to a difference of 791 deaths during the coronavirus period. However, expressed in relative terms, the increase in the income gradient has been more modest. The estimated RII increases from 1.8 to 1.9 for men and from 2.1 to 2.3 for women.

Finally, the bottom panels show the mortality rates for individuals in nursing homes, who are excluded from the other panels. Interestingly, we do not find a clear income gradient in mortality rates for individuals in the baseline years. As is well known, mortality increased most starkly for this group, but it did so uniformly across income groups.

[10] In particular, with the estimated coefficient of the loglinear regression equal to β, we estimate the mortality ratio between the first and the tenth decile to be equal to $\frac{1}{(1+\beta)^9}$.

Figure 3.2: Mortality Rates in March-May By Gender/Age/Income



Notes: These figures show the average mortality rate by income decile in March-May of 2015-2019, with a 95% confidence interval, and in March-May of 2020. Panels A-D show mortality rates for all Belgian individuals, excluding people living in collective households or households with more than 10 individuals. Panels E and F show mortality rates for Belgian inhabitants aged 65 or older and living in nursing homes. These individuals are ranked based on their individual income within the corresponding age-gender group in the Belgian population, but to control for differential selection into nursing homes the results in Panels E and F are residualized on age.

III.B Distributional Pattern of Excess Mortality

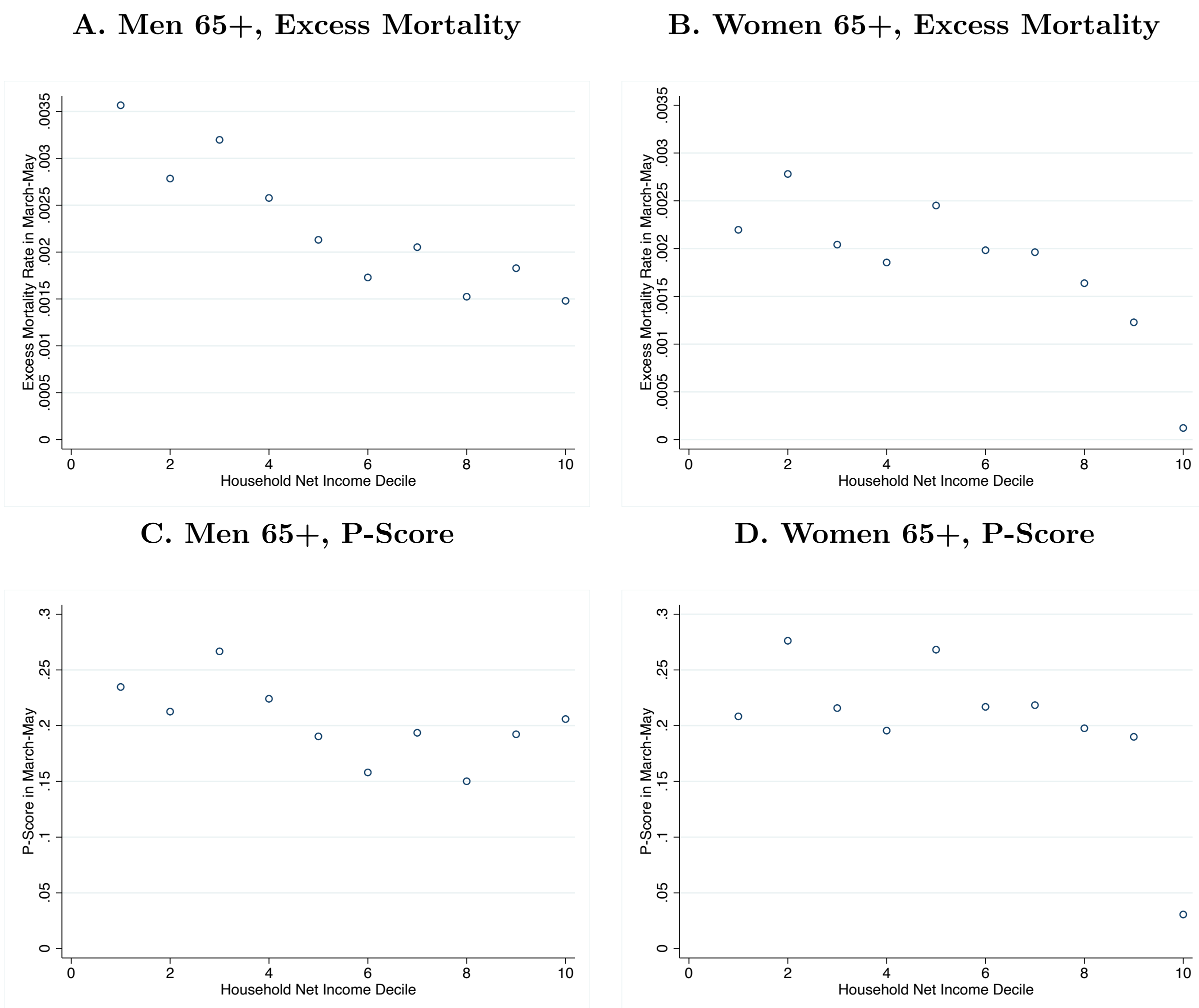
Several studies have analyzed the relationship between Covid-19-related mortality and socioeconomic status, arguing that the incidence of the pandemic falls disproportionately on low-income individuals. Our analysis of income gradients - before and during the Covid-19 crisis for different groups - nuances this view and provides a new perspective. We already noted above that different pictures emerge when presenting the gradient by means of absolute (SII) or relative (RII) differences across the income scale. Both measures correspond to a different 'inequality equivalence' when looking at changes, the SII being invariant to equal absolute changes in mortality rates due to Covid-19, whereas the RII is invariant to equal percentage changes across the income scale. The 'choice' of presenting excess mortality as an absolute difference or as a relative change between the baseline years and the Covid-19-period then boils down to the choice of an absolute or relative perspective for the income gradient.

The top panels of Figure 3.3 show excess mortality - expressed in absolute terms - for each household income decile in the male and female populations older than 65. Consistent with the earlier observation that the SII increased during the Covid-19 crisis, excess mortality, measured in absolute differences, is decreasing with income. The differences are substantial. Using the estimated linear income gradients in Appendix Table A3.2, the estimated excess mortality is 326 out of 100,000 in the bottom decile vs. 131 in the top decile for men. The corresponding numbers are 269 vs. 96 for women.[11] This corroborates the argument that the mortality incidence of the Covid-19 crisis falls disproportionately on lower income households. The nuance is that the difference in excess mortality by income is entirely driven by the elderly. In the younger age group the negative income gradient in all-cause mortality has basically remained the same, while in the group of nursing home residents there has been no meaningful relationship between income and mortality, neither before nor during the Covid-19 crisis.

The bottom panels of Figure 3.3 show excess mortality relative to baseline mortality - commonly referred to as P-scores (see Aron and Muellbauer (2020)) - for each income decile within the same subgroups. The relationship between the relative mortality increase and income is less precise and less pronounced overall. This corresponds to the small and insignificant change in the RII for both men and women, providing a new perspective on how much inequality has increased due to the Covid-19 crisis.

[11] The difference in observed excess mortality is even larger, especially for women, as the observed excess mortality in the top decile is an outlier.

Figure 3.3: Income Gradient in Absolute and Relative Excess Mortality March-May 2020



Notes: Panels A-B plot the excess mortality rate by income decile in March-May 2020 for individuals aged 65 or older, excluding people living in collective households, or households with more than 10 individuals. Panels C-D show the excess mortality fraction (P-score) for the same groups of individuals, where the P-Score is defined as excess mortality in 2020 divided by average mortality in 2015-2019 within the associated group.

In principle it should not come as a surprise that choosing to use relative measures like the RII or absolute measures like the SII, can lead to different conclusions. Whereas many authors conclude that the best way out of this uncomfortable choice of measure is to present several of them - illustrated by the numerous other measures described in Mackenbach and Kunst (1997) - others point to the inescapable need to depart from the purely descriptive stance. They plead in favor of making the implicit value judgements in the chosen inequality measure explicit by following a more axiomatic route, inspired by the development in inequality or poverty measurement in the economic discipline.[12] Especially in the health economics context, this more axiomatic approach has been fruitful in unveiling the impact of using bounded variables (like mortality, which is bounded between 0 and 1), or the attractiveness of specific axioms, like the 'mirror axiom'. The latter imposes that, whether one chooses to measure inequality in terms of an 'attainment' (e.g. 'surviving'), or in terms of 'shortfall' (e.g. 'dying'), one should obtain the same inequality ordering in distributional comparisons.[13] When following the index proposed by Erreygers (2009), satisfying the mirror axiom, we again conclude that inequality has increased during the Covid-19 months (see Appendix Table A3.2).

Besides the different normative perspectives, the obvious reason why the choice of measure matters so much empirically is the simple fact that mortality rates are so unequal during the baseline years. Framed differently: due to the strong baseline income gradient of mortality, the impact of the Covid-19 crisis on inequality is less clear cut. While it has not decreased by either of our measures, how much it has increased critically depends on the measurement of inequality.

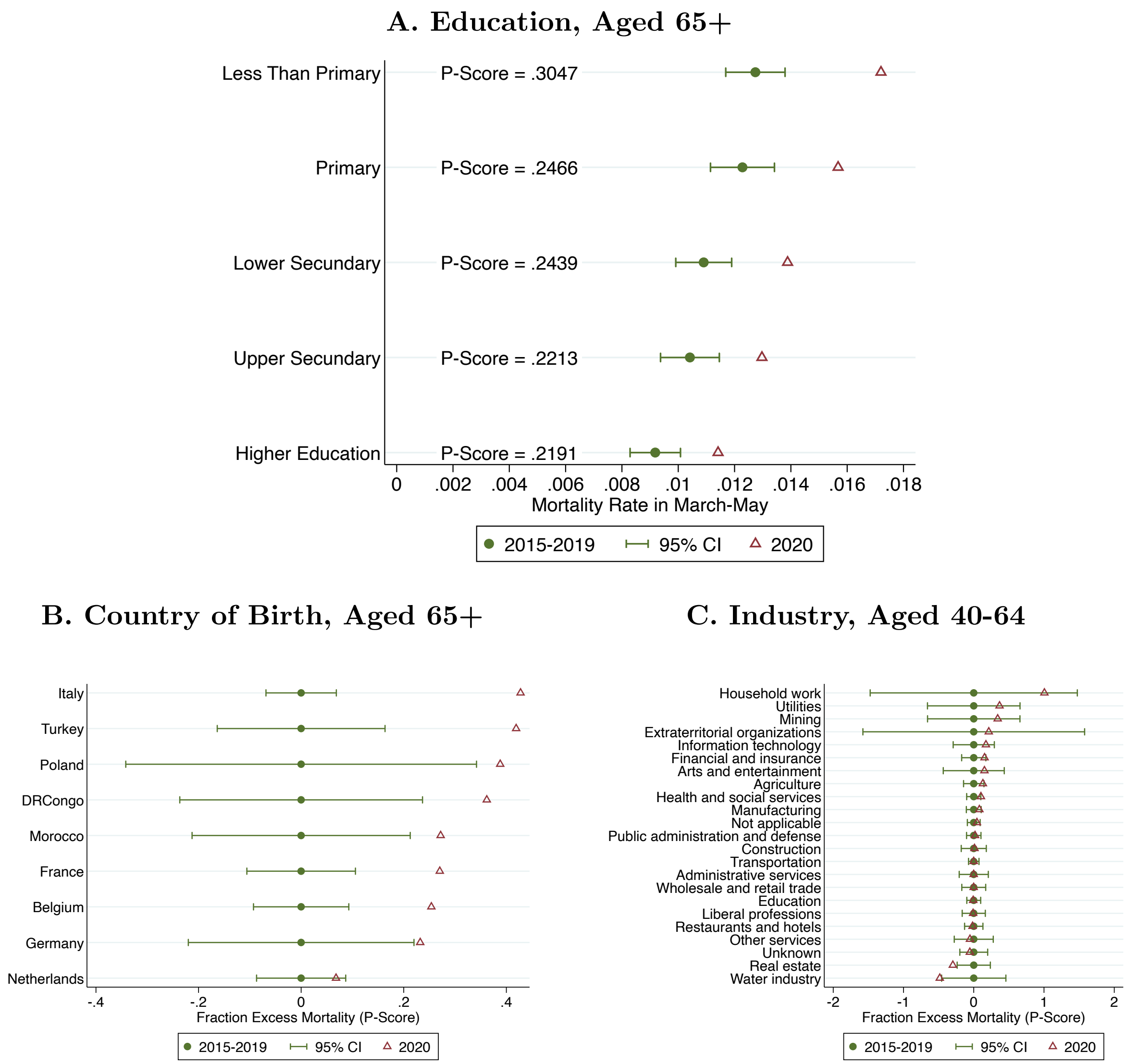
III.C Other Socioeconomic Determinants

An important strand of the literature on socioeconomic differences in health points to education as the go-to indicator of socioeconomic status. The reason for this is both pragmatic and fundamental. Education is often known in survey data, and as education is obtained early in life, it is a arguably less endogenous to health than income as a socioeconomic indicator. Panel A of Figure 3.4 clearly shows how, for the elderly, the negative educational gradient in mortality becomes stronger during the Covid-19 crisis and the change is more pronounced than for the income gradient. Indeed, we find a negative educational gradient in excess mortality during the Covid-19 pandemic, both when expressed in absolute and relative terms. The mortality rate was 30.47% higher in March-May 2020 compared to the baseline years for elderly who did not complete primary school, while for elderly who

[12] This is most markedly pronounced in the title of the paper by Kjellsson, Gerdtham and Petrie (2015) 'Lies, Damned Lies, and Health Inequality Measurements. Understanding the Value Judgements'. The descriptive nature of measures like SII or RII on the contrary, is revealed by labelling the estimated coefficients of the underlying regressions as the *least false parameter* (Moreno-Betancur et al. (2015) p.519), emphasising that these parameters not necessarily correspond to an estimate of a "true" model underlying the data.

[13] As shown by Erreygers (2009) and Erreygers and Van Ourti (2011), imposing the mirror axiom drastically reduces the choice of inequality measures to measures which are 'absolute' instead of 'relative', i.e. inequality is unaffected by equal additions or subtractions of the outcome variable across the income scale. The fact that one cannot satisfy scale invariance, when imposing the mirror principle is easily seen from the fact that a distributional change which keeps the ratio's $m(i)/m(j)$ constant cannot simultaneously keep the ratio $(1-m(i))/(1-m(j))$ constant, where we use the example of mortality rates bounded between 0 and 1.

Figure 3.4: Excess Mortality by Education, Country of Birth and Industry



Notes: Panel A shows mortality rates (with 95% confidence intervals) in March-May 2015-2019 and March-May 2020 by educational level for individuals aged 65 and older. Excess mortality in percentages (P-Score) is also indicated on the figure. Panel B shows excess mortality fractions in March-May 2020 and 95% confidence intervals for 2015-2019 by country of birth for individuals aged 65 and older. Panel C shows excess mortality fractions in March-May 2020 and 95% confidence intervals for 2015-2019 by industry for individuals aged 40-64. Samples in all panels exclude individuals living in collective households, or households with more than 10 individuals. Average mortality rates (also used in the computation of the P-score) are the weighted average of mortality rates by age, where population-based weights are taken for each age. Such a calculation makes sure that there is no influence of age-related composition differences between origins on the plotted mortality rate differences or P-scores.

completed higher education the increase was smaller at 21.91%. For individuals under 65, just like for the income gradient, the relationship between education and mortality remains largely unchanged during the Covid-19 pandemic, as shown in Appendix Figure A3.2.

We briefly consider two other socioeconomic factors in Panels B and C of Figure 3.4:

First, several authors have documented the large burden of the pandemic on minorities in the US and UK (Bertocchi and Dimico (2020), Gross et al. (2020), McLaren (2020), Chowkwanyun and Reed Jr (2020), Price-Haywood et al. (2020) and Chen and Krieger (2020)). While we do not observe race in our data, we do observe country of birth. Panel B of Figure 3.4 shows the relative increases in mortality (P-values) for Belgian residents aged 65+ by country of birth for the 9 most represented countries as country of birth among the elderly in Belgium. The mortality increases among Belgian residents born in Italy (42.77%), Turkey (41.91%) and Poland (38.80%) are larger than among those born in Belgium (25.39%), Germany (23.21%) and Netherlands (6.80%).[14] We also investigate how much of this association is mediated by income. Appendix Figure A3.3 shows coefficients in a regression of excess mortality on country of birth after residualizing on income deciles. In Panel A, we find that the high excess mortality of people of Italian ethnicity is fully explained by differences in income. However, for individuals born in Morocco, Poland and Turkey the excess mortality is high, even conditional on income. That is, the estimated excess mortality is 0.33%, 0.16% and 0.13% higher for individuals born in these respective countries, above and beyond the potential differences in excess mortality explained by their income.

Second, while we do not observe the occupation of workers, we do observe the industry they work in. Focusing on individuals between 40-64 years old, Panel C of Figure 3.4 shows substantial dispersion in the relative increases in mortality across industries, but for none of the industries is the difference between the mortality rate during the Covid-19 crisis and the years before highly significant. This is not too surprising given the lack of significant excess mortality in that age group as a whole. Interestingly, the only sector where we do find marginally significant positive excess mortality is the health and social services sector (10.06%), where workers have arguably been more exposed to the virus. This potential explanation seems to be confirmed when we control for income (see Panel B Appendix Figure A3.3). Conditional on income, the association with excess mortality is highest for individuals working in utilities (2.1%) and in health and social services (1.1%), but lowest in industries that have been shut down during the lockdown, like real estate (-1.8%) and restaurants and hotels (-2.0%).

[14] When looking at excess mortality fractions for individuals aged 40-64 by country of birth in Appendix Figure A3.2, we generally find few groups with significantly positive excess mortality. One exception is the high and significant P-score of 52.79% for 40-64 aged Congolese-born individuals.

Table 3.2: Household Income Gradient vs. Municipality Income Gradient

	Dependent Variable:						
	Indiv. Mortality in March-May (0/1)						Municip. Mortality Rate in March-May
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Household Income	-0.00419***	-0.00101***	-0.00097***	-0.00092***	-0.00094***	-0.00094***	
	(0.00006)	(0.00006)	(0.00006)	(0.00006)	(0.00006)	(0.00006)	
Year 2020 X Log Household Income	-0.00173***	-0.00041***	-0.00025	-0.00019	-0.00021	-0.00023	
	(0.00015)	(0.00015)	(0.00015)	(0.00015)	(0.00015)	(0.00015)	
Log Per Capita Municipality Income				-0.00355***	-0.00406***	-0.00397***	-0.00419***
				(0.00020)	(0.00030)	(0.00031)	(0.00027)
Year 2020 X Log Per Capita Municipality Income				-0.00446***	-0.00221***	-0.00006	-0.00395***
				(0.00054)	(0.00078)	(0.00081)	(0.00082)
Constant	0.05270***	0.02124***	0.02055***	0.06243***	0.07031***	0.06580***	0.05099***
	(0.00058)	(0.00054)	(0.00055)	(0.00187)	(0.00266)	(0.00276)	(0.00266)
Age-Time FE	NO	YES	YES	YES	YES	YES	NO
Municipality-Time FE	NO	NO	YES	NO	NO	NO	NO
Municipality Controls	NO	NO	NO	NO	YES	YES	NO
Number of Cases Control	NO	NO	NO	NO	NO	YES	NO
Observations	12,156,397	12,156,396	11,619,380	11,613,489	11,613,489	11,608,535	3,372
Adjusted R-squared	0.00069	0.01202	0.01219	0.01207	0.01210	0.01211	0.24614

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Columns (1) - (6) regress individual mortality in March-May on log household income, a year 2020 dummy, their interaction as well as other variables depending on the specification. The sample includes individuals aged 65 and older, but excludes people living in collective households, or households with more than 10 individuals. Column (1) is the basic specification. Column (2) adds age-time fixed effects, which are dummies for every age, interacted with year 2020. Column (3) adds municipality-time fixed effects, which are dummies for every municipality, interacted with year 2020. Column (4) uses both log household income as well as log per capita municipality income as controls, together with age-time fixed effects. Column (5) adds demographic municipality controls, including the fraction of 65+ living in single households, the fraction of 65+ that are Belgian-born, the density (inh/km2), and the fraction of 65+ older than 75, as well as the interactions of these with year 2020. Column (6) adds controls for the number of Covid-19 cases per 1000 in each municipality, as well as the interaction with year 2020. Specification (7) regresses the yearly mortality rate among 65+ in March - May of each municipality on log income, a year 2020 dummy, and their interaction. In all columns, only observations from years 2015-2020 are used.

IV Individual vs. Local Effects

Our results so far show that an individual's mortality is highly correlated with his or her household income and that this correlation increased further in the first months of the Covid-19 crisis. An individual's income is, however, related to many other factors, in particular the location that he or she lives in. The pandemic has struck differently across locations with differences in the inflow, propagation and thus exposure to infections, but also with potential differences in access to hospitals and in response to the outbreak of the pandemic.

Most research studying the relationship between income or other socioeconomic factors and mortality during the Covid-19 crisis has been limited by data availability and needed to rely on aggregate measurements at different geographic levels (see Table 3.1). While the geographic inequality in the incidence of Covid-19 and how this correlates with income at the local level is important by itself, one should be cautious when drawing any inference about the role of individual socioeconomic determinants based on geographic variation. This would only be valid in the absence of local factors and geographic sorting on income. We illustrate this in Table 3.2, which reports the estimates from a regression of mortality over the March-May period on log income, allowing the relation to differ in the Covid-19 year 2020. Columns (1) and (7) compare the estimates when running this regressions at the individual vs. municipality level. Regressing individual mortality on log household income in column (1), we confirm the negative gradient we found before and how this negative gradient becomes significantly stronger in 2020. During the baseline years, the relationship between mortality and income is similar when measured at the individual level and the municipality level. However, this negative effect is more pronounced during the Covid-19 crisis when using municipality income than when using household income. In the former case, it almost doubles, while in the latter case, it increases by less than half. Hence, we would drastically overestimate the importance of socioeconomic factors at the individual level for excess mortality when using income measured at the municipality level.[15]

The individual income and mortality data allow us to go further and to separate the relationship between mortality and individuals' income from where individuals live and study how the role of individual vs. local factors changed during the crisis.[16] Column (2) in Table 3.2 repeats the regression of individual mortality on log household income, but includes age fixed effects in line with our graphical results shown before. This reduces the estimated gradient substantially, but simplifies its interpretation as it no longer captures the strong correlation between age and both income and mortality. In column (3) we add municipality times year fixed effects. Controlling for local factors, the estimates of the income effect at the household level remain very similar during the baseline years, but the increase in the income gradient during the crisis decreases and loses significance (p-value = 0.10). The invariance of the estimate in the baseline years when using only within-municipality variation indicates that the income gradient of mortality in Belgium is not driven by location effects. That is, the income gradient

[15] Appendix Table A3.4 repeats the same analysis for the group of individuals aged 40-64. While there are no differences in excess mortality depending on household income, the effect of municipality income during the crisis is marginally significant (p-value = 0.052).

[16] Our decomposition exercise also relates to the separation of selection vs. place effects in explaining the geographic inequality in mortality (Finkelstein, Gentzkow and Williams (2019)).

seems to reflect a relation between mortality and income itself rather than the effects of where individuals with different income live (e.g., accessibility and quality of healthcare). However, this is different during the pandemic. While imprecisely estimated, the smaller interaction terms in columns (2) and (3) suggest that location effects explain about half of the stronger relation between mortality and household income during the Covid crisis. This indicates that location effects have been important during the crisis indeed, but they cannot fully explain the stronger income gradient in mortality either.[17]

We can shed further light on the mechanisms underlying the stronger location effects. Instead of adding municipality fixed effects, column (4) adds average income at the municipality level, allowing again its relation with mortality to differ in 2020. Controlling for household income, individual mortality is higher in municipalities with lower average income and this municipality effect more than doubles during the pandemic. Column (5) in Table 3.2 shows how the municipality income effect during the pandemic is reduced when one controls for other demographic controls at the municipality levels, including the population density, share of elderly, share of elderly living in single households and share of immigrants. Interestingly, the extra effect of municipality income in 2020 disappears when we explicitly control for the number of Covid-19 infections in column (6), suggesting that location effects are important for infections, but not necessarily for case-fatality rates. The relationship between mortality and individual income, however, is robust to the specific controls for local factors. This again illustrates that any inference about individual relationships from analysis at a geographical level is difficult, but particularly so during a pandemic which plays out at the local level.

V Discussion

This paper relates high-quality individual data on mortality to socioeconomic factors and contributes to a better understanding of the impact of the pandemic on the socioeconomic gradient of mortality. We showed that there exists a significant and negative income gradient in excess mortality during the Covid-19-period in Belgium for the elderly. However, this - strongly negative - gradient is comparable to the gradient in all-cause mortality in non-pandemic times. The Covid-19 crisis might stall the trend of narrowing absolute (but not relative) mortality inequality, as documented recently for European

[17] Note that when regressing mortality on income quartiles instead, again allowing for an interaction with a year 2020 dummy, the estimated interaction is also reduced when including municipality-time fixed effects, but the reduction is smaller and the interaction terms remains significant, as shown in Appendix Table A3.3.

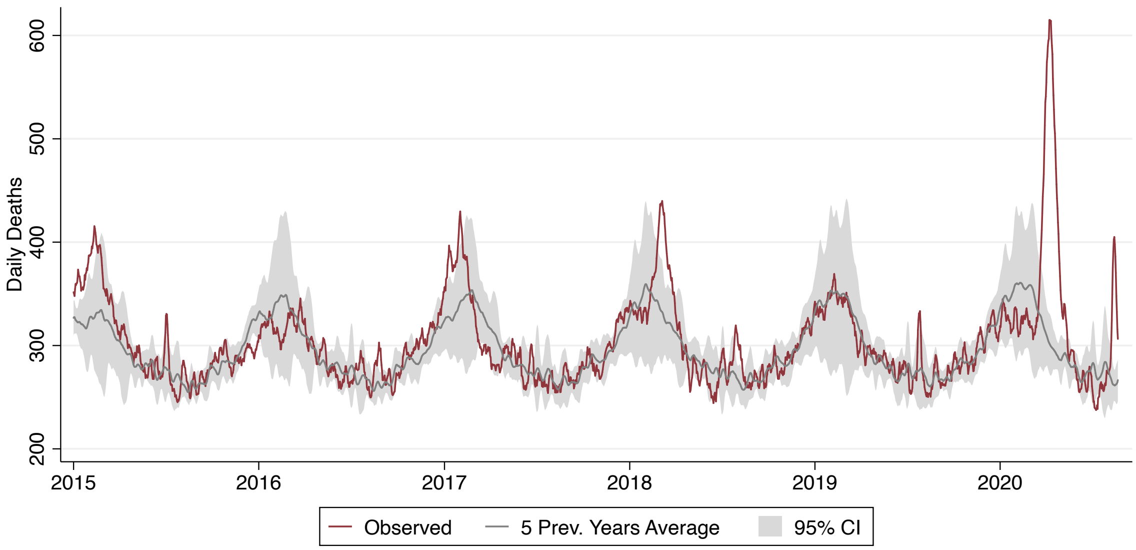
countries in Mackenbach et al. (2019).

The reasons for potential socioeconomic differences and thus a socioeconomic gradient in the incidence and mortality of Covid-19 are heavily debated. Despite the higher likelihood of high-income individuals to import the virus due to international travel, as shown in Pluemper and Neumayer (2020), several papers hint at higher transmission rates among individuals with low socioeconomic status once the illness is widespread within a country (e.g., Desmet and Wacziarg (forthcoming)). Brandily et al. (2020) mention poor housing conditions and higher occupational exposure as the most likely mechanisms causing the higher burden for the poor in France, while McLaren (2020) stresses the importance of higher transit exposure among the less well off. Papageorge et al. (2020) argue that individuals of lower socioeconomic status typically have less flexible work arrangements and a lack of outside space at home, which in turn are correlated with less protection against a pandemic.

Due to the specific data-availability in the Belgian case, our current analysis faces important limitations. First, by using mortality data, we cannot separate the income gradient in infection (e.g., due to differences in employment or social contacts) from the one in case fatality risk (e.g., due to an income gradient in Covid-19 mortality risk factors). Clearly, linking the available data on hospitalizations, prior health diagnoses and test results would allow important progress to be made. Second, by considering mortality, we potentially miss out on important differences in morbidity, physical health, and mental well-being. Again, linking the available data from health records or surveys would allow researchers to provide a more comprehensive picture of the unequal consequences of the ongoing crisis.

VI Appendix Tables and Figures

Figure A3.1: Covid-related Spike in Deaths in March-May 2020



Notes: This plot shows the daily 7-day moving averages of the number of deaths recorded in Belgium. Also plotted is the average daily 7-day moving average of mortality in the 5 previous years, together with 95% confidence intervals.

Table A3.1: Summary Statistics

	40-64	65+	Nursing Home Residents, 65+
	Mean	Mean	Mean
Demographics			
Male	49.97%	44.77%	24.97%
Age	52	75	86
Died in March-May 2020	0.10%	1.19%	9.64%
Education Level			
Missing	11.60%	8.01%	13.44%
Less Than Primary	1.52%	5.23%	7.60%
Primary	6.09%	20.45%	33.90%
Lower Secundary	16.95%	26.80%	23.98%
Upper Secundary	33.15%	20.79%	12.75%
Higher Education	30.69%	18.72%	8.34%
Household Income			
Mean	48,409	34,487	
p10	14,710	15,480	
Median	46,420	29,700	
p90	86,730	61,850	
Personal Income			
Mean	26,389	19,495	18,653
p10	10,440	8,380	12,870
Median	24,090	17,490	17,040
p90	50,000 (capped)	33,420	26,910
Municipality			
Per Capita Income	18,501	18,732	18,477
Number of Residents	66,343	64,415	72,120
Observations	3,740,619	2,130,114	100,829

Notes: This table shows summary statistics for three subsamples of Belgian citizens in 2020. Household and personal income are measured in 2017. Nursing home residents' household income is not included as in our data residents within one nursing home are counted as belonging to the same household. Municipality per capita income and number of residents are measured in 2017.

Table A3.2: Regression and Inequality Estimates

	Aged 40-64		Aged 65+		Aged 65+ in Nursing Homes	
A. Slope Estimates	Men	Women	Men	Women	Men	Women
Linear Regression						
2015-2019	-0.00021	-0.00010	-0.00066	-0.00055	0.00152	-0.00084
	(0.00003)	(0.00001)	(0.00008)	(0.00011)	(0.00106)	(0.00041)
2020	-0.00020	-0.00011	-0.00088	-0.00075	0.00286	-0.00015
	(0.00004)	(0.00002)	(0.00009)	(0.00016)	(0.00202)	(0.00067)
Log-linear Regression						
2015-2019	-0.169	-0.139	-0.061	-0.077	0.030	-0.015
	(0.01)	(0.006)	(0.008)	(0.020)	(0.020)	(0.007)
2020	-0.158	-0.152	-0.067	-0.089	0.031	-0.002
	(0.015)	(0.012)	(0.007)	(0.025)	(0.020)	(0.007)
B. Inequality Measures						
Slope Index of Inequality (SII)						
2015-2019	185	93	596	499	-1368	758
2020	184	100	791	672	-2572	131
Relative Index of Inequality (RII)						
2015-2019	5.30	3.86	1.76	2.05	0.76	1.15
2020	4.71	4.43	1.86	2.31	0.76	1.02
Erreygers-index						
2015-2019	0.0014	0.0007	0.0044	0.0037	-0.0100	0.0056
2020	0.0013	0.0007	0.0058	0.0049	-0.0189	0.0010

Notes: This table provides information on the distributional pattern of mortality in 2015-2019 and in 2020. Panel (A) provides slope estimates and associated standard errors from a linear and log-linear regression of mortality rates on income deciles for both periods separately. Panel (B) shows several measures to evaluate the inequality in mortality in both periods. The calculation of SII - expressed per 100,000 - and RII are based on the estimated slopes in Panel (A).

Table A3.3: Income Gradient Within Municipalities

	Dependent Variable:		
	Mortality in March-May (0/1)		
	(1)	(2)	(3)
Income Q2	-0.0015***	-0.0014***	-0.0014***
	(0.0001)	(0.0001)	(0.0001)
Income Q3	-0.0020***	-0.0020***	-0.0019***
	(0.0001)	(0.0001)	(0.0001)
Income Q3	-0.0044***	-0.0044***	-0.0043***
	(0.0001)	(0.0001)	(0.0001)
Year 2020 X Income Q2	-0.0004	-0.0002	-0.0002
	(0.0002)	(0.0002)	(0.0002)
Year 2020 X Income Q3	-0.0009***	-0.0007***	-0.0007***
	(0.0002)	(0.0002)	(0.0002)
Year 2020 X Income Q4	-0.0014***	-0.0012***	-0.0011***
	(0.0002)	(0.0002)	(0.0002)
Constant	0.0117***	0.0121***	0.0464***
	(0.0001)	(0.0001)	(0.0020)
Municipality-Time FE	NO	YES	NO
Log Municipality Income Control	NO	NO	YES
Observations	12,156,397	11,619,381	11,613,490
Adjusted R-squared	0.0003	0.0005	0.0004

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table regresses mortality in March-May on a year 2020 dummy and on household income quartile dummies, as well as their interactions for individuals aged 65 or older, excluding people living in collective households, or households with more than 10 individuals. Only observations from years 2015-2020 are included. Column (2) adds fixed effects for every Belgian municipality and their interactions with year 2020. Column (3) controls for the log per capita income in each municipality, as well as their interactions with year 2020.

Table A3.4: Household Income Gradient vs. Municipality Income Gradient, Ages 40-64

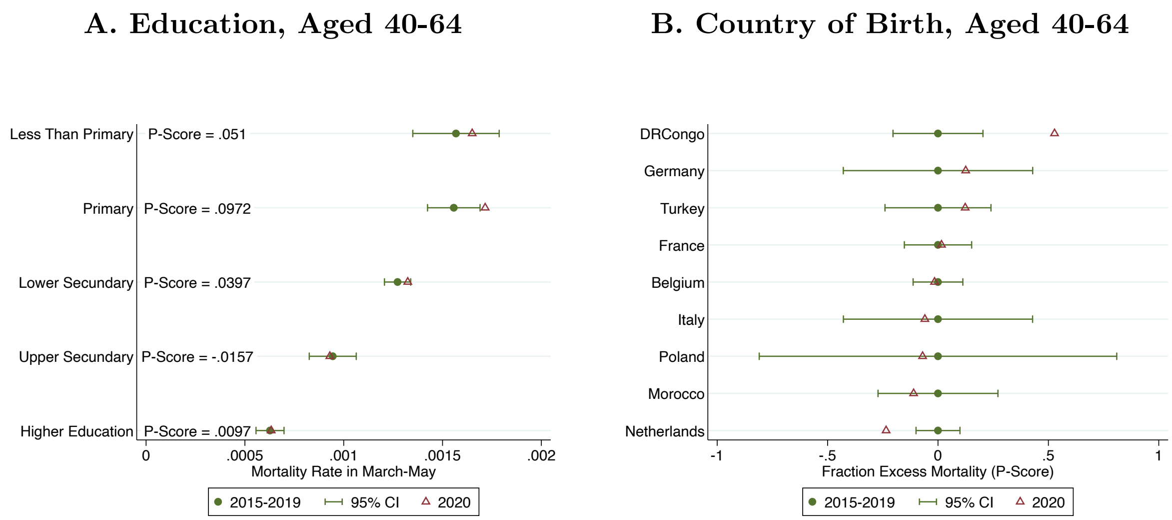
	Dependent Variable:						
	Indiv. Mortality in March-May (0/1)						Municip. Mortality Rate in March-May
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Household Income	-0.00059***	-0.00056***	-0.00054***	-0.00054***	-0.00054***	-0.00054***	
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	
Year 2020 X Log Household Income	-0.00001	-0.00002	-0.00002	-0.00001	-0.00000	-0.00000	
	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)	(0.00003)	
Log Per Capita Municipality Income				-0.00035***	-0.00063***	-0.00068***	-0.00087***
				(0.00005)	(0.00007)	(0.00007)	(0.00008)
Year 2020 X Log Per Capita Municipality Income				-0.00023*	-0.00012	-0.00007	-0.00018
				(0.00012)	(0.00017)	(0.00017)	(0.00025)
Constant	0.00726***	0.00695***	0.00672***	0.01057***	0.01236***	0.01281***	0.00953***
	(0.00013)	(0.00012)	(0.00012)	(0.00041)	(0.00059)	(0.00061)	(0.00080)
Age-Time FE	NO	YES	YES	YES	YES	YES	NO
Municipality-Time FE	NO	NO	YES	NO	NO	NO	NO
Municipality Controls	NO	NO	NO	NO	YES	YES	NO
Number of Cases Control	NO	NO	NO	NO	NO	YES	NO
Observations	20,766,260	20,766,260	20,766,260	20,755,375	20,755,375	20,745,695	3,372
Adjusted R-squared	0.00019	0.00060	0.00064	0.00061	0.00063	0.00063	0.10132

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Notes are similar to Table 3.2. However, in this table, the sample includes individuals aged 40 to 64, and still excludes people living in collective households, or households with more than 10 individuals. Municipality controls now consist of: the fraction of 40-64 year olds living in single households, the fraction of 40-64 year olds that are Belgian-born, the density (inh/km2), and the fraction of 40-64 year olds older than 55, as well as the interactions of these with year 2020.

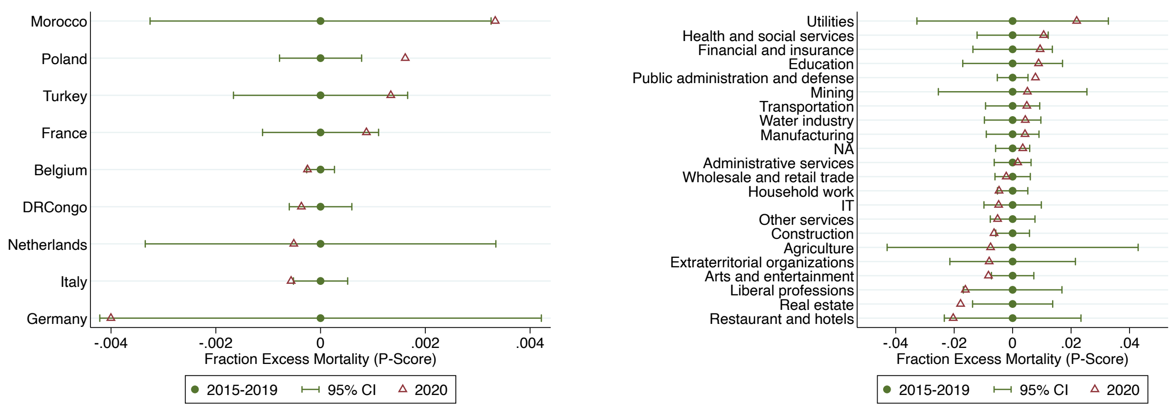
Figure A3.2: Excess Mortality by Education and Country of Birth



Notes: Panel A shows mortality rates (with 95% confidence intervals) in March-May 2015-2019 and March-May 2020 by educational level for individuals aged 40-64, excluding people living in collective households, or households with more than 10 individuals. Panel B shows excess mortality fractions in March-May 2020 and 95% confidence intervals for 2015-2019 by country of birth for individuals aged 40-64 and older, excluding people living in collective households, or households with more than 10 individuals. Calculation of mortality rates and P-score is similar as in Figure 3.4.

Figure A3.3: Excess Mortality by Country of Birth and Industry, Residualized for Income

A. Country of Birth, Aged 65+ **B. Industry, Aged 40-64**



Notes: This Figure is similar to Panel B and Panel C of Figure 3.4, yet here the *residualized* (on income) mortality rates are plotted. Panel A shows these coefficients in March-May 2020 and 95% confidence intervals for 2015-2019 by country of birth for individuals aged 65 and older, after residualizing for household income decile. Panel B shows these coefficients in March-May 2020 and 95% confidence intervals for 2015-2019 by industry for individuals aged 40-64, after residualizing for household income decile. Samples in all panels exclude individuals living in collective households, or households with more than 10 individuals.

Bibliography

Abaluck, Jason, and Abigail Adams. 2019. "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses." Yale Working Paper. 56

Abaluck, Jason, and Jonathan Gruber. 2011. "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." *American Economic Review*, 101(4): 1180–1210. 56

Abaluck, Jason, and Jonathan Gruber. 2016*a*. "Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance." *American Economic Review*, 106(8): 2145–84. 56, 100

Abaluck, Jason, and Jonathan Gruber. 2016*b*. "Improving the Quality of Choices in Health Insurance Markets." NBER Working Paper No. 22917. 56

Abedi, Vida, Oluwaseyi Olulana, Venkatesh Avula, Durgesh Chaudhary, Ayesha Khan, Shima Shahjouei, Jiang Li, and Ramin Zand. 2020. "Racial, Economic, and Health Inequality and COVID-19 Infection in the United States." *Journal of Racial and Ethnic Health Disparities*. 133

Abowd, John, Francis Kramarz, and David Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333. 53, 81

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2020. "Do immigrants assimilate more slowly today than in the past?" *American Economic Review: Insights*, 2(1): 125–41. 5, 6

Abramitzky, Ran, Leah Boustan, Elisa Jacome, and Santiago Perez. 2021. "Intergenerational mobility of immigrants in the United States over two centuries." *American Economic Review*, 111(2): 580–608. 3, 5, 6, 32

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration." *Journal of Political Economy*, 122(3): 467–506. 5, 6

Adams-Prassl, Abi, Teodora Boneva, Marta Golin, and Christopher Rauh. 2020. "Inequality in the impact of the coronavirus shock: Evidence from real time surveys." *Journal of Public Economics*, 189: 104245. 131

Akee, Randall, and Maggie R Jones. 2019. "Immigrants' Earnings Growth and Return Migration from the U.S.: Examining their Determinants using Linked Survey and Administrative Data." National Bureau of Economic Research Working Paper 25639. 2, 5

Algan, Yann, Christian Dustmann, Albrecht Glitz, and Alan Manning. 2010. "The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom." *Economic Journal*, 120(542): F4–F30. 2

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky. 2019. "Regressive sin taxes, with an application to the optimal soda tax." *The Quarterly Journal of Economics*, 134(3): 1557–1626. 57

Andersen, Steffen, John Y. Campbell, Kasper Meisner Nielsen, and Tarun Ramadorai. 2020. "Sources of Inaction in Household Finance: Evidence from the Danish Mortgage Market." *American Economic Review*, 110(October): 3184–3230. 58

Ansala, Laura, Olof Aslund, and Matti Sarvimaki. 2018. "Immigration history, entry jobs, and the labor market integration of immigrants." IFAU - Institute for Evaluation of Labour Market and Education Policy Working Paper Series 2018:20. 2

Ansala, Laura, Ulla Hämäläinen, and Matti Sarvimäki. 2019. "Age at arrival, parents and neighborhoods: understanding the educational attainment of immigrants' children." *Journal of Economic Geography*, 20(2): 459–480. 6

Aron, Janine, and John Muellbauer. 2020. "Measuring excess mortality: the case of England during the Covid-19 Pandemic." INET Oxford Working Paper No. 2020-11. 143

Ashraf, Badar Nadeem. 2020. "Socioeconomic conditions, government interventions and health outcomes during COVID-19." *Covid Economics*, 37: 141–162. 133

Aslund, Olof, Anders Bohlmark, and Oskar Nordstrom Skans. 2015. "Childhood and family experiences and the social integration of young migrants." *Labour Economics*, 35: 135–144. 21, 25

Atkinson, Anthony. 1970. "On the Measurement of Inequality." *Journal of Economic Theory*, 2: 244–263. 55, 96

Aydemir, Abdurrahman, Wen-Hao Chen, and Miles Corak. 2009. "Intergenerational Earnings Mobility among the Children of Canadian Immigrants." *The Review of Economics and Statistics*, 91(2): 377–397. 5

Bachas, Natalie, Peter Ganong, Pascal J Noel, Joseph S Vavra, Arlene Wong, Diana Farrell, and Fiona E Greig. 2020. "Initial Impacts of the Pandemic on Consumer Behavior: Evidence from Linked Income, Spending, and Savings Data." National Bureau of Economic Research Working Paper 27617. 131

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." *Quarterly Journal of Economics*, 130(4): 1623–1667. 51, 67

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018. "The Economic Effects of Social Networks: Evidence from the Housing Market." *Journal of Political Economy*, 126(6): 2224–2276. 57

Bakker, Linda, Jaco Dagevos, and Godfried Engbersen. 2017. "Explaining the refugee gap: a longitudinal study on labour market participation of refugees in the Netherlands." *Journal of Ethnic and Migration Studies*, 43(11): 1775–1791. 6

Banerjee, Abhijit, Amy Finkelstein, Rema Hanna, Benjamin A Olken, Arianna Ornaghi, and Sudarno Sumarto. 2019. "The Challenges of Universal Health Insurance in Developing Countries: Evidence from a Large-scale Randomized Experiment in Indonesia." National Bureau of Economic Research Working Paper 26204. 100

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2018. "Estimating Risk Preferences in the Field." *Journal of Economic Literature*, 56(2): 501–64. 66

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua Teitelbaum. 2013. "The Nature of Risk Preferences: Evidence from Insurance Choices." *American Economic Review*, 103(6): 2499–2529. 56

Bauer, Philipp, and Regina T. Riphahn. 2007. "Heterogeneity in the intergenerational transmission of educational attainment: evidence from Switzerland on natives and second-generation immigrants." *Journal of Population Economics*, 20(1): 121–148. 6

Becker, Gary S, and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of political Economy*, 87(6): 1153–1189. 29

Bertocchi, Graziella, and Arcangelo Dimico. 2020. "COVID-19, Race, and Redlining." *Covid Economics*, 38(4): 129–195. 147

Beshears, John, James Choi, David Laibson, and Brigitte Madrian. 2008. "How are Preferences Revealed?" *Journal of Public Economics*, 92: 1787–1794. 58

Beshears, John, James Choi, David Laibson, Brigitte Madrian, and Sean Wang. 2016. "Who is Easier to Nudge?" Harvard University Working Paper. 58

Beshears, John, James J. Choi, David Laibson, and Brigitte Madrian. 2018. "Behavioral Household Finance." *Handbook of Behavioral Economics*. 58

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. "Choose to Lose: Health Plan Choices from a Menu with Dominated Option." *The Quarterly Journal of Economics*, 132(3): 1319–1372. 51, 56, 57, 100

Bisin, Alberto, and Thierry Verdier. 2000. ""Beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits." *The Quarterly Journal of Economics*, 115(3): 955–988. 7

Bisin, Alberto, Giorgio Topa, and Thierry Verdier. 2004. "Religious intermarriage and socialization in the United States." *Journal of political Economy*, 112(3): 615–664. 7

Böhlmark, Anders. 2008. "Age at immigration and school performance: A siblings analysis using Swedish register data." *Labour Economics*, 15(6): 1366–1387. 21, 25

Bolotnyy, Valentin, and Cristina Bratu. 2018. "The Intergenerational Mobility of Immigrants and the Native-Born: Evidence from Sweden." 2, 5, 6, 32

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience theory of choice under risk." *The Quarterly journal of economics*, 127(3): 1243–1285. 51

Borjas, George J. 1985. "Assimilation, changes in cohort quality, and the earnings of immigrants." *Journal of labor Economics*, 3(4): 463–489. 5

Borjas, George J. 1992. "Ethnic Capital and Intergenerational Mobility." *The Quarterly Journal of Economics*, 107(1): 123–150. 5, 30

Borusyak, Kirill, and Xavier Jaravel. 2017. "Revisiting Event Study Designs." 25

Bouckaert, N., C. Maertens de Noordhout, and C. Van de Voorde. 2020. "Health System Performance Assessment: how equitable is the Belgian health system?" Health Services Research(HSR) Brussels: Belgian Health Care Knowledge Centre (KCE). KCE Reports 334. 136

Brandily, Paul, Clément Brébion, Simon Briole, and Laura Khoury. 2020. "A Poorly Understood Disease? The Unequal Distribution of Excess Mortality Due to COVID-19 Across French Municipalities." *medRxiv*. 132, 133, 151

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad. 2017. "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *The Quarterly Journal of Economics*, 132(3): 1261–1318. 54, 67

Brot-Goldberg, Zarek, Timothy Layton, Boris Vabson, and Adelina Wang. 2021. "The Behavioral Foundations of Default Effects: Theory and Evidence from Medicare Part D." University of Chicago Working Paper. 51, 56, 100, 121

Brown, Caitlin S, and Martin Ravallion. 2020. "Inequality and the Coronavirus: Socioeconomic Covariates of Behavioral Responses and Viral Outcomes Across US Counties." National Bureau of Economic Research Working Paper 27549. 133

Bundorf, Kate, Jonathan Levin, and Neale Mahoney. 2012. "Pricing and Welfare in Health Plan Choice." *American Economic Review*, 102(7): 3214–3248. 56

Calderón-Larrañaga, Amaia, Davide L Vetrano, Debora Rizzuto, Tom Bellander, Laura Fratiglioni, and Serhiy Dekhtyar. 2020. "High excess mortality during the COVID-19 outbreak in Stockholm Region areas with young and socially vulnerable populations." *medRxiv*. 133

Campbell, John, and Luis Viceira. 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press. 51, 66

Campbell, John Y. 2016. "Restoring Rational Choice: The Challenge of Consumer Financial Regulation." *American Economic Review*, 106(5): 1–30. 49

Cardon, James, and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *The RAND Journal of Economics*, 32(3): 408–427. 56

Carmichael, Fiona, Christian K Darko, Marco G Ercolani, Ceren Ozgen, and W Stanley Siebert. 2020. "Evidence on intergenerational income transmission using complete Dutch population data." *Economics Letters*, 189: 108996. 30

Carroll, Gabriel, James Choi, David Laibson, Brigitte Madrian, and Andrew Metrick. 2009. "Optimal Defaults and Active Decisions: Theory and Evidence from 401(k) Saving." *Quarterly Journal of Economics*, 12(4): 1639–1674. 58

Chen, Jarvis T., and Nancy Krieger. 2020. "Revealing the unequal burden of COVID-19 by income, race/ethnicity, and household crowding: US county vs ZIP code analyses." *Harvard Center for Population and Development Studies Working Paper Series*, 19(1). 133, 147

Chen, Jarvis T., Pamela D. Waterman, and Nancy Krieger. 2020. "COVID-19 and the unequal surge in mortality rates in Massachusetts, by city/town and ZIP Code measures of poverty, household crowding, race/ethnicity, and racialized economic segregation." Harvard Center for Population and Development Studies (HCPDS) Working Paper Volume 19, Number 2. 131, 133

Chen, Yiqun, Petra Persson, and Maria Polyakova. 2019. "The Roots of Health Inequality and The Value of Intra-Family Expertise." National Bureau of Economic Research Working Paper 25618. 57

Chetty, Raj, and Adam Szeidl. 2007. "Consumption Commitments and Risk Preferences." *The Quarterly Journal of Economics*, 122(2): 831–877. 67

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*." *The Quarterly Journal of Economics*, 133(3): 1107–1162. 22

Chetty, Raj, John N Friedman, Nathaniel Hendren, Michael Stepner, and The Opportunity Insights Team. 2020*a*. "How Did COVID-19 and Stabilization Policies Affect Spending and Employment? A New Real-Time Economic Tracker Based on Private Sector Data." National Bureau of Economic Research Working Paper 27431. 131

Chetty, Raj, John N. Friedman, Søren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen. 2014*a*. "Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark." *The Quarterly Journal of Economics*, 129(3): 1141–1219. 49, 51, 57

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association Between Income and Life Expectancy in the United States, 2001-2014." *JAMA*, 315(16): 1750–1766. 131, 138, 140, 141

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2020*b*. "Race and economic opportunity in the United States: An intergenerational perspective." *The Quarterly Journal of Economics*, 135(2): 711–783. 1, 3, 4, 5, 7, 16, 29, 32, 34, 35

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014*b*. " Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *." *The Quarterly Journal of Economics*, 129(4): 1553–1623. 7, 29

Chiswick, Barry. 1978. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy*, 86(5): 897–921. 5

Chowkwanyun, Merlin, and Adolph L Reed Jr. 2020. "Racial health disparities and Covid-19—caution and context." *New England Journal of Medicine.* 147

Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *The American Economic Review*, 97(3): 745–788. 56, 66

Cronqvist, Henrik, and Richard Thaler. 2004. "Design Choices in Privatized Social Security Systems: Learning from the Swedish Experience." *American Economic Review*, 94(2): 424–428. 58

Decoster, A., K. Dedobbeleer, and S. Maes. 2017. "Using Fiscal Data To Estimate The Evolution Of Top Income Shares In Belgium From 1990 To 2013." KU Leuven, Department of Economics, Discussion Paper Series DPS17.18. 136

Desmet, Klaus, and Romain Wacziarg. forthcoming. "Understanding Spatial Variation in COVID-19 across the United States." 132, 133, 151

Douven, Rudy, Annette Zeilstra, Anne-Fleur Roos, and Annemiek Verrips. 2019. "Verbeteropties Gemeentepolis." CPB Netherlands Bureau for Economic Policy Analysis CPB Policy Brief. 60

Drefahl, Sven, Matthew Wallace, Eleonora Mussino, Siddartha Aradhya, Martin Kolk, Maria Brandén, Bo Malmberg, and Gunnar Andersson. 2020. "Socio-demographic risk factors of COVID-19 deaths in Sweden: A nationwide register study." Stockholm Research Reports in Demography. 131, 132, 133, 134

Dustmann, Christian, and Tommaso Frattini. 2011. "Immigration: the European experience." *Centro Studi Luca d'Agliano Development Studies Working Paper*, , (326). 5

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2010. "Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market." *Econometrica*, 78(3): 1031–1092. 56

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2015. "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D." *Quarterly Journal of Economics*, 130(2): 841–899. 67

Einav, Liran, Amy Finkelstein, Sendhil Mullainathan, and Ziad Obermeyer. 2018. "Predictive Modeling of U.S. Health Spending in Late Life." *Science*, 360(6396): 1462–1465. 51, 68, 103

Einav, Liran, Amy Finkelstein, Stephen Ryan, Paul Schrimpf, and Mark Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *American Economic Review*, 103(1): 178–219. 56

Enthoven, Alain, Alan Garber, and Sara Singer. 2001. "Near-Universal Coverage Through Health Plan Competition: An Insurance Exchange Approach." *Covering America: Real Remedies for the Uninsured, Editors J. Meyer and E. Wicks*, Washington DC: 155–172. 49

Epple, Dennis, and Richard E. Romano. 2011. "Chapter 20 - Peer Effects in Education: A Survey of the Theory and Evidence." In . Vol. 1 of *Handbook of Social Economics*, , ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 1053–1163. North-Holland. 57

Ericson, Keith. 2014. "Market Design When Firms Interact with Inertial Consumers: Evidence from Medicare Part D." *American Economic Journal: Economic Policy*, 6(1): 38–64. 56

Ericson, Keith Marzilli, and Justin R Sydnor. 2018. "Liquidity Constraints and the Value of Insurance." Working Paper. 51, 67, 79

Erreygers, G., and T. Van Ourti. 2011. "Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice." *Journal of Health Economics*, 30: 685–694. 145

Erreygers, Guido. 2009. "Correcting the concentration index." *Journal of Health Economics*, 28: 504–515. 145

EuroMOMO. 2020. "Graphs and maps." https://www.euromomo.eu (accessed September 11, 2020). 134, 138

Fadlon, Itzik, and Torben Heien Nielsen. 2019. "Family Health Behaviors." *American Economic Review*, 109(9): 3162–91. 57

Falcke, Swantje, Christoph Meng, and Romy Nollen. 2020. "Educational mismatches for second generation migrants. An analysis of applied science graduates in the Netherlands." *Journal of Ethnic and Migration Studies*, 46(15): 3235–3251. 6

Fang, Hanming, Michael Keane, and Daniel Silverman. 2008. "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market." *Journal of Political Economy*, 116(2): 303–350. 56, 57

Finkelstein, Amy, Matthew Gentzkow, and Heidi L Williams. 2019. "Place-based drivers of mortality: Evidence from migration." National Bureau of Economic Research. 149

Finkelstein, Amy, Nathan Hendren, and Erzo Luttmer. 2019. "The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment." *Journal of Political Economy*, 127(6): 2836–2874. 51

Fryer, Roland G, Lisa Kahn, Steven D Levitt, and Jörg L Spenkuch. 2012. "The plight of mixed-race adolescents." *Review of Economics and Statistics*, 94(3): 621–634. 7, 33

Galton, Francis. 1886. "Regression towards mediocrity in hereditary stature." *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15: 246–263. 29

Gross, Cary P, Utibe R Essien, Saamir Pasha, Jacob R Gross, Shi-Yi Wang, and Marcella Nunez-Smith. 2020. "Racial and Ethnic Disparities in Population-Level Covid-19 Mortality." *Journal of General Internal Medicine*, 1–3. 147

Gruber, Jonathan, Benjamin R Handel, Samuel H Kina, and Jonathan T Kolstad. 2020. "Managing Intelligence: Skilled Experts and AI in Markets for Complex Products." National Bureau of Economic Research. 96

Hammarstedt, Mats, and Mårten Palme. 2012. "Human capital transmission and the earnings of second-generation immigrants in Sweden." *IZA Journal of Migration*, 1(1): 4. 6

Handel, Benjamin. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, 103(7): 2643–2682. 51, 121

Handel, Benjamin, and Jonathan Kolstad. 2015*a*. "Getting the Most From Marketplaces: Smart Policies on Health Insurance Choice." *Brookings Hamilton Project Discussion Paper 2015-08*. 56, 96, 100

Handel, Benjamin, and Jonathan Kolstad. 2015*b*. "Health Insurance For Humans: Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review*, 105(8): 2449–2500. 51, 56, 98

Handel, Benjamin, and Joshua Schwartzstein. 2019. "Behavioral Economics and Health Care Markets." *Handbook of Behavioral Economics.* 56

Handel, Benjamin R, Jonathan T Kolstad, and Johannes Spinnewijn. 2019. "Information frictions and adverse selection: Policy interventions in health insurance markets." *Review of Economics and Statistics*, 101(2): 326–340. 56, 98

Hastings, Justine, Chris Neilson, Anely Ramirez, Unika Shrestha, and Seth Zimmerman. 2013. "(Un)informed College Choice: Evidence From Chile." Brown University, Working Paper. 49

Hermansen, Are Skeie. 2016. "Moving Up or Falling Behind? Intergenerational Socioeconomic Transmission among Children of Immigrants in Norway." *European Sociological Review*, 32(5): 675–689. 6

Ho, Kate, Joseph Hogan, and Fiona Scott Morton. 2017. "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program." *The RAND Journal of Economics*, 48(4): 877–905. 56, 122

Ito, Koichiro. 2015. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review*, 104(2): 537–563. 49

Jung, Juergen, James Manley, Vinish Shrestha, et al. 2020. "Coronavirus Infections and Deaths by Poverty Status: Time Trends and Patterns." 131, 132, 133

Ketcham, Jonathan, Claudio Lucarelli, Eugenio Miravete, and Christopher Roebuck. 2012. "Sinking, Swimming, or Learning to Swim in Medicare Part D?" *The American Economic Review*, 102(6): 2639–2673. 56

Ketcham, Jonathan D., Claudio Lucarelli, and Christopher A. Powers. 2015. "Paying Attention or Paying Too Much in Medicare Part D." *American Economic Review*, 105(1): 204–33. 56

Ketcham, Jonathan D., Nicolai V. Kuminoff, and Christopher A. Powers. 2019. "Estimating the Heterogeneous Welfare Effects of Choice Architecture." *International Economic Review*, 60(3): 1171–1208. 56

Kim, Sage J., and Wendy Bostwick. 2020. "Social Vulnerability and Racial Inequality in COVID-19 Deaths in Chicago." *Health Education & Behavior*, 47(4): 509–513. PMID: 32436405. 133

Kjellsson, G., U-G. Gerdtham, and D. Petrie. 2015. "Lies, damned lies, and health inequality measurements: understanding the value judgments." *Epidemiology*, 26: 673–680. 145

Kling, Jeffrey, Sendhil Mullainathan, Eldar Shafir, Lee Vermeulen, and Marian Wrobel. 2012. "Comparison Friction: Experimental Evidence from Medicare Drug Plans." *Quarterly Journal of Economics*, 127(1): 199–235. 51

Knittel, Christopher R, and Bora Ozaltun. 2020. "What Does and Does Not Correlate with COVID-19 Death Rates." National Bureau of Economic Research Working Paper 27391. 132, 133

Koszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4): 1133–1165. 66

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review*, 97(4): 1047–1073. 72, 121

Kroneman, Madelon, Wienke Boerma, M van den Berg, P Groenewegen, J de Jong, and E. van Ginneken. 2016. "Netherlands: Health System Review." *Health Systems in Transition*, 18(2): 1–240. 58

Lemmermann, Dominique, and Regina T Riphahn. 2018. "The causal effect of age at migration on youth educational attainment." *Economics of Education Review*, 63: 78–99. 21

Lubotsky, Darren. 2007. "Chutes or ladders? A longitudinal analysis of immigrant earnings." *Journal of Political Economy*, 115(5): 820–867. 2

Lucassen, Jan, and Rinus Penninx. 1997. "Newcomers. Immigrants and their Descendants in the Netherlands 1550-1995." 8

Mackenbach, J.P., and A.E. Kunst. 1997. "Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe." *Soc Sci Med.*, 44: 757–771. 140, 144

Mackenbach, J. P., J. R. Valverde, M. Bopp, H. Bronnum-Hansen, G. Costa, P. Deboosere, R. Kaledienne, K. Kovács, M. Leinsalu, P. Martikainen, G. Menvielle, M. Rodriguez-Sanz, and W. J. Nusselder. 2019. "Progress against inequalities in mortality: register-based study of 15 European countries between 1990 and 2015." *European Journal of Epidemiology*, 34: 1131– 1142. 131, 151

Madrian, B., and D. Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *Quarterly Journal of Economics*, 141(4): 1149–1187. 51, 58

Manski, Charles. 1999. *Identification Problems in the Social Sciences.* Harvard University Press. 53

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531–542. 81

Maturana, Gonzalo, and Jordan Nickerson. 2018. "Teachers Teaching Teachers: The Role of Workplace Peer Effects in Financial Decisions." *The Review of Financial Studies*, 32(10): 3920–3957. 57

McLaren, John. 2020. "Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census data." National Bureau of Economic Research. 147, 151

Molenberghs, Geert, Christel Faes, Jan Aerts, Heidi Theeten, Brecht Devleesschauwer, Natalia Bustos Sierra, Toon Braeye, Francoise Renard, Sereina Herzog, Patrick Lusyne, Johan Van der Heyden, Herman Van Oyen, Pierre Van Damme, and Niel Hens. 2020. "Belgian Covid-19 Mortality, Excess Deaths, Number of Deaths per Million, and Infection Fatality Rates (8 March - 9 May 2020)." *medRxiv*. 137, 138

Moreno-Betancur, M., A. Latouche, G. Menvielle, A.E. Kunst, and G. Rey. 2015. "Relative index of inequality and slope index of inequality: a structured regression framework for estimation." *Epidemiology*, 26: 518–27. 140, 145

Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity: Why having too little means so much.* Macmillan. 49

Neilsen, Christopher. 2017. "Targeted Vouchers, Competition Among Schools, and the Academic Achievements of Poor Students." Princeton Working Paper. 49

Newhouse, Joseph P. 1993. *Free for All? Lessons From the RAND Health Insurance Experiment.* Harvard University Press. 67

OECD. 2018. *Opportunities for All: A Framework for Policy Action on Inclusive Growth.* OECD Publishing, Paris. 136

OECD/European Observatory on Health Systems and Policies. 2017. *Belgium: Country Health Profile 2017.* OECD Publishing, Paris/European Observatory on Health Systems and Policies, Brussels. 136

Office for National Statistics. 2020. "Deaths involving COVID-19 by local area and socioeconomic deprivation: deaths occurring between 1 March and 31 July 2020." 133

Papageorge, Nicholas W, Matthew V Zahn, Michèle Belot, Eline van den Broek-Altenburg, Syngjoo Choi, Julian C Jamison, and Egon Tripodi. 2020. "Socio-demographic factors associated with self-protecting behavior during the Covid-19 pandemic." *Covid Economics*, , (40): 1–45. 151

Piton, Céline, and François Rycx. 2021. "A Broken Social Elevator? Employment Outcomes of First-and Second-generation Immigrants in Belgium." *De Economist*, 1–47. 2, 6

Pluemper, Thomas, and Eric Neumayer. 2020. "The COVID-19 Pandemic Predominantly Hits Poor Neighborhoods, or does it? Evidence from Germany." *medRxiv*. 151

Polyakova, Maria. 2016. "Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D." *American Economic Journal: Applied Economics*, 8(3): 165–195. 56

Portes, Alejandro, and Min Zhou. 1993. "The new second generation: Segmented assimilation and its variants." *The annals of the American academy of political and social science*, 530(1): 74–96. 2

Price-Haywood, Eboni G, Jeffrey Burton, Daniel Fort, and Leonardo Seoane. 2020. "Hospitalization and mortality among black patients and white patients with Covid-19." *New England Journal of Medicine.* 147

Raifman, Matthew A., and Julia R. Raifman. 2020. "Disparities in the Population at Risk of Severe Illness From COVID-19 by Race/Ethnicity and Income." *American Journal of Preventive Medicine*, 59(1): 137–139. 134

Remmerswaal, Minke, Jan Boone, and Rudy Douven. 2019. "Selection and Moral Hazard Effects in Healthcare." CPB Netherlands Bureau for Economic Policy Analysis Working Paper. 51, 71

Sá, Filipa. 2020. "Socioeconomic Determinants of COVID-19 Infections and Mortality: Evidence from England and Wales." Institute of Labor Economics (IZA) Policy Paper 159. 133

Schippers, Edith. 2016. "Kamerbrief van 24 November 2016 over Onderzoek naar het Vrijwillig Eigen Risico." Ministerie van Volksgezondheid, Welzijn en Sport Ministerial Letter to the Parliament. 50, 60

Shue, Kelly. 2013. "Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers." *The Review of Financial Studies*, 26(6): 1401–1442. 57

Sorenson, Alan. 2006. "Social Learning and Health Plan Choice." *The RAND Journal of Economics*, 37(4): 929–945. 53, 57

Sydnor, Justin. 2010. "Over(insuring) Modest Risks." *American Economic Journal: Applied Economics*, 2(4): 177–199. 51, 56, 121

Tegunimataka, Anna. 2020. "The Intergenerational Effects of Intermarriage." *Journal of International Migration and Integration*, 1–22. 7

Thijssen, Lex, Bram Lancee, Susanne Veit, and Ruta Yemane. 2021. "Discrimination against Turkish minorities in Germany and the Netherlands: field experimental evidence on the effect of diagnostic information on labour market outcomes." *Journal of Ethnic and Migration Studies*, 47(6): 1222–1239. 6

Tubadji, Annie, Don J Webber, and Fred Boy. 2020. "The determinants of the differential exposure to COVID-19 in New York City and their evolution over time." *Covid Economics*, , (13): 50–70. 133

Van Ark, Tamara. 2021. "Kamerbrief van 25 Februari 2021 over Gerapporteerde Onderzoek in 'Slechte eigenrisicokeuzes maken de zorgverzekering duurder voor laagopgeleiden'." Ministerie van Volksgezondheid, Welzijn en Sport Ministerial Letter to the Parliament. 50

van de Beek, Jan, Hans Roodenburg, Joop Hartog, and Gerrit Kreffer. 2021. "Grenzeloze verzorgingsstaat: De gevolgen van immigratie voor de overheidsfinancien." 6

Van den Berg, Gerard J, Petter Lundborg, Paul Nystedt, and Dan-Olof Rooth. 2014. "Critical periods during childhood and adolescence." *Journal of the European Economic Association*, 12(6): 1521–1557. 25

van Kleef, Richard C, Frank Eijkenaar, and René CJA van Vliet. 2019. "Selection incentives for health insurers in the presence of sophisticated risk adjustment." *Medical Care Research and Review*. 59

van Ours, Jan C, and Justus Veenman. 2010. "How interethnic marriages affect the educational attainment of children: Evidence from a natural experiment." *Labour Economics*, 17(1): 111–117. 7, 32

Van Rie, T., and I. Marx. 2014. "Belgium: when growing background inequalities meet resilient institutions." In *Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries' Experiences*. , ed. B. Nolan, W. Salverda, D. Checchi, I. Marx, A. McKnight, I.G. Tóth and H.G. van de Werfhorst, Chapter 6. Oxford:Oxford University Press. 136

Van Winssen, KPM, RC Van Kleef, and WPMM Van de Ven. 2015. "How profitable is a voluntary deductible in health insurance for the consumer?" *Health Policy*, 119(5): 688–695. 56

Van Winssen, KPM, RC Van Kleef, and WPMM Van de Ven. 2016. "Potential determinants of deductible uptake in health insurance: How to increase uptake in The Netherlands?" *The European Journal of Health Economics*, 17(9): 1059–1072. 56

Wiemers, Emily E, Scott Abrahams, Marwa AlFakhri, V. Joseph Hotz, Robert F Schoeni, and Judith A Seltzer. 2020. "Disparities in Vulnerability to Severe Complications from COVID-19 in the United States." National Bureau of Economic Research Working Paper 27294. 134

Williamson, Elizabeth J., Alex J. Walker, Krishnan Bhaskaran, Seb Bacon, Chris Bates, Caroline E. Morton, Helen J. Curtis, Amir Mehrkar, David Evans, Peter Inglesby, Jonathan Cockburn, Helen I. McDonald, Brian MacKenna, Laurie Tomlinson, Ian J. Douglas, Christopher T. Rentsch, Rohini Mathur, Angel Y. S. Wong, Richard Grieve, David Harrison, Harriet Forbes, Anna Schultze, Richard Croker, John Parry, Frank Hester, Sam Harper, Rafael Perera, Stephen J. W. Evans, Liam Smeeth, and Ben Goldacre. 2020. "Factors associated with COVID-19-related death using OpenSAFELY." *Nature*, 584(7821): 430–436. 131, 133

Zorlu, Aslan, and Joop Hartog. 2001. "Migration and Immigrants: The Case of the Netherlands." Tinbergen Institute Tinbergen Institute Discussion Papers 01-042/3. 7, 8

Zorlu, Aslan, and Joop Hartog. 2018. "The Impact of Language on Socioeconomic Integration of Immigrants." Institute of Labor Economics (IZA) IZA Discussion Papers 11485. 6