Mattias Öhman

Essays on Cognitive Development and Medical Care

This Page will be Replaced before Printing

UPPSALA
UNIVERSITET

Abstract Dummy Page.

Dedicated to my friends and family

Acknowledgments

I was never supposed to be an economist. When I moved to Umeå to study computer engineering I thought I was going to end up as a programmer. However, I became more and more interested in social science, and decided to drop out and study sociology in Linköping. But I knew that, to become a good social scientist, you must also take at least one undergraduate course in economics. It was during that course I realized that not all economists are evil, and I learnt that economics is a lot more than interest rates and taxes. So, I switched again, and moved to Uppsala to study economics, and later apply to the PhD programme. To my surprise I was accepted.

To be accepted to the PhD programme and to present a thesis is not the same thing, however. I would not be here today if not for my main supervisor Matz Dahlberg, who has been nothing but encouraging during all these years. I have always learnt something new at our meetings and discussions, and always felt more optimistic afterwards, even when I thought that all hope was lost. Thank you!

I am also very grateful to my co-supervisor, Erik Grönqvist, whom I first met writing my bachelor thesis. Erik really thinks like an economist. While I am glad that I still sometimes can find that somewhat puzzling, I also find it highly intellectually stimulating. I have learnt much from Erik's econometric skills and his knowledge about health economics. Thank you!

I am also grateful to my second co-supervisor, Louise von Essen, for accepting me in the U-CARE interdisciplinary research school, and for her support during these years. Thank you! To meet researchers and PhD students from other disciplines than economics has been stimulating, and I am convinced that increased interdisciplinary collaboration is necessary to improve science.

Thanks to my licentiate discussant Damon Clark, and to my final discussant Ingeborg Waernbaum. Your comments, questions and suggestions have greatly improved my essays.

When I started the PhD programme I did not know what to expect. It soon became apparent that the first-year courses were nothing like the undergraduate courses. I know that I could not have done it without the amazing PhD cohort of 2011. Anna, my fellow health economist in U-CARE, with whom I have struggled to convince psychologists and system developers that economics is not only about taxes. Eskil, without you I know that our first year would have been much harder. Jenny and Ylva, you actually dared to visit dMz. Johannes, for teaching us all about pensions. Jonas and Sebastian, for showing us that hard work pays off in the end. And last but not least, my

roommate Linuz. Without you this thesis would look a lot different – not only because you are one of my co-authors, but also because we have shared more laughs and "duh" than it is possible to remember. I think it is safe to say that we have the best decorated office in Ekonomikum. Thank you all, it would not have been the same without you!

It is not only my own PhD cohort who makes the department running smoothly. Thanks to the administration, Katarina, Stina, Ann-Sofie and Åke, for helping us with traveling expenses, computers and whatnot. All colleagues at the Department of Economics have also contributed to this experience. Thank you, Jon, Sebastian E, Linna, Jacob, Fredrik, Kristin, Lovisa, Gunnar, Dagmar, Georg, Maria, Adrian, Daniel, and everyone else for all lunches, fika, conferences and discussions. A special thanks must go to Gabriella and Evelina for their excellent choice of field, health economics.

I must also thank Per Engström, for letting me, Linuz and Ylva to hold a guest lecture on the philosophy of science, the history of economics, and modern research in economics for the undergraduate students once each semester. It has been both fun and inspiring, and I hope that it can continue.

Not everything that matters is about economics. Fortunately, I have many friends who reminds me of that by being completely uninterested when I talk about some identification issue for the fiftieth time. You are too many to mention here, so instead of thanking each and everyone of you, I dedicate this thesis to you. Perhaps I will now talk about something else. But do not count on it. I must, however, give a special thanks to Linn, who has heard about my troubles more times than anyone should have to endure.

I am very grateful to my family for all the love and support. You have always been supportive with whatever I wanted to do, regardless of it being playing video games, learn computer programming or do a PhD in economics. You have always been there for me.

One of my lecturers in Linköping, Jan Lindvall, once convinced me to keep studying economics by saying that I think more like an economist than a sociologist. It feels strange to contemplate how different my life would have been without those words.

Uppsala, October 2016
Mattias Öhman

Contents

Introduction

> *The aim of this study is to construct a model of the demand for the commodity "good health".*
>
> — Grossman (1972)

This thesis consists of four self-contained chapters, all of which are related to empirical health economics. Some readers might be quite surprised to hear that an evaluation of the effects on well-being of a screening for the disease abdominal aortic aneurysm can be considered economics. In this introduction, I hope to be able to explain how and why. I will briefly discuss the history of economics, how economists think about health, and how health economics is related to epidemiology. The introduction also includes an overview of the four chapters. I conclude with some final thoughts about economics as a discipline.

First, what is economics? For people outside the profession, the answer to that question might feel obvious. Economics is about taxes, economic growth, interest rates, unemployment, and so forth. However, for economists, the question has become increasingly harder to answer due to the developments within the field during the recent decades. Jacob Viner (1892-1970) is famously credited with the quote *"Economics is what economists do"* (Backhouse and Medema 2009). And perhaps, if one were to describe economics today and how it has changed since Adam Smith (1723-1790) published *An Inquiry into the Nature and Causes of the Wealth of Nations* in 1776, that may be the best definition one could come up with. In the two following sections I will discuss the history of economics and health economics.

1 A (very) brief history of economics

Adam Smith is often seen as the father of economics as a discipline with the aforementioned book *The Wealth of Nations*. Economic thought, however, is much older. One of the earliest works on "economics" is the Socratic dialogue *Oeconomicus* by the Greek philosopher Xenophon (431-404 BC), in which he discusses household management and agriculture. In his work *Cyropaedia*, Xenophon discusses the division of labor and the importance of market size; in small towns, the same individual must do everything by himself, while in larger cities, it is possible with specialization (Sandelin et al. 2008).

Much of the early economic thought was normative. The scholastic school, which flourished in Europe around the 11th century, for example, was interested in the "just price" and condemned the practice of charging interest. The focus changed somewhat with the mercantilists, which dominated the economic thought in Europe during the 15th to the 18th century. For the mercantilists, it was important to maximize the national wealth by accumulating precious metals. One of the ways to do so was to subsidize exports and have high tariffs on imports (Sandelin et al. 2008). The physiocratic school, which was developed around the 18th century, criticized the mercantilists' focus on the rulers' wealth. The physiocrats were perhaps the first school to see labor as the source of value, but according to them, that applied only to agricultural labor. With Adam Smith, all of this changed.

One of the revolutionary thoughts that Adam Smith had was that the welfare of the society can be maximized if individuals' are allowed to pursue their own interests, through the *Invisible hand* (Evensky 1993).[1] In contrast to the physiocrats, it was not only agricultural labor that created value. Instead of the mercantilists' focus on collecting gold and silver, individuals had a role to play.

With the turn to neoclassical economic thought, at the end of the 19th century, the individual was suddenly the *only* actor. The society consists of individuals and is a mere aggregate of the behaviors of these agents (Sandelin et al. 2008).[2] The neoclassical economists were pure "microeconomists", as compared with the earlier "macroeconomic" focus. Joseph Schumpeter (1883-1950), invented the term *methodological individualism* to describe this methodological view (Hodgson 2007). Further, individuals were assumed to be *rational* in the sense that they act to maximize their own utility or happiness. The break from the classical economics, developed by Smith, to the neoclassical economic theory is often called *the marginal revolution*. Marginalism allowed economic analysis, among other things, to be based on more sophisticated mathematical ground.

During the 20th century, much of the economic theory has been founded on neoclassical thought, with its marginalistic approach and rational utility maximizing individuals. The analysis became even more mathematical. This has sometimes been criticized, but has also allowed formulation of clear hypotheses which can be empirically tested.

Around 1960, economists began to study topics traditionally belonging to sociology using economic theory. Gary Becker (1930-2014) was one of the first to do so. Becker and Jacob Mincer (1922-2006) studied what they called "human capital", a term to describe an individuals' knowledge, skills, and

[1] As discussed in Evensky (1993), Smith did not believe that this would happen automatically. The success of creating a liberal society depended on the individuals' adherence to a common social ethics.

[2] This is not to say that the state does not exist, or that it could not have some role to play, but that the analysis must start with individuals.

health as a kind of capital.[3] According to the theory of human capital, individuals invest in, for example, education, to receive higher wages. If it would not be profitable in a life-cycle perspective, individuals would not invest. Becker applied this thinking not only to education, but to questions such as crime and drug addiction. Using the theory of human capital, economists began to invade other fields. Since the Chicago economist Edward P. Lazear[4] (1948-) invented the term it has become common to call this *economic imperialism* (Lazear 2000).

Simultaneously with the evolution of economics, statistical tools to test economic theories against empirical data were developed. A new field called econometrics, closely related to both economics and statistics, emerged (Boumans and Davis 2010). While the goal in statistics typically is prediction, econometricians test causal claims.[5] For example, fundamental in economic theory is the well-known laws of supply and demand, but it is a non-trivial task to estimate these functions. This is because they generally depend on common variables. In econometrics, this is the so-called *identification problem*. The same problem arises every time individuals can choose what to do (self-selection). Econometricians began to develop methods that allowed causal inference to address this problem. This development of the econometric thinking is sometimes called the "credibility revolution".[6]

My view of this development is that since economists assume that agents are rational and utility maximizing, the focus on the problem of self-selection – which is one of the biggest threat to any causal claim – comes naturally. When an economist want to understand the effects of, for example, a health insurance program, the first question that arises is not what is done in the program, but why an individual takes part of that program. The methodological individualism and rational choice view begs the question of *why* this is "utility maximizing" for the individual. This is not to say that we are free to choose our destiny (on the contrary, budget restrictions – not only monetary – are an important part of economic theory), but it raises legitimate questions on the possibility of heterogeneous effects, self-selection and moral hazard. The later occurs when an individual takes more risk because the cost is taken by someone else, such as in an insurance program.

[3] Even though Becker and Mincer may have been the most successful popularizers of the term, I have found that it has been used since 1916, at least, but with a different meaning (Boag 1916).

[4] Now at Stanford Graduate School of Business.

[5] This is perhaps to give the statistics literature to little credit. In fact, it was statisticians that analyzed randomized experiments and formulated the now dominant view in econometrics, the potential outcomes framework (Imbens and Wooldridge 2009). There are also many statisticians today that are interested in causality, such as Donald B. Rubin, Paul R. Rosenbaum and Tyler VanderWeele. However, while causality is the focus for *some* statisticians, it is the focus for *all* econometricians.

[6] The term "credibility revolution" is used by Angrist and Pischke (2010) in a comment to the critique by Leamer (1983) against the empirical work of that time. They argue that the methods of causal inference today are so developed that the "con" is taken out.

As the economic imperialism met the credibility revolution, we had economists who were not afraid of stepping into other fields, equipped with a toolbox of well-developed econometric methods. These economists claimed to be able to give causal answers to questions that they argued that others had not been able to give. Today, much of the empirical economics published explicitly or implicitly builds on the potential outcomes framework and the experimentalist approach (Angrist and Pischke 2009).

2 Economics and health

Since health is such an important part of the human capital, economists need to understand how an individuals' demand for health looks like, and how the health care market functions.

Health care is an interesting market for economists in itself.[7] There is a large literature on cost-benefit analysis, aimed at evaluating the best choice of medical treatments when there are at least two options to choose from and no strictly preferred option. However, a seminal article by Grossman (1972) had a different focus. Grossman claimed that "health" was an investment, in principle not different from other goods, and developed a model for the demand of health capital.[8] Health depreciates over time, so to stay healthy, an individual must keep investing in health. The efficiency of the production of health depends on variables that modify the price of health capital. For example, investments are more effective for highly educated people. The so-called Grossman model is still today the workhorse model for health economists.

Health investment models and empirical findings indicating that early life health is important for later labor market outcomes, have drawn economists' attention to "fundamental" factors such as cognitive and non-cognitive skill. Cognitive skill is what we usually call IQ, or intelligence, while non-cognitive skill refers to personality and emotional traits (Cunha et al. 2010).[9] The first two chapters in this thesis focus on these skills, either as outcomes or as explanatory variables, which explains the first part of the thesis title. There is a large and growing literature in economics studying these skills. One of the main findings is that non-cognitive skill is, at least, as important as cognitive skill. Both skills have been shown to be important predictors of future outcomes (see references in related chapters).

[7] See for example Arrow (1963), in which Kenneth Arrow (1921-) studied the role of asymmetric information in medical care, which has been cited over 7,000 times!

[8] As should be clear from the earlier discussion, Michael Grossman (1942-) was not the first to see health as human capital, see for example Mushkin (1962). He was, however, the first to construct an investment model of health.

[9] Non-cognitive skill is sometimes called "socioemotional skills". "Skills" and "abilities" are used interchangeably in the literature, but conceptually, "ability" refers to an innate capacity, while "skill" is something that can be trained.

Sometimes it is not ethically possible to conduct a randomized controlled trial to estimate the effect of a medical treatment. Here, the economists' knowledge of causal inference with observational data comes in handy. To study questions on, for example, well-being, economic theory is not always necessary. In fact, the evaluation instruments used are primarily from psychology and other disciplines. The identification problem is, however, the same. The last two chapters in this thesis concerns medical care, and these two chapters explain the second part of the title. Even if the two chapters do not build on economic theory, the two different methods that I use are common in economics to solve the identification problem.

2.1 Relationship to epidemiology

If economists sometimes leave economic theory and only use the econometric tools to answers questions on health with observational data, what is the difference between health economists and epidemiologists?

In a sense, empirical health economics is relatively close to epidemiology, at least the part of the literature that study mortality and outcomes of that sort. My view is that, indeed, health economists could very well be mistaken for epidemiologists if one only look at the questions studied. However, epidemiologists and economists do not use the same tools, and have different languages.

At the core, there is a fundamental difference; while the economists have an experimentalist approach to questions, epidemiologists are "model builders". My view of the differences is that, in practice, economists search for exogenous variation. If such can be found, he or she carries on and use this variation to answer the question at hand. The mechanisms at work are, somewhat, a black box. Epidemiologists search for credible mechanisms, but are not as concerned as the economist of finding exogenous variation. If there is a plausible mechanism, the epidemiologist tries to answer the question at hand. This is reflected in the Hill's criteria for causation (Hill 1965), published by epidemiologist Bradford Hill (1897–1991).

My understanding of these two different approaches is that they seem to originate from the two different traditions that we come from. Economists are worried about rational utility maximizing individuals who self-selects into treatment – which is why we need exogenous variation so that we can control how individuals choose – while the epidemiologist has a background in medical science, and is more concerned of the mechanisms at hand. The economist often lacks deep knowledge of the variables included in the regressions, but has a good knowledge on how to measure a causal effect. The epidemiologist has the medical knowledge, but in practice often settles with studying associations.

3 The chapters

By now, the reader should have the necessary background to understand how and why the chapters in this thesis are economics. In this section I give a short overview of each chapter.

The Effects of Fluoride in The Drinking Water

The thesis begins with studying a topic that has received a lot of attention in recent years: The effects of fluoride in the drinking water on cognitive ability.

There has been an intense public debate on the effects of fluoride in the water since many countries, such as the United Kingdom and the United States, fluoridate their water. There is an enormous amount of evidence that fluoride improves the dental health, speaking in favor of fluoridation, even if some argue that it is an involuntary treatment of the population.[10] However, a meta-study published in 2012 found that higher fluoride levels in the water is associated with lower cognitive ability, which sparked a new round of debate regarding fluoridation (Choi et al. 2012).

From an economist point of view, the studies reviewed all had very small samples, and lacked credible identification strategies. In this chapter, my co-author and I use the rich population-wide registers in Sweden combined with data on the fluoride levels in the Swedish drinking water to study the effects on health and labor market outcomes. Since many municipalities use more than one water source, the fluoride level differs randomly between relatively small geographical areas. This allows us to interpret our results causally.

We do not find any evidence of negative effects on cognitive or non-cognitive ability for the fluoride levels in the Swedish drinking water. We find *positive* effects on dental health, income and employment. Possibly, the effects on income and employment can be explained by the positive effects on dental health, in line with what has been suggested in earlier literature (Glied and Neidell 2010).

Be Smart, Live Long: The Relationships between Cognitive and Non-Cognitive Abilities and Mortality

In this chapter, I study the associations between cognitive and non-cognitive abilities and mortality. Economists have become increasingly more interested in early life health capital accumulation, as it has been shown to have large effects later in life, for educational attainment, labor market outcomes, criminal behavior, and so forth (Cunha et al. 2010; Lindqvist and Vestman 2011). The motivation behind this study is to see if these abilities are related with a severe outcome such as mortality. I also look at how good income and education capture these underlying skill measures. This is interesting as it is common in economics to use income and education as proxy measures for these skills.

[10] This is why fluoridation of the water has not been allowed in Sweden since the 70's.

The Swedish military enlistment (abolished 2009) measured the cognitive and non-cognitive skills of all enlisted. Enlistment was mandatory for all Swedish men at age 18-20.[11] Using register data, I have a population-wide dataset of about 700,000 men born between 1950 and 1965. I follow these individuals up till year 2009 and measure all-cause mortality, with and without controlling for income and education.

I find that both skills indeed are strongly associated with mortality. The earlier epidemiological literature has focused only on cognitive ability, and has therefore lacked an important skill dimension. However, the associations for both abilities are heterogeneous. I find no associations with mortality for individuals with college education or for those being at least a middle-income earner. For non-college educated and low-income earners, on the other hand, the associations are strong. The results suggest that income and education are inadequate as proxy measures for individuals in the lower part of the distributions.

Health Information and Well-Being: Evidence from an Asymptomatic Disease

The two earlier chapters revolved around cognitive development. The second part of this thesis turns to the medical care. In this essay, we examine how unexpected information about the health affect the well-being. How individuals react to health information (whether it may be positive or negative) is a highly debated topic, especially regarding screening-programs. Is it worth the cost, considering the potentially negative effects on well-being for individuals who receive information that they have a disease they did not know about before?

We study a specific screening-program for an asymptomatic disease, Abdominal Aortic Aneurysm (AAA), to which all 65-year-old men in Sweden are invited. There is an ongoing debate whether this screening-program should continue or not. The prevalence of the disease is, compared to the number of invited individuals, low, and even for those who have an aneurysm, the probability of dying of other causes is high (Johansson et al. 2015). On the other hand, since AAA is asymptomatic you are not aware of it, and if the aorta ruptures you are likely to die within a few minutes. However, surgery is only conducted on large AAAs. So what are the effects on well-being by being informed of that you are in the risk of having an aneurysm, but that you will not be treated?

Using the regression discontinuity (RD) design, we can estimate the causal effects on well-being of receiving this information. We find only very small and statistically insignificant effects on well-being. For the individuals who have a small aneurysm, we find *positive* effects on well-being. Why? These individuals will be under increased surveillance, and one possible explanation is that this has a calming effect. In a cost-benefit analysis of the screening-

[11] However, this practice was not enforced the end years of the enlistment.

program, the effects on well-being would therefore rather be on the benefit side than the cost side.

Myocardial Infarction, Antidepressants and Mortality
In the last chapter, I study another highly debated topic, the effects of antidepressants on mortality.

Depression has become a common illness in the western world the last decades. One consequence of this is that antidepressant medications have become among the most commonly used drugs in the world. In Sweden, almost 10 percent of the population use antidepressants. But do they have negative, potentially dangerous, side effects?

Depression is common among individuals who have experienced a myocardial infarction (MI), commonly known as heart attack (National Institutes of Health 2015). The most common treatment today is antidepressants. However, it has been established in the literature that the old tricyclic antidepressants (TCA) have cardiac effects and is contraindicated for MI patients. The newer SSRI antidepressants is considered to be more safe. But even so, some studies find that SSRI antidepressants may increase the risk of mortality (e.g. Tata et al. 2005). Most studies on this subject are either relatively small randomized trials, or large observational studies that only study associations. There is therefore a need for large-scale studies using methods that allow for a causal interpretation.

I use a matching technique.[12] The aim is to find a "statistical twin" for individuals in the treated group (individuals who receive SSRI antidepressants) in the untreated group. If there are no important unobserved characteristics, the difference between the treated and untreated groups can be interpreted as the causal effect of antidepressants on mortality. The Swedish health care quality registers are very rich on health variables, which allow for a credible use of a matching method strategy.

After analyzing several different matching specifications, I find no statistically significant effects on mortality of antidepressants, which suggest that SSRI antidepressants are, in this respect, safe to use for MI patients.

4 Concluding thoughts

Almost two decades ago, Lazear claimed that "[b]y almost any market test, economics is the premier social science" (Lazear 2000). Is he correct?

As a graduate student in economics, I may not be in the position to give an unbiased answer to this question. If we are to trust the revealed preference theory, this seems to be a common view amongst economists (Fourcade

[12]The method I use is called Propensity Score Matching. Rosenbaum and Rubin (1983) show that we can reduce the dimensionality problem of many variables down to a one-dimensional problem by using the likelihood of treatment instead of exact matching.

et al. 2015). I will conclude this introduction with some thoughts on this question.

As I have shown, economics has evolved quite a bit since the days of Adam Smith. Today, economists do not only study questions related to the national (or individual) economy; economics is a broad social science. It is true that economics and economists receive a lot of attention. Some positive, some negative. In that sense, economics is certainly the winner of the market test. Economic theory and methods are sophisticated tools to study a long range of questions. Empirical economists today are as much econometricians as economists, and we can formalize clear hypotheses and test them using methods that allow causal inference.

However, this does not necessarily mean that the answers from economic theory is always true, or that we blindly should trust our estimates. Deep institutional knowledge of the questions are needed for trustworthy answers. The economic imperialism, promoted by Lazear, cannot be without consideration of what is already known in other fields, and economists should not disregard the theories and methods of other disciplines without careful examination of the evidence. My feeling is that this is not always done.

In my studies, I have benefited – and depended – a lot from researchers in other fields. If not for them, I would not be as confident of the results as I am today.

References

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, New Jersey: Princeton University Press. ISBN: 9780691120348.

———. 2010. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." *Journal of Economic Perspectives* 24 (2): 3–30.

Arrow, Kenneth J. 1963. "Uncertainty and the welfare economics of medical care." *The American Economic Review* 53 (5): 941–973.

Backhouse, Roger E, and Steven G Medema. 2009. "Retrospectives on the definition of economics." *Journal of Economic Perspectives* 23 (1): 221–233.

Boag, Harold. 1916. "Human capital and the cost of the war." *Journal of the Royal Statistical Society* 79 (1): 7–17.

Boumans, Marcel, and John B Davis. 2010. *Economic Methodology: Understanding Economics as a Science.* Palgrave Macmillan. ISBN: 9780230218130.

Choi, Anna L, Guifan Sun, Ying Zhang, and Philippe Grandjean. 2012. "Developmental fluoride neurotoxicity: A systematic review and meta-analysis." *Environmental Health Perspectives* 120 (10): 1362–1368.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78 (3): 883–931.

Evensky, Jerry. 1993. "Retrospectives: Ethics and the invisible hand." *Journal of Economic Perspectives* 7 (2): 197–205.

Fourcade, Marion, Etienne Ollion, and Yann Algan. 2015. "The superiority of economists." *Journal of Economic Perspectives* 29 (1): 89–114.

Glied, Sherry, and Matthew Neidell. 2010. "The economic value of teeth." *The Journal of Human Resources* 45 (2): 468–496.

Grossman, Michael. 1972. "On the concept of health capital and the demand for health." *Journal of Political Economy* 80 (2): 223–255.

Hill, Austin Bradford. 1965. "The environment and disease: Association or causation?" *Proceedings of the Royal Society of Medicine* 58 (5): 295.

Hodgson, Geoffrey M. 2007. "Meanings of methodological individualism." *Journal of Economic Methodology* 14 (2): 211–226.

Imbens, Guido W, and Jeffrey M Wooldridge. 2009. "Recent developments in the econometrics of program evaluation." *Journal of Economic Literature* 47 (1): 5–86.

Johansson, Minna, Anders Hansson, and John Brodersen. 2015. "Estimating overdiagnosis in screening for abdominal aortic aneurysm: Could a change in smoking habits and lowered aortic diameter tip the balance of screening towards harm." *British Medical Journal* 350:h825.

Lazear, Edward P. 2000. "Economic imperialism." *The Quarterly Journal of Economics* 115 (1): 99–146.

Leamer, Edward E. 1983. "Let's take the con out of econometrics." *The American Economic Review* 73 (1): 31–43.

Lindqvist, Erik, and Roine Vestman. 2011. "The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment." *American Economic Journal: Applied Economics* 3 (1): 101–128.

Mushkin, Selma J. 1962. "Health as an investment." *Journal of Political Economy* 70 (5, Part 2): 129–157.

National Institutes of Health. 2015. "What is a heart attack?" Accessed September 16, 2016. `http://www.nhlbi.nih.gov/health/health-topics/topics/heartattack`.

Rosenbaum, Paul R, and Donald B Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70 (1): 41–55.

Sandelin, Bo, Hans-Michael Trautwein, and Richard Wundrak. 2008. *Det Ekonomiska Tänkandets Historia.* SNS Förlag. ISBN: 9789185695591.

Tata, LJ, J West, C Smith, P Farrington, T Card, L Smeeth, and R Hubbard. 2005. "General population based study of the impact of tricyclic and selective serotonin reuptake inhibitor antidepressants on the risk of acute myocardial infarction." *Heart* 91 (4): 465–471.

I. The Effects of Fluoride in the Drinking Water

Co-authored with Linuz Aggeborn

Acknowledgments: We would like to thank Erik Grönqvist, Eva Mörk, Matz Dahlberg, Mikael Elinder, Ronny Freier, Kaisa Kotakorpi, Melanie Luhrmann, Mattias Nordin and Adrian Adermon for helpful discussions, comments and suggestions, as well as Robin Djursäter, Liselotte Tunemar, Tomas Byström, Gullvy Hedenberg, Louise von Essen and John Wallert. We would also like to thank seminar participants at the Department of Economics at Uppsala University, Geological Survey of Sweden (SGU), U-CARE, the 72th IIPF conference in Lake Tahoe, the EEA conference in Geneva and IFAU. We gratefully acknowledge financial support from U-CARE.

1 Introduction

It is well-established that fluoride strengthens the tooth enamel and that application of fluoride on the surface of the teeth prevents caries, tooth decay and cavities. The use of fluoride in a wide range of dental products is therefore considered as an important mean to improve dental health. Because there is such a well-defined link between fluoride and healthy teeth, some countries artificially fluoridate the drinking water so that people are continuously exposed to higher levels than the natural level. Australia, Brazil, Canada, Chile, Malaysia, the United Kingdom and the United States are a few examples of countries that apply such a public policy (Mullen 2005). Other countries, such as Sweden, do not fluoridate the water, but the authorities choose not to reduce the fluoride level in the water cleaning process as long as it is below a certain limit. These public policies are, however, debated. Fluoride is deadly at high levels, and there is an emerging and much discussed epidemiological literature of potential negative side effects of long-term fluoride exposure for lower levels on the central nervous system. The hypothesis is that fluoride might function as a neurotoxin.

In comparison to dental products, drinking water containing fluoride is ingested, meaning that everyone drinking water is exposed to fluoride continuously for a long period of time. In this paper we investigate the causal effect of fluoride exposure through the drinking water on cognitive and non-cognitive ability, education and later labor market outcomes. We also study the long-established link between fluoride and dental health. To further investigate the effect of fluoride, we look at other health outcomes that may be connected to fluoride. We use a unique register dataset from Sweden together with drinking water fluoride data, where we exploit intra-municipality variation in fluoride to estimate the effect.

Earlier epidemiological studies have found evidence of negative side effects of fluoride, and the results have sparked a public debate regarding the potential dangers associated with fluoride in the water (e.g. Johnston 2014 in The Telegraph; Mercola 2013 in The Huffington Post).[1] A meta-study by Choi, Sun, et al. (2012) from Harvard School of Public Health reviewed 27 papers and concluded that exposure to high dosages of fluoride is associated with a reduction of almost half of a standard deviation in IQ among children.[2] The

[1] One indication that people tend to be very concerned with fluoridation is found in Lamberg et al. (1997). The local authorities in Finland decided that water fluoridation should stop at a given date, and this decision was communicated to the inhabitants. However, water fluoridation ceased one month earlier without notification to the public, but people still reported various symptoms in a survey.

[2] See Tang et al. (2008) for an earlier meta-study, which also show a negative relation between fluoride and IQ. Epidemiological papers published after or around Choi, Sun, et al. (2012) include Ding et al. (2011), Saxena et al. (2012), Seraj et al. (2012), Nagarajappa et al. (2013), Ramesh et al. (2014), Khan et al. (2015), Sebastian and Sunitha (2015), Kundu et al. (2015), Choi, Zhang, et al. (2015), Das and Mondal (2016) and Dey and Giri (2016) who all found

data from the reviewed papers originated from China and Iran. Several of these papers considered very high levels of fluoride which surpasses the recommendation from the World Health Organization (WHO) that fluoride should not exceed 1.5 mg/l in the drinking water (WHO 2011, p.42). However, some of the studies reported negative effects on cognitive development for levels below the recommended level. This is a cause for concern because these levels are present naturally in the drinking water in many parts of the world. Countries that fluoridate the drinking water also have fluoride within this range. Common problems with the studies reviewed by Choi, Sun, et al. (2012) are that the analyses were based on small samples with poor data quality, and without clear identification strategies.[3]

Our paper is to our knowledge the first to study the effects of fluoride in a large-scale set-up with individual register data. We have access to a rich panel of Swedish register data which enables us to investigate the effect of fluoride in a more credible way and with a much larger sample than earlier studies. Sweden has a natural variation of fluoride in the drinking water which stems foremost from the bedrock under the water sources. The fluoride level in our data is hence not endogenous to any policy decision. The fluoride level in the Swedish drinking water ranges between 0 and 4 mg/l in our dataset, and there is often variation within municipalities which we exploit to estimate the casual effect. In comparison to China and Iran, Sweden has a well-supervised water supply system, meaning that other drinking water hazards that can affect cognitive development are not likely to be present. Fluoride in Sweden is generally not considered to be a problem unless the level exceeds 1.5 mg/l.[4] Since our data include a variation in fluoride in the lower spectra, our results are more policy relevant for countries that artificially fluoridate the drinking water, because water authorities seldom add fluoride so that the level exceeds 1.5 mg/l. There is no evidence of any differences between artificially fluoridated drinking water and water with a natural occurrence of fluoride (Harrison 2005; John 2002), meaning that our results should be valid for countries with comparable artificial fluoride levels.

As economists, we are interested in the connection between fluoride, cognitive and non-cognitive ability, education and labor market outcomes for at least two reasons. First, fluoridation of the drinking water is a common public health program, and it is important that the effectiveness of such a policy is

or discussed negative effects of fluoride on IQ. Additionally, Malin and Till (2015) found a positive association between fluoridated water and the prevalence of ADHD in the U.S.. See also Li et al. (2016) for a study on fluorosis and cognitive impairment.

[3]There are some studies that point in the other direction. Broadbent et al. (2015) follows approximately 1,000 individuals in an observational study from New Zeeland. The authors find no negative effect on IQ from living in an area in the city of Dunedin with artificial fluoridation. The main critique against this study is that artificial water fluoridation may be an endogenous policy variable.

[4]The absolute majorities of the Swedish water plants has fluoride levels below 1.5 mg/l.

evaluated. Second, economists have in an increasing degree become interested in early determinants of health and human capital, and its long-run effects on labor market outcomes. Our paper is connected to this literature on human capital development where we study a treatment that millions of people are affected by all over the world: fluoride in the drinking water.

Our results confirm the positive link between fluoride and dental health. However, in contrast to earlier studies, we find a zero-effect of fluoride on cognitive ability, non-cognitive ability and education (measured by test scores on a national math test). We also find a zero-effect on related health outcomes. Our point estimates with regard to cognitive ability are much more precisely estimated compared to earlier studies and always close to zero. We find evidence that fluoride is a positive factor for later labor market outcomes, which indicates that better dental health is a positive factor on the labor market.

The rest of the paper is organized as follows. In the next section we review related papers, followed by a short medical background for why fluoride might have an effect on the central nervous system. Next, we provide a simple conceptual framework on how we should think about fluoride in the drinking water as a public health policy. Our identification strategy is mainly based upon the variation in fluoride which stems from an exogenous variation in the bedrock, so in section 5, we present the necessary geological background and information on how we have mapped drinking water data to the individuals. In section 6, we describe our data material. Our identification strategy and econometric set-up are discussed in section 7 followed by descriptive statistics in the same section. The empirical results are then presented, next robustness checks and lastly our conclusions. Additional results and figures are presented in the appendix.

2 Earlier literature

In this section we review the literature regarding early determinants for health and their long-run effects. We explicitly focus on papers that have studied drinking water.

Currie (2011) provides an excellent overview of this research field with a special emphasis on determinants at birth and in utero. Economists acknowledge that health during childhood is an important determinant for success on the labor market (Currie 2009). Case, Lubotsky, et al. (2002) and Currie and Stabile (2003) provide evidence for the connection between health and socioeconomic status. Case, Fertig, et al. (2005) present the conclusion that health during one's early years seems to be connected to (among others) socioeconomic status and one's education once becoming an adult. Smith (2009) has also demonstrated this link empirically, and found that poor health before age 16 is negatively associated with future income, wealth and labor supply.

Cognitive development is part of individuals' health, and earlier research have shown that cognitive ability and non-cognitive ability are very adequate explanatory variables for basically everything that we consider as positive individual labor market outcomes (e.g. Heckman et al. 2006, Lindqvist and Vestman 2011). Cunha and Heckman (2007) create a theoretical model concerning cognitive and non-cognitive ability and Cunha and Heckman (2009) emphasize that there are "critical" and "sensitive" windows when cognitive and non-cognitive abilities are more affected by environmental factors. See also Cunha, Heckman, and Schennach (2010). According to the authors both cognitive and non-cognitive ability are very important factors for later achievements in life. This view is confirmed in Lindqvist and Vestman (2011) and Öhman (2015), who use the results from the Swedish draft tests for cognitive and non-cognitive ability and show that they are very good predictors for education, income and mortality. If fluoride has negative effects on cognitive development, this adds a piece to the puzzle why some individuals are more successful than others on the labor market.[5]

We are not aware of any other paper that has employed large individual register datasets to estimate the effect of fluoride on cognitive development specifically. In a recent unpublished paper, Heck (2016) studies the effects of water fluoridation on health and education with U.S. survey data. He finds that fluoridated water prevents caries in deciduous teeth, but no effects on education and general health. A limitation in this study is that education is measured only at the county level. The main critique is that water fluoridation is a result of a policy choice, making the identification less clear.

Earlier papers in economics have focused on other potential hazards and their effects on health and cognitive ability. Currie, Graff Zivin, et al. (2013) study the effect of mothers' consumption of polluted drinking water (broadly defined) during pregnancy on birth weight of the offspring with data from New Jersey. They find that the birth weight is negatively affected by contaminated water for mothers with a low education. Zhang (2012) uses Chinese data to study the effect of providing monitored and safe drinking water from a water plant to the population. The author finds a positive effect on the ratio of weight and height for both children and adults and some evidence of less illness among adults.[6] Galiani et al. (2005) study whether privatization of water supply in Argentina improved water quality, and find that children mortality decreased if an area was provided with drinking water from a private provider. Feigenbaum and Muller (2016) study lead and explicitly how people were treated with lead originating from the drinking water pipes. The authors study homicide incidence and find a positive effect of lead, i.e., an increased incidence of homicide.

[5] A seminal paper by Grossman (1972) presents a framework for individual health investment. Fluoride may affect an individual's health before he or she can make an active investment choice.

[6] The author briefly discuss fluoride in the Chinese drinking water but do not study this explicitly.

Aizer et al. (2016) study reductions of lead levels in Rhode Island for cohorts born between 1997 and 2005. They use variation in lead in buildings due to policy implementations as an instrument, and find significant positive effects on children's reading test score in third grade for lower lead levels. Lead has also been studied with regards to air pollution. Nilsson (2009) investigates the long-term effects of lead on labor market outcomes. The author uses time variation from the time period when lead in gasoline was reduced together with Swedish geographical data on lead levels in the environment, and concludes that a reduction in lead exposure in early life has positive effects on cognitive ability, education and labor market outcomes. In a similar paper, Grönqvist et al. (2014) conclude that the reduction in lead exposure also reduce criminal behavior. Other economic papers have studied air pollution in general. Schlenker and Walker (2015) study pollution from airports in California and find that prevalence of respiratory deceases, heart diseases and asthma increase among the inhabitants, especially among children and older people, if carbon monoxide emission increases. In Jans et al. (2014) the authors study air pollutants' effect on child health. Periods of inversions seems to affect children from high-income families 40 percent less than children from low-income families.

It might be that fluoride in the drinking water has negative side effects on cognitive ability, but the net effect on income still is positive because the effect on dental health is so large. Glied and Neidell (2010) found that women living in areas whose water was fluoridated had higher incomes, where the effect seems to be stronger according to the authors for those with a poor socioeconomic status.[7]

3 Medical background

In this section we shortly review the medical discussion about fluoride and its effects on health.

Sodium fluoride (NaF), from now on called fluoride, is a toxic compound which exists naturally in the environment. WHO acknowledge a deadly dose of fluoride to be about 5-10 grams depending on the body weight (Liteplo et al. 2002, p.100). Fluoride intake from the drinking water is absorbed and transmitted throughout the blood system (Fawell et al. 2006, p.29-30). When large amounts of fluoride are ingested it has a number of toxic effects on the body. For example, approximately 100,000 individuals in the Assam region in India have been taken ill with kidney failure stiff joints and anemia and

[7] Näsman, Ekstrand, et al. (2013) also apply Swedish drinking water data, but from an earlier time period. Cohorts born between 1900 and 1919 are included in their study where the authors study the effects on hip fracture incidence. The authors find no indications that fluoride induces hip fractures. Näsman, Granath, et al. (2016) use the same dataset to study the effects on myocardial infarctions and find no effects on this outcome either.

as a result of very high natural levels of fluoride in the water (WHO 2015). Gessner et al. (1994) discuss a case in Alaska where individuals in a small village accidently were exposed to extremely high levels of fluoride (up to 150 mg/l) due to a malfunctioning water pump. One individual died and many became very ill as a result of fluoride poisoning.

Water fluoridation is a highly debated issue (Richards 2002; Peckham and Awofeso 2014). Researchers have called for more research on the subject, where Grandjean and Landrigan (2014) argue for a global initiative for more research on potential neurotoxins, including fluoride. Mullenix et al. (1995) was one of the first papers testing the hypothesis that fluoride exposure also has effects on the central nervous system. The researchers exposed randomly selected rats to different fluoride treatments (including fluoridation of the drinking water), and concluded that the rats' brain tissue can store fluoride and that fluoride can pass through the blood-brain barrier. They found that a higher concentration of fluoride in the brain tissue induced behavioral changes meaning that fluoride functions as a neurotoxin in rats. Chioca et al. (2008) also conducted laboratory rat experiments and concluded that high exposure of fluoride through the drinking water induced impaired memory and learning. Whether fluoride can pass the blood-brain barrier in humans is debated. Chioca et al. (2008) state that a one-time high consumption of fluoride does not seem to pass the blood-brain barrier. Hu and Wu (1988), however, found fluoride to be present in the cerebrospinal fluid, which surrounds the brain among humans. Consuming water with fluoride is an example of a long-term consumption and the question is whether this consumption of fluoride can pass the barrier.

Lower dosages of fluoride have, on the other hand, beneficial effects on dental health (see Griffin et al. (2007) and Twetman et al. (2003) for reviews). For that reason, fluoride is added to toothpaste and in some countries to the drinking water. Fluoride is also present naturally in tea leaves and in low concentration in the food (Liteplo et al. 2002, p.5).

Given that fluoride is both a lethal and dangerous compound at higher dosages, and improves dental health at lower dosages, it is important to find the optimal level. There has been a consensus that fluoride only has adverse effects above the threshold level of 1.5 mg/l (WHO 2004). In light of recent epidemiological findings reviewed in Choi, Sun, et al. (2012) this threshold could be questioned.

4 Conceptual framework

We present a simple and short conceptual framework in this section on how we can think about water fluoridation as a public policy.

Fluoride is a potential neurotoxin that may have a negative effect on cognitive ability, but is known to have a positive effect on dental health. The

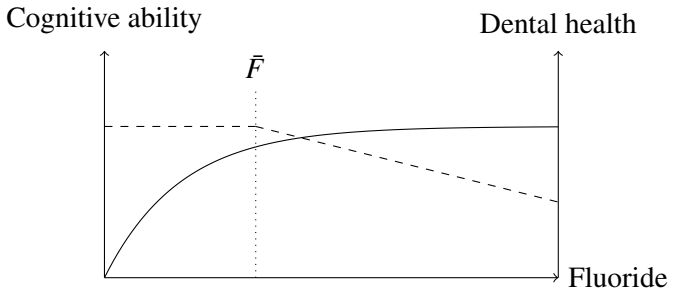


Figure 1. The effects of fluoride on dental health (solid line) and cognitive ability (dashed line).

policy maker must decide on the cost-benefit of fluoridation in comparison to other alternatives. For example, fluoridation of the water can be less expensive than publicly subsidized dental checkups and teeth repairs, thus making it an effective public policy.

It is on the one hand unlikely that the general public would accept fluoridation if it is dangerous for the health in any known way. On the other hand, for economists, the optimal level of fluoride is where the marginal cost equal the marginal benefit. If the positive effect on dental health is very large with only a very small negative effect on cognitive ability, the net effect could still be positive.

Figure 1 illustrates the policy makers problem in a single figure.

The effects of neurotoxins often take the form of a hockeystick where exposure above a certain level becomes dangerous (Nilsson 2009). The effect of fluoride on dental health on the other hand probably follows a concave function where the marginal benefits on fluoride become smaller for higher levels. We investigate whether $\bar{F}$ exists in the Swedish drinking water. Based on this, it is possible to do a cost-benefit analysis of the optimal fluoride level if the fluoride level is found to have a negative effect on human capital development. If the fluoride level is not found to have a negative effect on human capital development for the levels of fluoride we consider, the cost-effectiveness of water fluoridation may instead solely be evaluated based on the effects on dental health and the cost of fluoridation. This is possible because countries that fluoridate the water normally do not add more than the WHO recommendation of 1.5 mg/l. To find whether $\bar{F} < 1.5$ mg/l is also important for countries with no artificial fluoridation since they may reduce the fluoride level in the water cleaning process.

5 Exogenous variation in fluoride: Geological background

In this part of the paper we discuss how fluoride varies exogenously in Sweden. We also discuss how we map the drinking water data to individuals' place of residence.

The natural level of fluoride in the drinking water depends on geological characteristics, especially the type of bedrock under a water source (SGU 2013, p.81). Fluoride is both tasteless, without odor and without any color for the levels we consider in this paper, implying that individuals cannot know whether they are drinking water with lower or higher levels of fluoride (WHO 2001).

There are different types of bedrock, providing different levels of fluoride to the water. Soil bedrock is associated with lower levels of fluoride in comparison to stone bedrocks such as granite. Greywacke bedrock also yields higher levels of fluoride. Especially water from drilled bedrock wells usually contains higher levels of fluoride (SGU 2013, p.81,84). Rainfall usually contains low levels of fluoride (Edmunds and Smedley 2013, p.313).[8] In Sweden, water sources are situated on different types of bedrock, thus yielding different fluoride levels. For a detailed description about fluoride and its natural geological occurrence, see Edmunds and Smedley (2013) and SGU (2013).

The fluoride level is, from our perspective, an exogenous variable that is constant for a very long time because the bedrock is constant. Hence, the water authorities have no possibility to manipulate the natural levels of fluoride in raw water. The water authorities may reduce the fluoride levels in the water cleaning process, but this is not done in Sweden unless the level exceeds 1.5 mg/l.[9]

Each municipality in Sweden is responsible for the public drinking water. Because municipalities often have different water sources situated on different types of bedrock, there is a within-municipality variation in fluoride.[10] Each municipality in Sweden is divided into several SAMS (Small Areas for Market Statistics) by Statistics Sweden. We make use of these SAMS when we estimate the effect of fluoride. A SAMS consists of approximately 750 individuals in the year 2011, with median 520. There are almost 9,300 SAMS in Sweden in comparison to 290 municipalities.[11] The large majority in Sweden drinks water from the municipal water plants. However, some individuals

[8] One of the main sources of fluoride in rain is volcanic emissions (Edmunds and Smedley 2013, p.314), but there are no active volcanoes in Sweden.

[9] In our data collecting process from the Swedish municipalities, nothing indicates that water authorities lowered the fluoride if it was below 1.5 mg/l.

[10] Augustsson and Berger (2014) show that there is a variation in the fluoride level in private wells in Kalmar county in Sweden.

[11] The reader should note that SAMS are not something that the public in general is aware of. Municipalities, however, are administrative areas that exist in the publics mind.

Table 1. *Bedrock analysis*

	F. (0.1 mg/l)
Mix of stone and soil bedrock	2.983***
	(0.526)
Stone bedrock	4.085***
	(0.214)
Constant	3.057***
	(0.129)
R^2	0.1729
Observations	1,788

Notes: The dependent variable is fluoride which is expressed in 0.1 mg/l. Standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The benchmark is "soil bedrock". The analysis is based on the entire SGU dataset.

have private wells for which we do not have data. Approximately 1.2 million people of Sweden's total population of approximately 10 million drink water from private wells (Livsmedelsverket 2015).

We have information on fluoride levels for the outgoing drinking water from the water plants supervised by the municipalities. There are 1,726 water plants in our final data where we have manually designated a coordinate for the water plant based on the supplementary information we have from SGU and from the municipalities (our two data sources for the fluoride data, we return to our data sources in the data section below). We also have information about the bedrock for the corresponding water source for the water plants. The variable is categorical where bedrock is classified into three broader categories: Soil bedrock, a mix between soil bedrock and stone bedrock and stone bedrock.

In Table 1 we verify that the fluoride level in the drinking water depends on the bedrock. The benchmark bedrock is soil bedrock and we include dummies for the other two categories. It is clear that the mixed bedrock as well as the stone bedrock yields higher fluoride levels in comparison to soil bedrock, which is exactly what we expect. Note that these three categories include different subtypes of bedrock (granite, greywacke et cetera) meaning that there is variation in fluoride within each category.

Some municipalities do not have a water plant within its borders. These municipalities have been dropped from the analysis together with those municipalities where we do not have any information regarding fluoride. In total, data from 261 municipalities are included. We know in which SAMS an individual lived for a given year, but we cannot observe the exact geographical coordinate for the location where the individual lived within a SAMS.[12] Thus, we need a mapping protocol for how to assign fluoride data for each SAMS.[13]

[12] Such data would abolish the anonymous structure of the Swedish individual register data, since population address registers are public information in Sweden.

[13] Since we cannot observe the exact location within a SAMS, we cannot distinguish on the household level who drinks the water from the municipal water plants and the private wells. We return to this issue in the robustness analysis.

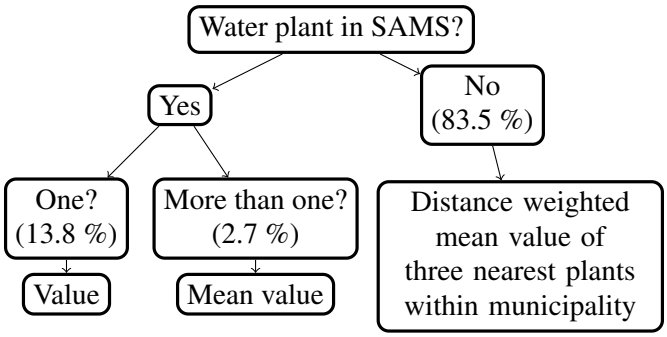


Figure 2. Water plants mapping. Percentage of SAMS in parenthesis.

We map the fluoride level to SAMS using the mapping protocol illustrated in Figure 2. We indicate the share of SAMS in each category in parenthesis.

For SAMS that have a water plant within the borders we assign the fluoride level of that water plant to all individuals that lived in the area. If there is more than one water plant within the SAMS border, we take the mean fluoride level. For SAMS without a water plant within the borders, we calculate the geographical center point of the SAMS, and assign a mean of the fluoride level for the three closest water plants (triangular polygon) using the inverse distance as a weight. We assess this mapping protocol by first looking at the effect of fluoride on dental outcomes for which we expect to see an effect of fluoride. By looking at dental health measures, we also address whether the variation in fluoride in our data is enough to estimate effects.

Figure 3a displays the raw variation in fluoride for those SAMS with a least one water plant. White areas are thus SAMS without a water plant. Figure 3b shows the variation in fluoride between SAMS after our mapping.

6 Data

In this section we present the data material.

In short, we have register data at the individual level for all outcomes and covariates except dental health. The dental health data is only available on the SAMS level for each cohort from age 20 for the years 2008 and 2013, and comes from The National Board of Health and Welfare. We observe place of residence for all individuals of age 16 and older on the SAMS level.[14] In order to track individual's place of residence before age 16 we link them to their parents, and use the mother's place of residence as a proxy. Our treatment period for fluoride consumption spans between birth and up to the year when

[14]For some individuals and years, SAMS codes are missing. We have imputed SAMS codes from $t-1$ or $t+1$ in these cases if municipal code is the same.

(a) SAMS with at least one water plant (b) Final mapping

Figure 3. Mapping of fluoride data.

we measure the outcome variable.[15] We include cohorts born between 1985 and 1992 in our data.

6.1 Fluoride data

Fluoride data is measured for each water plant, and there are in total 1,726 water plants supervised by the municipalities in our data set. This data comes from two sources: Drinking water data from Swedish Geological Survey (SGU) and drinking water data from the municipalities. We use the SGU data or the municipal data depending on which data set that has the earliest available drinking water data for a given municipality. The SGU data starts in 1998. For some municipalities data is only available for later years.[16] We have contacted each of Sweden 290 municipalities to complement the SGU data set. We asked the municipalities to provide us with additional data from 1985. If data were not available, we asked them whether they have changed any of their water sources since 1985.[17]

It should be noted that the fluoride level is constant back in time because the bedrock has not changed. The fluoride level should only be different if (1) the municipality has changed the water source (which is rare), or, (2) installed any purification for fluoride (which they do not do unless the level exceeds 1.5 mg/l). We collapse the fluoride data into a single measure for each water plant, meaning that we take the average when we have data from several years for a water plant. Variation between the years should be due to variation in the measurement validity for individual data points, meaning that an average measure is more accurate. The reader should note this means that for the very few cases where purification has been installed, we take the average for *all* years available.[18] We drop all individuals who have ever lived in a municipality between birth and age 16 for which we do not have fluoride data. We

[15] There are some inconsistencies in the register data. For example, we have dropped all individuals with multiple birth years, duplicate observations, individuals not in both the LOUISE database and the multigenerational database. We also drop individuals that have immigrated to Sweden during childhood since we need to track their fluoride level from birth. Their parents may, however, have immigrated before the individual's birth.

[16] We only use the observations from the SGU data regarding drinking water and not the observations for "raw water".

[17] Not all municipalities have kept their statistics from 1985 and some have not been able to answer our questions. In the robustness analysis, we rerun all specifications but only include municipalities where we are sure that they use the same water source since 1985.

[18] In 2003, the Swedish Food Agency abolished the possibilities to give exceptions for fluoride levels above 1.5 mg/l to 6 mg/l. There were fewer than 100 water plants before 2003 with a median level higher than 1.5 mg/l (Persson and Billqvist 2004). Those plants provided water to approximately 0.26 % of the Swedish population (Svenskt Vatten 2016). After 2003, there is a single limit set to 1.5 mg/l (SGU 2013, p.82). 1.3 mg/l to 1.5 mg/l yielded a note prior of 2003, but was considered safe and did not result in general purification of the water. Children below half a year old was recommended to drink such water with moderation.

choose age 16 because this is the age for which me measure our first outcome variable.

6.2 Individual level data

The data for the individuals originates from several sources which we briefly discuss in this section.

As an outcome for education we use results from the national test taken at age 16. We focus on the basic points result on the math test. This is due to two reasons. First, this is the variable where we have the most detailed data, and, second, it should be a fairly good proxy variable for cognitive ability. The data comes from Statistics Sweden (SCB). We have results for those born in 1987 and later.

The cognitive and non-cognitive ability measures come from the Swedish military enlistment. For more detailed information about the enlistment, see Öhman (2015). Conscription was obligatory for men between 18-20 years old in Sweden until its abolishment in 2009. Those who declined their call to conscription were punished; however, this practice was not enforced in the end years of the Swedish draft. Conscription involved testing of cognitive and non-cognitive ability and the individual's physical health. Cognitive ability was measured by a test where the purpose was to measure the underlying intelligence, often called the *g* factor. This was done by using four sub-tests: verbal, spatial, logical and technical knowledge. The overall test score was then standardized into a single measure on a scale between 1 and 9, according to a Stanine scale. The non-cognitive ability was assessed by a psychologist during a half-hour interview with the conscript. The psychologist's goal was to evaluate the person's ability to function in a war scenario. Those who were keen to take initiative and who were well-balanced emotionally ended up with a high score. The psychologist also considered the individual's ability to deal with stressful situations. The overall assessment was a score according to the Stanine scale. Öhman (2015) shows that both these measures are good predictors for individual outcomes later in life. We only include men born before 1988 when estimating these outcomes since we only have access to this data for those years.

In the end years of the Swedish enlistment, there was a theoretical possibility of strategic manipulation of test results. Individuals who scored low on the tests were not always forced to do military service meaning that the incentives to perform well were less clear for later cohorts. However, the Stanine distribution is relative to others enlisting in the same cohort, so we should still be able to capture meaningful differences in cognitive ability and non-cognitive ability within a cohort (see Figure A2 in the appendix). We can also test this by looking at the correlation between this test score and the test score for the same individual on the national math test. In the latter case, the individual

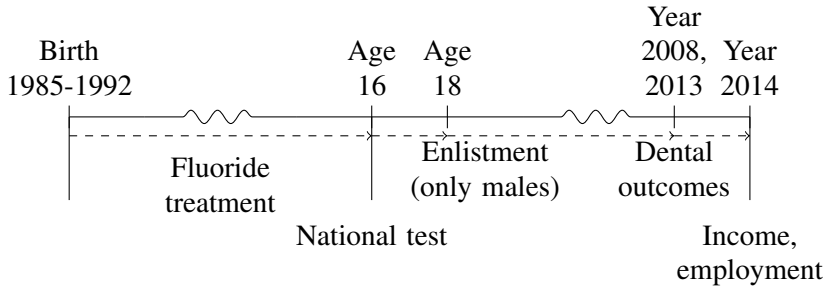


Figure 4. Timeline of measurement.

has clear incentives to perform well since final grade in math from junior high school depends on this test result. The correlation between these two tests is 0.43. We conclude that strategic manipulation on the military enlistment test does not seem to be a big concern.

Income is measured in 2014 (the last year available), and the data comes from the Swedish tax agency through Statistics Sweden. The variable is defined as gross income for all individuals that have earned any income throughout a year. We exclude all individuals that have earned less than 1,000 Swedish kronor (about $120 in 2016) during a year for this outcome. Employment status is measured in November the year 2014. An individual is coded as employed if he or she has worked at least one hour during a week.

Our main outcome variables are cognitive and non-cognitive ability, points on the national math test and labor market outcomes. In order to investigate other manifestations of how fluoride affects human capital development, we also look at health outcomes related to the brain. Data on health comes from the prescribed drug register, the inpatient and the outpatient registers. We look at prescription medicines for of ADHD, psychoses and depression which is available for 2005-2009. We also look whether the individual has a diagnosis from either the inpatient register or the outpatient register (both available for 1987-2010) for diagnoses classified within the ICD10-chapter for psychiatric illnesses (chapter F) or neurological diseases (chapter G). There has been a discussion in the earlier medical literature whether fluoride is associated with osteoporosis and hip fracture, see Näsman, Ekstrand, et al. (2013). To connect to this earlier medical literature, we also estimate the effect on skeleton and muscular diseases (chapter M). For all these health outcomes, we create dummy variables for whether an individual received a diagnosis or were prescribed medicines for any of the years available in these health registers.

Figure 4 illustrates the timing of the outcome variables and the fluoride treatment.

7 Empirical strategy

This section contains a presentation of our identification strategy and a discussion about potential threats to identification. The section also includes a presentation of the econometric set-up and descriptive statistics.

We estimate the causal effect of fluoride exposure through the drinking water on dental health cognitive ability, non-cognitive ability, education, employment status and income. We also estimate the effect of fluoride on a set of other health outcomes. The ideal experiment with maximal internal validity would be to randomly assign fluoride to individuals. Due to randomization, the fluoride levels would be independent of individual characteristics, which enable a causal interpretation of the results. Since it is not possible to randomly assign fluoride intake from birth, we need to rely on a quasiexperimental design.

We use exogenous variation in fluoride within municipalities in Sweden to estimate the effect. This enables us to control for unobservable characteristics on the municipal level which could also be determinants for the outcomes we study. Hence, our main identifying variation in fluoride stems from an exogenous geographical variation in the bedrock within municipalities.

In addition to using within-municipality variation in fluoride, we also exploit a second source of variation stemming from individuals' moving patterns. To move or not is undoubtedly endogenous, but as long as the choice of moving and the moving location is not dependent on fluoride or other variables correlated with fluoride, this yield an exogenous variation in the intensity of fluoride treatment which depends on the number of years in different SAMS. It is very unlikely that people self-select into SAMS based on the fluoride level. It is difficult to obtain information about the fluoride level since there is no comprehensive open dataset in Sweden. People cannot be aware of fluoride in the drinking water because fluoride is tasteless. We confirm that the choice to move is not dependent on the fluoride level in various tests in Table A3 presented in section D in the appendix. We also use data from Google Trends in Table A10 and conclude that people overall do not search more for information about fluoride in those regions where the fluoride level is higher.

7.1 Threats to identification

The first threat concerns our use of geological variation in fluoride. Because the bedrock is constant, the fluoride level in the drinking water is also constant over the years. If we would consider large geographical areas and use the variation between these areas, fluoride might not be independent of the outcome variables. As an illustrating example, assume that fluoride is negative for cognitive ability. If people are living in the same place over the generations, fluoride might have an effect on the regional labor market or the educational system because people on average have a lower cognitive ability. An individual's income would then be a function of individual background

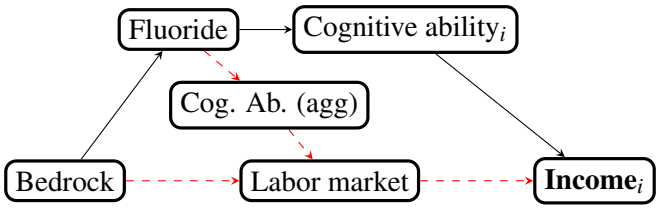


Figure 5. Relationships between the bedrock, fluoride level, cognitive ability and income.

characteristics but also the general labor market situation in the area. Since the labor market has adjusted to a lower cognitive ability pool, the individual wage level will on average be lower. It may also be the case that the bedrock in itself can affect the labor market. For example, specific bedrock might be more suitable for mining, which could affect the structure of the regional labor market and, hence, the labor market outcome for a specific individual. Figure 5 illustrates the main identification problem in this setting using the long-run outcome income as an example.

If our identification strategy relied on between-municipality variation, this would have been a concern. The key to identifying the causal effect of fluoride exposure is to have small geographical units between which there is a variation. We argue that Sweden's SAMS are sufficiently small and that fluoride is independent of the outcome between these small areas. Given the use of SAMS level data, the red dashed lines in Figure 5 are blocked.

A second threat to identification would be that municipalities deliberately provide certain SAMS with fluoridated water because municipalities have some inside information about the dangers of fluoride. We demonstrate in Table A4, A5, A6 and A7 in the appendix that this is not the case. There is no evidence that the provided drinking water fluoride level is dependent on predetermined characteristics in any clear way.

A third threat to our empirical strategy would be that people do not drink tap water but instead bottled water, meaning that our fluoride data is not accurate for the actual level of fluoride exposure. In general, Swedes drink the tap water and there are no general recommendations not to drink tap water. This is also confirmed by sales data for bottled water. Table A9 in the appendix display the total sales of bottled water per inhabitants in Sweden from 1994 to 2015. The average sales between these years are 20.3 liter per inhabitants and year. The recommended consumption of water for an individual is between 2-4 liters per day in a country with temperate climate (Fagrell 2009). This equals a yearly consumption between 730 and 1460 liters per person. The share of bottled water sales is thus only 1.4-2.8 percent of total yearly consumption of water. It is also likely that individuals during childhood drink less bottled water in

comparison to the entire population. We thus conclude that bottled water is not a threat to our empirical strategy.[19]

A fourth threat concerns self-selection for the outcome variables. There are missing values for the cognitive and non-cognitive test taken during conscription. There are also some missing values for individuals that wrote the math test on the national test in ninth grade. Imagine that fluoride is negative for cognitive ability and that some individuals as a result of being exposed to lower levels of fluoride have a possibility to avoid conscription or the math test because they are more intelligent. We would then have self-selection into who is taking the conscription test and the math test. In Table A8 in the appendix, we demonstrate that this is not the case. Whether or not we have a result from the cognitive or non-cognitive ability test or the math test does not depend on the individual fluoride treatment level.

The fifth threat is about biological inheritance of cognitive ability. Assume that fluoride is negative for cognitive ability and that cognitive ability affected by fluoride can be passed on to the offspring. The effect of fluoride on the cognitive ability of the offspring is then an inherited factor, resulting in an overestimation of the effect of fluoride exposure in the present generation. This line of thought requires that environmental cognitive factors can be transmitted. The field of epigenetics concerns environmental factors that can switch genes on and off, and then be transgenerationally transmitted. Fluoride can be stored within the body which may *potentially* switch genes on or off that are related to cognitive ability. We test if such a transmission effect is present by also running all of our specifications for adoptees only. Adoptees have not inherited genes from their adoptive parents, so the effect of fluoride in this case purely stems from variation in fluoride exposure in the present generation. We discuss this in more detail in the robustness analysis.

The sixth threat to identification is related to nurture. Assume that parents exposed to high levels of fluoride develop lower cognitive ability resulting in bad parenting skills, which in turn affects our measure of cognitive ability in the present generation. Luckily, we have a rich set of generational covariates where we can control for fathers' cognitive and non-cognitive ability measured in the same way during their enlistment. We also have covariates for parents' income and education. We can thus control for nurture effects.

7.2 Econometric set-up

The fluoride level for each individual is a weighted average for the number of years a person lived within a specific SAMS. For non-movers, their fluoride

[19] Avoidance behavior due to information in line with the discussion in Neidell (2009) and Zivin et al. (2011) is unlikely since fluoride is not considered to be a hazard for levels below 1.5 mg/l. The sales data for bottled water confirms that people – on the aggregate level – does not seem to substitute tap water to bottled water in Sweden.

level is simply the fluoride level for their SAMS between birth and up until the year when we measure the outcome variable. People may thus have lived in the same SAMS, moved between SAMS within a municipality, or moved between municipalities. We include municipality fixed effects for where the person was born since there are several differences between municipalities that may also be determinants for our outcomes. To control for age effects we include cohort fixed effects. In addition, we add municipality fixed effects for place of residence in 2014 when we measure income and employment status, since the wage structure and the possibility of employment differs throughout Sweden. We also run two subsample specifications. Those who move could experience multiple treatments; for example, a person moving to a different municipality changes school. In the first sub-sample specification, we analyze the effect of fluoride for the non-movers only, i.e., individuals who have lived in the same SAMS. In the second specification, we analyze only those who move within a municipality but between different SAMS at least once.

We estimate the following regression equation:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \beta_3 W_s + \beta_4 W_p + \tau_m + \gamma_m + \lambda_c + u_i \tag{1}$$

where Y_i is the outcome variable measured at the individual level (except for dental outcomes where it is measured for each SAMS and cohort). X_i is the amount of individual fluoride exposure, taking into account moving, for each individual. W_i is a vector of covariates on the individual level. We also include aggregated covariates on SAMS level, W_s to control for peer effects. W_p designates parental covariates. τ_m designates birth municipal fixed effects, γ_m equals municipal fixed effects in 2013 and λ_c designates cohort fixed effects. u_i is the error term. β_1 is the treatment effect of interest. The reader should note that we run several specifications where we add covariates and fixed effects sequentially. For cognitive ability, non-cognitive ability and math points, we never include municipal fixed effects in 2014 since these outcomes are measured at an earlier age.

Most SAMS do not have a water plant within the borders, meaning that the fluoride level that we assign to a SAMS is not independent on the fluoride level of the other SAMS within the same municipality. Therefore, we choose to cluster the standard errors on the birth municipal level because municipalities are responsible for the drinking water. This clustering level is our benchmark and we use it throughout the paper. In the regression tables in the result section, we also add standard errors clustered at other levels. The main variation in fluoride is on the SAMS level so we also cluster the standard errors on the birth SAMS level. In addition, we calculate standard errors clustered at the local labor market region in accordance with the definitions from Statistics Sweden.[20] In a fourth standard error specification, we calculate spa-

[20]There are 73 local labor market regions in Sweden which are statistical areas for commuting regions. These standard errors are based upon place of residence in 2014 and we only estimate them when we look at personal income and employment status in 2014.

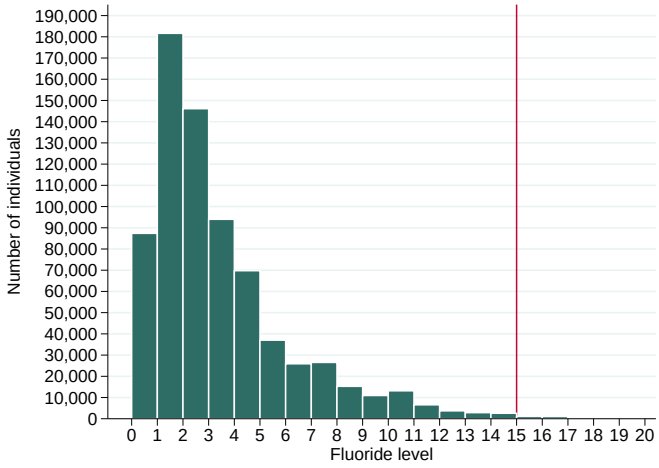


Figure 6. Histogram of fluoride levels below 2 mg/l (in 0.1 mg/l).

tial adjusted standard errors in line with Conley (1999) and use 10 kilometers as a spatial cut-off. These standard errors are based upon Euclidian distance, and the clustering structure is specified to last up until 10 kilometers from the center point of each SAMS. It can be argued that geographical distance is a more natural clustering level since individuals living far from each other are less dependent than those who live close, in comparison to municipalities and labor market regions which are administrative constructed entities.

7.3 Descriptive statistics

In this subsection we present descriptive statistics. Figure 6 presents a histogram of the frequency of individuals who are treated with the corresponding level of fluoride, expressed in 0.1 mg/l. The level displayed in the histogram is the actual individual treatment level taken into account moving patterns between different SAMS and municipalities. The histogram displays treatment up until age 16 which is when our first outcome variable is measured. The WHO recommendation of maximum 1.5 mg/l in the drinking water is marked with a red line.[21]

Our identification is based on an exogenous variation in fluoride stemming from a variation in the bedrock. In Table 2, we present some detailed descriptive statistics of the standard deviation in fluoride levels within and between municipalities. It is clear from the table that there is variation within municipalities, but also between municipalities. The combined variation is used to

[21] Those few cases above 1.5 mg/l originates from the earlier exceptions for higher levels mentioned in the data section. We cut the histogram at 2 mg/l because there are so few observations above 2 mg/l.

Table 2. *Standard deviation decomposition of fluoride*

	Mean	SD
Fluoride (0.1 mg/l)	3.53	
Overall		3.25
Between		2.95
Within		1.89
Observations	8,597	

Notes: Between and within variation on municipal level.

Table 3. *Descriptive statistics of main outcome variables*

	Mean	SD
Annual income in SEK	183,804	143,198
Employment status	0.73	0.44
Cognitive ability	5.01	1.93
Non-cognitive ability	4.75	1.82
Number of basic points math test	26.18	8.57

estimate the effect of fluoride where we consider people's moving patterns within and between municipalities as an additional source of variation.

Table 3 presents the mean and standard deviations for our five main outcomes of interest. The equivalent Table A2 for dental outcomes and the other health outcomes (Table A1) can be found in the appendix. Cognitive and non-cognitive ability are only measured for men and are centered on 5 with a standard deviation of about 2, which follows the Stanine definition. 73 percent of the individuals in our sample are employed, which is close to the population share of employed. The maximum number of points on the math test is 45, and the mean is about 26 points.

Table 4 presents descriptive statistics of the covariates. The sample is balanced on gender (49 percent women). More than 90 percent have at least high school education in 2014. Only 5 percent is married, which is not surprising given that the individuals in the sample are relatively young. We also include covariates for parents' level of education and income (mean real wage between 1985 and the last year available) for the parents, and whether they are immigrants. Income for the parents are specified as log income in the regressions, but displayed as real income in Table 4.[22] We are also able to include cognitive and non-cognitive ability from the enlistment for the father

[22] Böhlmark and Lindquist (2005) find that current income is not as good measure of lifetime income as the widespread use would imply. See also the discussion in Engström and Hagen (2015). To minimize bias we use all available years of income for the parents.

Table 4. *Descriptive statistics of covariates*

	Mean	SD	Outcomes	Set
Gender	0.49	0.50	All	Small
Individual at least high school	0.92	0.27	Income, employment	Small
Marital status	0.07	0.26	All	Large
Father at least high school	0.82	0.39	All	Large
Father's income	242,878	151,121	All	Large
Father's cognitive ability	5.07	1.90	All but non-cog. ability	Large
Father's non-cognitive ability	5.15	1.75	All but cog. ability	Large
Father immigrant	0.09	0.29	All	Large
Mother at least high school	0.89	0.31	All	Large
Mother's income	158,827	86,940	All	Large
Mother immigrant	0.10	0.30	All	Large
Both parents immigrants	0.04	0.21	All	Large
Cohort education (birth)	12.03	0.58	All	Large
Cohort education (school start)	12.03	0.25	All	Large
Cohort education (16 years age)	12.03	0.25	All	Large
Observations	728,074			

Notes: Explanatory variables used in the estimations. Small set covariates are also included in the large set covariates. Cohort education variables (last three in the table) are means for cohorts and SAMS.

as covariates. However, the enlistment data starts 1969 so older fathers are not included. To capture peer-effects, we measure the mean education among individuals included in the data for each cohort and SAMS for three time points. We measure the individuals' education as grown-ups in 2014 and then aggregate for each cohort and SAMS for where the individuals were born, where they started school (at 7 years of age) and where they lived at age 16. We include a dummy for whether an individual has graduated from high school when we estimate the effect on income and employment, but not when measuring cognitive ability, non-cognitive and the number of math points since these are measured before graduation.[23] We have grouped our covariates into two groups: Small set and Large set. Table 4 therefore also indicates which covariate is included in each group.

8 Results

In this section we present the results. We start by looking at the effects on dental health, and then present the results for our main outcomes. Next, we present the results for our additional health outcomes, followed by a section of results for the non-linear specifications. The section is ended with a comparison with earlier studies.

[23] Whether to graduate or not from high school could be a bad control. However, whether an individual graduates from high school is influenced by several other factors than cognitive ability and at the same time, graduation from high school is important for later labor market status. Therefore, we choose to include it when studying income and employment status.

Table 5. *Dental outcomes*

	Visit	Repair	RiskEvaluation	DiseasePrevention	DiseaseTreatment	RootCanal
2013	-0.6554	-0.3369	-0.6882	-0.8453	-0.3506	-0.0292
	(0.2987)**	(0.1103)***	(0.3015)**	(0.4309)*	(0.1389)**	(0.0172)*
	<0.0879>***	<0.0555>***	<0.0906>***	<0.0835>***	<0.0757>***	<0.0156>*
2008	-0.6356	-0.2290	-0.6765	-0.4337	0.1093	-0.0300
	(0.2935)**	(0.0683)***	(0.3204)**	(0.2238)*	(0.1056)	(0.0197)
	<0.0949>***	<0.0589>***	<0.0974>***	<0.0764>***	<0.0646>*	<0.0168>*

Notes: Standard errors in parenthesis clustered at the municipal level. Standard errors in <> are clustered on the SAMS level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The number of observations for the year 2013 is 7,622. The number of observations for the year 2008 is 7,606. Fluoride expressed in 0.1 mg/l. The dependent variable is displayed at the top of each column.

8.1 Effects of fluoride on dental health

If our strategy of mapping statistics from water plants to individual register data on the SAMS level has worked, we expect to see a positive effect of fluoride on dental health. We have dental outcomes for each cohort for each SAMS. The average number of individuals in a SAMS per included cohorts in our dental data set is approximately 16.

We have a set of variables that measure various dental outcomes. We present the results for a subset of these variables below that we judged was closely related to fluoride. The results for all additional outcomes are presented in Table A11 section E in the appendix. The variables we focus on here are visits to a dental clinic, tooth repairs, disease evaluation, prevention and treatment and root canal. Given that fluoride is good for dental health, we expect to find negative estimates for these variables. All these variables are expressed as share in percentage points; for example the share of 20 years old in a given SAMS that had a tooth repaired during a year. For a more detailed description about the variable abbreviations we use for the outcome variables in this section, see Table A2 in the appendix.

We divide our regression results into two separate tables. In Table 5 we run unweighted regressions where we look at the connection between fluoride and the aggregated measure of these six variables on the SAMS level. For this analysis, we focus on the 20 year olds which is the earliest cohort available. We can be more sure that the 20 year olds have not moved from a given SAMS in comparison to later cohorts. In Table 6 we run weighted regressions where we work with our full dataset. For this analysis, individuals from cohorts in the data analysis for the main outcomes are included. In this case, each individual has a unique fluoride treatment depending on moving patterns and the aggregated fluoride level on the SAMS level thus corresponds to those living in a SAMS.[24]

Table 5 clearly displays a negative effect of fluoride level for these outcomes. The reader may find the results both for the 2008 sample and the 2013 sample in Table 5. The point estimates are large and often statistically signifi-

[24] SAMS is not yet available for 2013 LOUISE data set. We have used SAMS for the individual in 2011 in this case.

Table 6. *Dental outcomes*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Visit	-0.2903	-0.0655	-0.0118	-0.0164	0.0067	-0.0052	-0.0011
	(0.1605)*	(0.0458)	(0.0433)	(0.0428)	(0.0343)	(0.0357)	(0.0360)
	<0.0386>***	<0.0178>***	<0.0195>	<0.0194>	<0.0187>	<0.0206>	<0.0206>
Repair	-0.0776	-0.0682	-0.0598	-0.0575	-0.0697	-0.0595	-0.0640
	(0.0600)	(0.0256)***	(0.0317)*	(0.0316)*	(0.0277)**	(0.0294)**	(0.0279)**
	<0.0134>***	<0.0105>***	<0.0138>***	<0.0138>***	<0.0140>***	<0.0152>***	<0.0152>***
RiskEvaluation	-0.3032	-0.0671	-0.0126	-0.0174	0.0062	-0.0042	0.0002
	(0.1685)*	(0.0478)	(0.0444)	(0.0438)	(0.0345)	(0.0360)	(0.0364)
	<0.0400>***	<0.0184>***	<0.0198>	<0.0198>	<0.0190>	<0.0208>	<0.0208>
DiseasePrevention	-0.5169	-0.1318	-0.1154	-0.1186	-0.0748	-0.0613	-0.0607
	(0.2741)*	(0.0619)**	(0.0553)**	(0.0547)**	(0.0348)**	(0.0383)	(0.0384)
	<0.0462>***	<0.0161>***	<0.0174>***	<0.0174>***	<0.0161>***	<0.0185>***	<0.0185>***
DiseaseTreatment	-0.0656	-0.0217	-0.0072	-0.0060	-0.0168	-0.0247	-0.0250
	(0.0996)	(0.0388)	(0.0340)	(0.0340)	(0.0282)	(0.0294)	(0.0296)
	<0.0280>**	<0.0152>	<0.0180>	<0.0180>	<0.0176>	<0.0195>	<0.0195>
RootCanal	-0.0051	-0.0138	-0.0159	-0.0145	-0.0182	-0.0137	-0.0156
	(0.0126)	(0.0058)**	(0.0077)**	(0.0076)*	(0.0070)***	(0.0072)*	(0.0071)**
	<0.0042>	<0.0041>***	<0.0051>***	<0.0051>***	<0.0052>***	<0.0059>**	<0.0059>***
Small set covariates	No	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	No	Yes
Fe. birth muni.	No	No	Yes	Yes	Yes	Yes	Yes
Fe. cohort	No	No	No	Yes	Yes	Yes	Yes
Fe. muni. 2014	No	Yes	No	No	Yes	Yes	Yes
Sample	All	All	All	All	All	Col 7	All

Notes: Standard errors in parenthesis clustered at the municipal level. Standard errors in <> are clustered on the SAMS level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Outcomes on each row. The number of observations ranges between 472,287 (col 6 and 7) and 725,004.

cant. If we take the first estimate in Table 5 as an example, the share of visits is decreased by approximately 6.6 percentage points if fluoride is increased by 1 mg/l. This should be considered as a large effect. The outcome that should be closest related to fluoride is tooth repair, which is displayed in column 2. If fluoride would increase with 1 mg/l, the share of 20 year olds that had a tooth repaired would be decreased approximately 3.4 percentage points considering the 2013 sample. Again, this effect is large, especially for this cohort. 20 year olds should on average have healthy teeth, but we still find these effects of fluoride. Root canal treatment is generally a treatment for more serious conditions caused by caries. We find a negative point estimate for this outcome (which is expected), but the coefficients are only statistically significant on the 10 percent level. This is again expected given that root canal treatment should be generally rare among those who are 20 years old. DiseaseTreatment is positive for 2008, but negative and large for the 2013 sample. It is important to note that comparisons across the years should not be done with this data, since definitions of treatments and diagnostics have somewhat altered across the years.

The results presented in Table 6 point in the same direction as the ones in Table 5, but the point estimates are generally smaller in size. The reason for this is probably because we consider the average treatment of fluoride between birth and up until we measure dental outcomes. Fluoride needs to be continuously applied to teeth and fluoride exposure in later years should be more important than the fluoride level that the individual was exposed to several years ago. People tend to move away from their parents after age 20, meaning

that the average fluoride level is more representative when measured at age 20 (Table 5) since people probably move more often when they are 21-28 in comparison to when they are 0-20. We focus on the 2013 data sample in Table 6. In the appendix, the reader may find results for additional outcomes and the equivalent results for the 2008 sample in Tables A12, A13 and A14.

The share of repairs is the most well-defined variable where we really expect to find an effect, and the results for this variable are stable across different specifications and points in the expected direction. If we consider column 7 where all covariates and fixed effects are included, the share of individuals that had a tooth repaired would decrease by approximately 0.6 percentage points if fluoride increased by 1 mg/l. This effect is smaller than the one found in Table 5, but still large considering that fluoride needs to be applied continuously to the teeth. What our results indicate – which is interesting in itself – is that fluoride treatment throughout the entire life has long run positive effects on dental health. Root canal treatment is now often statistically significant, which is expected since we have included older cohorts. Although the point estimates are not always statistically significant for the dental health outcomes, they almost always points in the expected negative direction.[25]

The overall conclusion after considering the results in Table 5-6 and the additional results presented in the appendix is that out mapping strategy seems to have worked. Generally, we find negative and often statistically significant results for fluoride on these outcomes; especially if we consider the 2013 sample.[26]

8.2 Main results

In this subsection we present our main results. We begin by looking at cognitive ability, non-cognitive ability and points at the math test taken in ninth grade. Then we move on and investigate the effect of fluoride on more long-term outcomes where we look at income and employment status. In this subsection we present the linear specifications. There are, however, reason to believe that the effect may be non-linear, and that fluoride become dangerous above a certain level. We estimate the non-linear effects in the next subsection.

[25] We can conclude that the coefficients for the 2008 specification are generally smaller in size and less precisely estimated. A reform was implemented in July 2008 that gave 20-29 years old a special dental care benefits. Given that people in their 20's usually have lower incomes, the benefit probably allowed people between 20 and 29 to visit the dentist regularly, which could potentially explain that the results are less clear for 2008.

[26] For two of the variables, we find results that point in the opposite directed that we expected for some of the specifications. These variables are median of intact teeth and median of remaining teeth. See the results in the appendix. After further consideration, we conclude that these outcomes are not suitable for this age group. Wisdom teeth are developed in this age, meaning that the median of remaining and intact teeth are mostly influenced wisdom teeth incidence. See section E for a discussion and for additional analysis on these two outcomes.

Table 7. *Cognitive ability*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride up until age 18 (0.1 mg/l)	-0.0088	-0.0028	-0.0028	-0.0021	0.0045	0.0030	0.0205
	(0.0082)	(0.0051)	(0.0051)	(0.0052)	(0.0038)	(0.0053)	(0.0078)***
	<0.0030>***	<0.0038>	<0.0038>	<0.0045>	<0.0040>	<0.0056>	<0.0084>**
	{0.0086}	{0.0046}	{0.0045}	{0.0052}	{0.0041}	{0.0054}	{0.0088}**
Mean	5.0067	5.0067	5.0067	5.0222	5.0222	5.0897	4.9246
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	Col 5	All	SAMS stayers	SAMS movers
R^2	0.0002	0.0216	0.0239	0.0282	0.1718	0.1683	0.1802
Observations	81,776	81,776	81,776	51,203	51,203	20,513	19,178

Notes: Standard errors in parenthesis are clustered at the municipal of birth. Standard errors in <> are clustered on the SAMS of birth. Standard errors in curley brackets are Conley standard errors with a cut-off of 10 km, centered on each SAMS. *** p < 0.01, ** p < 0.05, * p < 0.1.

Let us begin with cognitive ability, measured in a Stanine scale. In this case we only include males in our specifications and consider a fluoride treatment between birth and age 18. In Table 7 we present the point estimates for fluoride and three types of standard errors. The first standard error in parenthesis is clustered on the birth municipality. The standard errors within <> are clustered on the birth SAMS level. The standard errors in curly brackets are spatial adjusted standard errors in line with Conley (1999). The first column does not include any covariates or fixed effects. In the following two columns we add fixed effects. When we include covariates for fathers' cognitive ability our sample is reduced since we only have data on fathers' cognitive ability from 1969. To make the samples comparable with and without the covariates we run column 4 with the same sample as if we had included covariates which we do in column 5. We run two subsample analyses where we only focus on those individuals that have not moved from a municipality between birth and age 18. In column 6, we run an analysis for those who have lived in the same SAMS in a municipality for the entire period 0-18. In column 7 we restrict our sample to those who have moved, but only within a municipality.

Looking at the point estimates, they are all very small and often not statistically significant different from 0. Sometimes the point estimates are negative and sometimes they are positive, but always very close to 0. Fluoride is expressed in 0.1 mg/l. If we take the point estimate from column 5, which is equal to 0.0045, this means that cognitive ability is increased by 0.045 Stanine points if fluoride is increased by 1 mg/l (a large increase in fluoride). This should be considered as a zero-effect on cognitive ability. A Stanine point roughly equals 6-8 IQ points.[27]

Let us move on to non-cognitive ability. The point estimates are once again very close to 0 and often not statistically significant. If we do the same calculation as before with an increase in fluoride by 1 mg/l, the non-cognitive score would increase by 0.154 Stanine points according to column number 5. In this column, the point estimate is actually statistically significant, but the

[27]IQ measure with population mean of 100 and a standard deviation of 15. See Öhman (2015).

Table 8. *Non-cognitive ability*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride up until age 18 (0.1 mg/l)	0.0026	0.0058	0.0059	0.0109	0.0154	0.0087	0.0353
	(0.0058)	(0.0046)	(0.0046)	(0.0050)**	(0.0050)***	(0.0067)	(0.0148)**
	<0.0026>	<0.0037>	<0.0037>	<0.0046>**	<0.0045>***	<0.0069>	<0.0094>***
	{0.0054}	{0.0043}	{0.0043}	{0.0051}**	{0.0048}***	{0.0066}	{0.0126}***
Mean	4.7340	4.7340	4.7340	4.7754	4.7754	4.9214	4.6953
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	Col 5	All	SAMS stayers	SAMS movers
R^2	0.0000	0.0175	0.0176	0.0214	0.0784	0.0791	0.0934
Observations	66,375	66,375	66,375	41,636	41,636	16,731	15,425

Notes: Standard errors in parenthesis are clustered at the municipal of birth. Standard errors in <> are clustered on the SAMS of birth. Standard errors in curley brackets are Conley standard errors with a cut-off of 10 km, centered on each SAMS. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 9. *Math points*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until age 16 (0.1 mg/l)	-0.1031	-0.0296	-0.0269	-0.0269	-0.0435	-0.0163	-0.0184	-0.0191
	(0.0354)***	(0.0126)**	(0.0125)**	(0.0125)**	(0.0144)***	(0.0119)	(0.0133)	(0.0204)
	<0.0099>***	<0.0093>***	<0.0092>***	<0.0092>***	<0.0102>***	<0.0085>*	<0.0118>	<0.0165>
	{0.0355}***	{0.0116}**	{0.0115}**	{0.0115}**	{0.0128}***	{0.0096}*	{0.0120}	{0.0164}
Mean	26.2059	26.2059	26.2059	26.2059	26.4900	26.4900	27.2221	26.0441
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
R^2	0.0013	0.0229	0.0403	0.0403	0.0431	0.1643	0.1472	0.1723
Observations	499,892	499,892	499,892	499,892	336,827	336,827	139,149	127,062

Notes: Standard errors in parenthesis are clustered at the municipal of birth. Standard errors in <> are clustered at the SAMS of birth. Standard errors in curley brackets are Conley standard errors with a cut-off of 10 km, centered on each SAMS. *** p < 0.01, ** p < 0.05, * p < 0.1.

result should be interpreted as a negligible effect because of the very small estimated coefficient. In economic terms, the effect is zero.

For the next outcome variable – the number of points at the math test taken in the ninth grade – we have data for both males and females. In this case we also have data for additional cohorts in comparison to the first two outcomes. Fluoride treatment now takes place between birth and age 16. The average score was approximately 26. All of the point estimates are negative in this case and some of the estimated coefficients are statistically different from zero. The size of the point estimates are, however, very small. In the first four columns we have almost 500,000 observations so it is not surprising that some of our results are statistically significant. The important part is economic significance. Let us focus on column 6 where we have included all covariates and all fixed effects. If fluoride is increased by 1 mg/l (again, this is a large increase), the number of points on the math test should decrease by less than 0.2 points. This decrease is less than 1 percent of the average number of points on the test which was 26 points. In economic terms, this effect should be considered as a zero-effect.

We may thus conclude that we cannot reject the null hypothesis that fluoride does *not* have a negative effect on cognitive development.

Table 10 and 11 studies outcomes which are more long-term: Log annual income and employment status in 2014. These are the outcome variables for which we have the largest number of observations. Given the zero-results for the three variables above, we do not expect to find a negative effect on these

Table 10. *Annual log income in SEK*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	0.0053	0.0035	0.0040	0.0052	0.0040	0.0042	0.0030	0.0019
	(0.0031)*	(0.0014)**	(0.0014)***	(0.0016)***	(0.0014)***	(0.0014)***	(0.0021)	(0.0040)
	[0.0023]**	[0.0026]	[0.0028]	[0.0016]***	[0.0017]**	[0.0019]**	[0.0021]	[0.0038]
	<0.0007>***	<0.0008>***	<0.0008>***	<0.0008>***	<0.0010>***	<0.0010>***	<0.0010>***	<0.0010>***
	{0.0031}*	{0.0010}***	{0.0011}***	{0.0012}***	{0.0012}***	{0.0012}***	{0.0019}	{0.0025}
Mean	11.9124	11.9124	11.9124	11.9124	11.9229	11.9229	11.8452	11.9544
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
R^2	0.0002	0.0065	0.0528	0.0967	0.0997	0.1066	0.1289	0.1197
Observations	634,793	634,793	634,793	634,793	419,162	419,162	72,089	150,458

Notes: Individuals with a yearly income below 1,000 SEK are excluded. Standard errors in parenthesis are clustered at the municipal of birth. Standard errors in brackets are clustered at the local labor market area defined by Statistics Sweden (SCB). Standard errors in <> are clustered at the SAMS of birth. Standard errors in curley brackets are Conley standard errors with a cut-off of 10 km, centered on each SAMS. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 11. *Employment status*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	0.0021	0.0016	0.0018	0.0023	0.0019	0.0020	0.0016	0.0018
	(0.0013)*	(0.0006)**	(0.0006)***	(0.0007)***	(0.0006)***	(0.0006)***	(0.0010)	(0.0016)
	[0.0008]***	[0.0011]	[0.0012]	[0.0005]***	[0.0006]***	[0.0007]***	[0.0010]	[0.0014]
	<0.0003>***	<0.0003>***	<0.0004>***	<0.0004>***	<0.0004>***	<0.0004>***	<0.0007>**	<0.0008>**
	{0.0013}*	{0.0004}***	{0.0005}***	{0.0005}***	{0.0005}***	{0.0005}***	{0.0008}**	{0.0010}*
Mean	0.7346	0.7346	0.7346	0.7346	0.7459	0.7459	0.7129	0.7582
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
R^2	0.0002	0.0069	0.0322	0.0662	0.0661	0.0752	0.0778	0.0789
Observations	728,074	728,074	728,074	728,074	474,556	474,556	81,867	170,142

Notes: Standard errors in parenthesis are clustered at the municipal of birth. Standard errors in <> are clustered at the SAMS of birth. Standard errors in brackets are clustered at the local labor market area defined by Statistics Sweden (SCB). Standard errors in curley brackets are Conley standard errors with a cut-off of 10 km, centered on each SAMS. *** p < 0.01, ** p < 0.05, * p < 0.1.

long-term outcomes. It is, however, possible that fluoride has a positive effect, because of better dental health for the individuals. In the two tables we add an additional standard error calculation where the standard errors in brackets are clustered at the local labor market area in 2014. We also add an additional set of municipal fixed effects for where the individual lives in 2014. Fluoride is measured between birth and the year 2014.

Looking at log income, we have often statistically significant point estimates and the coefficients are always positive. If we look at column 6, the point estimate equals 0.0042, meaning that income increases by 4.2 percent if fluoride increases by 1 mg/l. This is not a negligible effect and the estimate should be considered as economically significant.

Let us continue to the last outcome. Employment status is a dummy variable taking the value 1 if the individual is defined as employed in 2014. In column 6, the point estimate for fluoride is 0.002 and statistically significant. If fluoride is increased by 1 mg/l, then the probability that the person is employed is increased by 2 percentage points. This result thus point in the same direction as the results for log income where both these results are significant in economic terms.

In the last two tables we looked at income and employment status for all included cohorts born 1985-1992. One objection is that the included cohorts are only 22-29 years old when income and employment status are measured, meaning that the estimates are not representative for the lifetime income and probability of being employed. In the subsample analysis below, we restrict

Table 12. *Annual log income in SEK (subsample)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride up until year 2014 (0.1 mg/l)	-0.0006	0.0057	0.0062	0.0048	0.0043	0.0042	0.0044
	(0.0012)	(0.0017)***	(0.0018)***	(0.0035)	(0.0025)*	(0.0042)	(0.0033)
	[0.0012]	[0.0025]**	[0.0019]***	[0.0061]	[0.0027]	[0.0033]	[0.0031]
	<0.0008>	<0.0018>***	<0.0019>***	<0.0034>	<0.0024>*	<0.0039>	<0.0031>
	{0.0012}	{0.0017}***	{0.0018}***	{0.0035}	{0.0024}*	{0.0039}	{0.0031}
Mean	12.1639	12.1520	12.3967	11.7976	12.2209	12.3500	12.1347
Birth cohort FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	Yes	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	Yes	Yes	Yes	Yes	Yes	Yes
Sample	All	No. Coll., all	No Coll., men	No Coll., women	Coll., all	Coll., men	Coll., women
R^2	0.0000	0.1195	0.0417	0.0394	0.0562	0.0761	0.0509
Observations	216,779	80,849	47,825	33,024	53,757	21,527	32,230

Notes: Individuals with a yearly income below 1,000 SEK are excluded, and individuals born 1988 or later. Standard errors in parenthesis are clusterem

In conclusion, we find zero-effects on cognitive and non-cognitive ability. We also find zero-effects for the number of math points. These results indicate that fluoride does not have adverse negative effect on cognitive development for the fluoride levels we consider. We also find that fluoride has positive effects on log income and employment status which could indicate that better dental health is a positive factor on the labor market. We investigate the reduced form results for income and employment status further below.

Interpreting the reduced form effect for labor market outcomes

The initial hypothesis that we wanted to test was whether fluoride has negative effects on human capital development. Log income and employment status was considered as alternative outcomes also measuring human capital development later in life. We could however not reject the null hypothesis that the effect was zero for cognitive and non-cognitive ability or math points on the national test. What we do in this subsection is that we run an IV analysis for dental health on labor market outcomes using fluoride as an instrument for dental health. This is however not an IV in the strict sense where we argue that the effect of the instrument only goes through the instrumented variable. We have already presented a potential second pathway that goes through human capital development where the hypothesis was that fluoride may be a neurotoxin. We merely use the IV as a method to interpret the size of the reduced form where we estimate the effect of dental health on labor market outcomes. Dental health status is only available to us on the aggregate level for each SAMS and cohort. We therefore collapse out data on later labor market status and fluoride to the same level to make the estimates interpretable. Given that the data is collapsed, we cannot include individual covariates or any fixed effects anymore. We choose to focus on dental repairs in the IV analyses since dental repairs have such clear connection to fluoride.

In Table 14 the IV for log income is presented. The reader may both find the OLS, the first stage, the reduced form and the 2SLS for this collapsed data set. The F-values for the first stage is presented at the bottom of the table. Two different analyses are presented. In the first part of the table, we run the analysis for all available cohorts. In the second part, we restrict the analysis to those who are 27-29 years old. The average share of repairs is about 18 percent (with a median of 17 percent).

Considering the full sample in Table 14, we find that when dental repairs increases by 1 percentage point, income decreases by 2 percent on the same aggregate level. This effect is clearly economically significant. This indicates that fluoride improves labor market outcomes through better dental health. The reduced form estimate in Table 14 equals 0.0034, meaning that when fluoride increases by 1 mg/l, income increases by 3.4 percent. This estimate may be compared to Glied and Neidell (2010), who find that women who drinks fluoridated water on average earn 4 percent more. The effect on income may also be compared to estimated education premiums. Card (1999) conducts

Table 14. *Annual log income in SEK*

	OLS Log income	FS Repair	RF Log income	2SLS Log income
Repair	0.0005			-0.0208
	(0.0002)***			(0.0282)
	<0.0002>***			<0.0071>***
Fluoride		-0.1625	0.0034	
		(0.0830)*	(0.0033)	
		<0.0325>***	<0.0009>***	
F stat. Municipality		3.83		
F stat. SAMS		25.07		
Sample	All			
Repair	0.0000			0.2420
	(0.0002)			(2.4793)
	<0.0003>			<1.1406>
Fluoride		-.0122	-0.0030	
		(0.1225)	(0.0019)	
		<0.0572>	<0.0015>*	
F stat. Municipality		0.01		
F stat. SAMS		0.05		
Sample	1985-1987			

Notes: Individuals with a yearly income below 1,000 SEK are excluded. Standard errors in parenthesis are clustered at the municipal level. Standard errors in <> are clustered at the SAMS level. *** $p < 0.01$, ** $p < 0.05$, * < 0.1.

Table 15. *Employment status*

	OLS Employment	FS Repair	RF Employment	2SLS Employment
Repair	0.0005			-0.0151
	(0.0001)***			(0.0175)
	<0.0001>***			<0.0040>***
Fluoride		-0.1673	0.0025	
		(0.0844)**	(0.0019)	
		<0.0326>***	<0.0004>***	
F stat. Municipality		3.93		
F stat. SAMS		26.33		
Sample	All			
Repair	0.0004			-0.0610
	(0.0001)***			(0.3942)
	<0.0001>***			<0.1661>
Fluoride		-0.0218	0.0013	
		(0.1247)	(0.0013)	
		<0.0577>	<0.0007>*	
F stat. Municipality		0.03		
F stat. SAMS		0.14		
Sample	1985-1987			

Notes: Standard errors in parenthesis are clustered at the municipal level. Standard errors in <> are clustered at the SAMS level. *** $p < 0.01$, ** $p < 0.05$, * < 0.1.

a meta-study reviewing several papers that have used different techniques to estimate the causal effect of education. The return of one additional year of education seems to be associated with an increase in income by approximately 6-10 percent, considering the IV estimates in the review study. If the share of dental repairs increases by 1 percentage point, the income is reduced be 2 percent according to our results. This corresponds to a quarter of a year longer education. For employment status, we find estimates going in a similar direction. If dental repairs increase by one percentage point, the probability of being employed on the same aggregated level is decreased by 1.5 percentage point considering the full sample. When we restrict the analysis to only those who are 27-29 years old, the F-values for the first stage is extremely small, making the IV uninterpretable. We have the same problem when we cluster the standard errors on the muncipal level.[28]

The question is what the causal channel looks like. The estimated effect could be interpreted as a beauty-effect. Given that we found larger effects for non-academics in the earlier reduced form analyses, one explanation might be that people working in the service sector – which is not uncommon for this age-group – are more sensitive to bad looking teeth. This is probably not the entire explanation however. Having bad dental health is probably associated with pain, and individuals with dental problems should on average be more sick and more absent from work. This could explain why they earn less and are less likely to be employed.

8.3 Additional outcomes: Health status

The purpose of this paper is primarily to study human capital development where we have focused on cognitive and non-cognitive abilities, education and labor market status. Given that we did not find any negative effects of fluoride on these outcomes, it is not likely that a negative effect of fluoride would manifest itself on more serious health outcomes. It is however interesting to see if this really is the case. In Table 16 and 17 we run the analysis on the prescription of medicines for ADHD, depression and psychoses. We also run the analysis for diagnoses from the outpatient and the inpatient registers. We look at psychiatric diagnoses and neurological diagnoses. We also estimate the effect on diagnoses for muscular and skeleton diseases to connect to the discussion whether fluoride has an effect on osteoporosis. All outcome variables are defined as dummy variables for whether the individual was prescribed or diagnosed sometimes during the measurement period. The ATC and ICD codes that we use can be found in appendix Q.

[28] One explanation for why we no longer find the same effect in the reduced form or in the first stage is probably because our data is now collapsed where each cohort and SAMS have an equal weight in the regressions. For some SAMS and cohorts, many individuals are included, and in others, far fewer individuals are included.

Table 16. *Prescription of medicine*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ADHD medicine	0.0000	-0.0001	-0.0001	-0.0002	-0.0001	-0.0002	-0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)*	(0.0001)*	(0.0001)**	(0.0001)*	(0.0002)
	<0.0001>	<0.0001>*	<0.0001>	<0.0001>**	<0.0001>**	<0.0001>***	<0.0001>*	<0.0002>
Antidepressants	0.0003	0.0000	-0.0001	-0.0002	-0.0002	-0.0003	-0.0005	-0.0002
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)**	(0.0005)
	<0.0001>**	<0.0002>	<0.0002>	<0.0002>	<0.0002>	<0.0002>**	<0.0002>**	<0.0004>
Antipsychotics	0.0000	-0.0000	-0.0001	-0.0001	-0.0001	-0.0001	-0.0000	0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
	<0.0000>	<0.0001>	<0.0001>	<0.0001>	<0.0001>	<0.0001>	<0.0001>	<0.0001>
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Fe. birth muni.	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fe. cohort	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Fe. muni. 2013	No	No	No	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	Col 7	All	SAMS stayers	SAMS movers

Notes: Standard errors in parenthesis clustered at the municipal of birth. Standard errors in <> clustered on the SAMS of birth. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Outcomes on each row. The number of observations ranges between 292,307 and 724,945.

Table 17. *Diagnosis*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mental retardation in childhood	0.0006	-0.0001	-0.0001	-0.0003	-0.0003	-0.0005	-0.0004	-0.0002
	(0.0006)	(0.0004)	(0.0004)	(0.0004)	(0.0003)	(0.0004)	(0.0003)	(0.0008)
	<0.0002>***	<0.0002>	<0.0002>	<0.0002>	<0.0002>	<0.0002>**	<0.0002>	<0.0005>
Neurological diseases	0.0001	0.0001	0.0001	-0.0000	-0.0000	-0.0000	0.0001	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
	<0.0001>	<0.0001>	<0.0001>	<0.0001>	<0.0001>	<0.0001>	<0.0002>	<0.0003>
Musculoskeletal diseases	-0.0006	-0.0005	-0.0005	-0.0006	-0.0006	-0.0006	-0.0003	-0.0005
	(0.0004)	(0.0002)**	(0.0002)**	(0.0003)**	(0.0002)**	(0.0002)**	(0.0003)	(0.0006)
	<0.0002>***	<0.0002>**	<0.0002>**	<0.0002>***	<0.0002>***	<0.0002>**	<0.0003>	<0.0005>
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Fe. birth muni.	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fe. cohort	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Fe. muni. 2013	No	No	No	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	Col 7	All	SAMS stayers	SAMS movers

Notes: Standard errors in parenthesis clustered at the municipal of birth. Standard errors in <> clustered on the SAMS of birth. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Outcomes on each row. The number of observations ranges between 292,307 and 724,945

It is clear from the first table that there is a zero-effect of fluoride on the probability of being prescribed any of these medicines. The point estimates are not always statistically significant and always small in size. Taking the estimate in the sixth column as an example, the probability of receiving ADHD medicines is decreased by 0.2 percentage points if fluoride is increased by 1 mg/l. In economic terms, this effect is a zero-effect.

The same picture emerges with diagnosis. The estimated effects are small and often statistically insignificant.

In conclusion, we do not find that fluoride has any effects on these health outcomes. This further strengthens our argument that fluoride does not have any negative effects for levels below 1.5 mg/l on human capital development or health outcomes related to human capital development. It is also interesting that we do not find any effects on diagnoses for muscular and skeleton diseases, which has been a question also discussed in connection to fluoride.

8.4 Non-linear effects

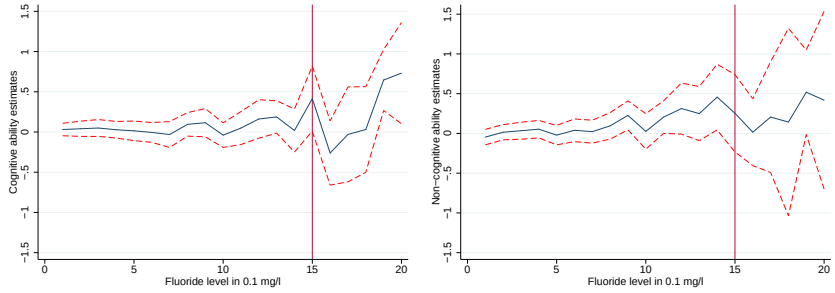
There are reasons to believe that a potential neurotoxic effect of fluoride on the central nervous system is not linear. As with many toxic compounds, small amounts do not yield any dramatic damage, but the effects manifest

itself above a certain threshold. We therefore continue our analysis and look for non-linear effects.

In Figures 7-9 the effect for each fluoride level is displayed. We have created dummy variables taking the value 1 for each 0.1 fluoride level and then included these in a regression. When we run the regressions, all fixed effects and all covariates are included just as in column 6 in the earlier tables. We then plot the effect for each 0.1 mg/l in a figure. Fluoride in our data is between 0 and 4 mg/l, but we have very few observations above the threshold level of 1.5 mg/l, meaning that the estimated effect is very noisy for high levels. In the figures, we have therefore cut the individual fluoride treatment level at 2 mg/l. The blue lines in the figures are the plotted point estimates and the red dashed lines are 95 % confidence intervals. The conclusion is that the effect up until 1.5 mg/l is always close to zero. In line with the earlier results for log income and employment status, the line in the figures seem to increase when closing on 1.5 mg/l, which indicate a positive effect of fluoride through dental health for higher levels. Also in line with the main analysis, the point estimates for the number of math points are sometimes statistically significant. The size of the point estimates are small, and the effect does not seem to be significant when considering fluoride levels close to 1.5 mg/l, which we would expect if fluoride had a negative effect on cognitive development.

The corresponding figures for dental health and other health outcomes may be found in the appendix (Figure A3 and A4). For the other health outcomes, the results are stable around zero. If we look at dental repairs and disease prevention, we can see an improvement of the dental health for fluoride levels up till 1 mg/l (fewer repairs, less preventions). However, for the other results, there are no evidence of an increasing effect higher fluoride levels. In section H in the appendix, we also present regression tables where we run the regressions with dummy variables for each quartile value in the fluoride distribution. In the tables, we run the exact same specifications for each outcome variable as in the tables in the last section when we looked at linear effects. The conclusion is, again, that there are no indications that fluoride has an effect other than zero for cognitive ability, non-cognitive ability and math points. For math points, we have some statistically significant, negative point estimates for the third quartile dummy. For the fourth quartile however, the point estimates are insignificant and positive for all specifications which we expect if fluoride does not have a negative effect on these outcomes. With regard to log income and employment status, we find positive and statistically significant results for the fourth quartile, which again points towards the explanation that fluoride has a positive effect through dental health – especially for higher levels of fluoride.[29]

[29] We have also created corresponding non-linear effects tables for dental outcomes. These tables are available from the authors upon request.



(a) Cognitive ability estimates (b) Non-cognitive ability estimates

Figure 7. Non-linear effects for ability measures.

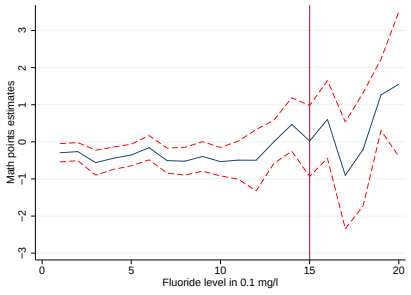
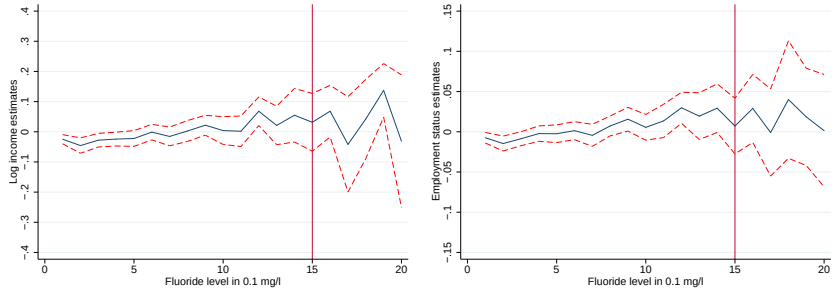


Figure 8. Non-linear math points estimates.



(a) Log income estimates (b) Employment estimates

Figure 9. Non-linear effects labor market outcomes.

Table 18. *Comparison with earlier studies*

Study	Obs.	*F.*	CI 95 %
Our study: No cov. or f.e.	81,776	0.05-4.10	-0.1296, 0.0386
Our study: Cov. and f.e.	51,203	0.05-4.10	-0.0156, 0.0626
Chen et al. (1991)	640	0.89-4.55	-0.41, -0.10
Lin et al. (1991)	119	0.34-0.88	-1.01, -0.28
Xu et al. (1994)	129	0.80-1.80	-1.35, -0.52
Yang et al. (1994)	60	0.50-2.97	-1.01, 0.02
Li et al. (1995)	907	1.02-2.69	-0.70, -0.39
Zhao et al. (1996)	320	0.91-4.12	-0.76, -0.31
Yao et al. (1997)	502	0.40-2.00	-0.61, -0.25
Lu et al. (2000)	118	0.37-3.15	-0.98, -0.25
Hong et al. (2001)	117	0.75-2.90	-0.85, -0.03
Wang et al. (2001)	60	0.50-2.97	-1.01, 0.02
Xiang et al. (2003)	512	0.18-4.50	-0.82, -0.46
Seraj et al. (2006)	126	0.40-2.50	-1.28, -0.50
Li et al. (2009)	80	0.96-2.34	-0.94, 0.08
Poureslami et al. (2011)	119	0.41-2.38	-0.77, -0.04

Notes: F is fluoride level in mg/l. This table consists of the results of comparable studies presented in Table 1 and Figure 2 on page 1364-1366 in Choi, Sun, et al. (2012). Note that these studies have not considered a continuous measure of fluoride.

8.5 Comparison with earlier studies

Are our estimated results for cognitive ability really zero? One way to evaluate a zero-result is to look at earlier studies which have found statistically significant results and compare the precision of the estimates. In Table 18, we have summarized the results for the reviewed papers in Choi, Sun, et al. (2012). We have only included the papers which study fluoride levels that are roughly equal to the levels we consider. Because earlier papers only have considered cognitive ability, we can only compare this outcome variable. To make our results comparable to the other papers, we have normalized cognitive ability around 0. The reader should note that we have not read the original articles since most of them are printed in Chinese or Persian. Instead, the comparison below is based on Choi, Sun, et al. (2012).[30]

In comparison to earlier papers, our study is based on a much larger sample, and our point estimates are much more precise. Earlier papers have found negative and statistically significant effects in many cases, but our results are always very close to 0. Our 95 % confidence intervals include the zero both with and without fixed effects and covariates.

[30] Since we have not read the original research articles, we do not cite them in the reference list. See Choi, Sun, et al. (2012) for details about these papers.

Broadbent et al. (2015) also claim to find a zero-result. Their confidence intervals are, however, much broader than ours. They estimate a 95 % confidence interval for the effect of living in a high fluoride (0.7-1 mg/l) area in comparison to those living in a low fluoride area (0-0.3 mg/l) on cognitive ability (with covariates) to be (-3.49, 3.20) for those between 7 and 13 years old and between (0.02, 5.98) for those at age 38. In this case, cognitive ability is measured in IQ points with a mean of 100. If we translate our estimates to IQ points, roughly by replacing the Stanine scores with the corresponding IQ[31], our confidence intervals are (-1.8560, 0.5546) for the specifications without covariates or fixed effects and (-0.2267, 0.8919) for the specifications with all covariates and fixed effects, when fluoride is increased by 1 mg/l.

Based on the assessment of the earlier literature, we are confident to claim that we have estimated a zero-effect on cognitive ability.

9 Robustness analysis

In this section we discuss the results from various robustness checks.

First we address the potential threat to our identification strategy that fluoride as an environmental factor can switch certain genes on and off in accordance with the idea in epigenetics. To test if this is a problem, we rerun all our specifications only including individuals that were adopted in section I in the appendix. The estimates are more noisy in this case since we are left with fewer observations. We find mixed results on income and employment, but no statistically significant negative results. There are no indications of any negative effect human capital development.

We use a mapping protocol to assign water plant data on fluoride in the drinking water to SAMS. Since we cannot observe the exact coordinate where an individual lives, we will have some measurement error with regard to those who drink water from a private well. All we know is if an individual live in a specific SAMS for a given year.[32] The probability that an individual consume the drinking water provided by the municipality should increase when the SAMS is small and/or when the distance from the water plant to the center of the SAMS is small. Smaller SAMS equals more densely populated areas. We have run all of our specifications in section J and K in the appendix where we look at subsamples in our data for various sizes of SAMS and various distances between the nearest water plant and the center point of the SAMS. We have plotted these estimates in graphs presented in the appendix. In conclusion, the point estimates does not seem to differ in a systematic way when just considering smaller SAMS and shorter distances, which is reassuring.

[31] See Table 1 in Öhman (2015).

[32] In a theoretical scenario where we have severe measurement error, we would have bias in our estimates towards 0. This is not likely given our results for dental health, however.

We do not have water statistics for each year from 1985 for all municipalities. We have therefore contacted all municipalities and asked them if they have changed their water sources after 1985. Because the bedrock is constant, they level of fluoride should also be constant from 1985 if the water source is the same. All municipalities do not have exact information regarding their water sources, and we have not received confirmation from all of them. In section L in the appendix, we also run a specification including only those municipalities where we have data from 1985 or where we have received a clear confirmation (conservative judgement) that the municipality has not changed their water sources after 1985. The results for cognitive and non-cognitive ability are in economic terms still zero. The estimated coefficients for math points are negative and sometimes statistically significant (as in the main analysis), but very small in size. For log income and employment status, we estimate positive coefficients as in the main analysis, but the estimated point estimates are generally smaller in magnitude in this specification.

We also run specific analysis only for those only born in 1985 in section M for labor market outcomes. The results point in the same direction as in the main analysis for employment, but is more mixed for income. The specifications with all covariates and fixed effects point in the same direction as in the main analysis.

We also run a specification where we only look at those SAMS which had one and only one water plant and where we have full information from 1985 from the municipalities in section N. In this specification we only include those who have not moved. In this case we are left with much fewer observations. For cognitive ability, non-cognitive ability and math points, there is still no evidence of any negative effects. For log income and employment status, the point estimates varies between different specifications and we no longer have statistically significant results. This is again probably a result of having fewer observations and thus lower statistical power.

We have also run an analysis for an alternative income measure in section O in the appendix. In the main analysis we look at a measure for income from employment. In the alternative specification, we run the same analysis for a measure for income from employment and business income (förvärvsinkomst). These results point in the same direction as the ones in the main analysis.

Finally, we have run specifications where we have included mother fixed effects. The variation in fluoride now stems from different moving patterns of a family where siblings have been exposed to different fluoride levels throughout life because they have resided in different areas for different amount of time. The reader should note that this specification is very demanding and forces the comparisons in the regressions to be very selective. If we take cognitive ability for instance, the variation in fluoride now stems from brothers born between 1985-1987 where the family has moved between their respective births and age 18. The empirical results points in different directions

depending on the outcome variable. For math points, we find no evidence of any negative effects. For cognitive ability and non-cognitive ability, the estimates are not statistically significant, but the point estimates are negative and large. For income and employment status, we have some negative, very large and statistically significant effects, but the point estimates moves towards zero when other fixed effects and covariates are included and becomes statistically insignificant.

Overall, while the results are mixed in our robustness checks, we are confident to conclude that we find support for our main analysis. The reader should bare in mind that when testing many different specifications for different subsamples, one can expect to find some that show different results.

10 Conclusions

We have investigated the effects of fluoride on outcomes related to the central nervous system and more long-term labor market outcomes. We find a zero-effect of fluoride on cognitive ability, non-cognitive ability and points on the national test in math. We also find a zero-effect of the probability of being prescribed medicines for ADHD, depression or psychiatric conditions as well as the probability of being diagnosed for psychiatric illnesses, neurological illnesses or muscular or musculoskeletal diseases. For income and employment status we found evidence of a positive effect of fluoride, which would be in line with the explanation that better dental health is a positive factor on the labor market. We began our analysis by first investigating the dental health effects of fluoride, and could confirm the long well-established positive relationship.

Our paper is to our knowledge the first large scale empirical study with individual register data to assess the effects of fluoride in the drinking water. Earlier studies, which have found a negative effect of fluoride on cognitive ability, rely on much smaller samples originating from countries with poorer data quality. In addition, these papers have usually not applied credible identification strategies. That said, earlier studies have sometimes focused on higher levels of fluoride than the levels we consider in this paper. It may be that higher levels of fluoride in the drinking water have negative effects on cognitive ability. However, in comparison, our paper is more policy relevant for developed countries, because water authorities seldom consider fluoridating the drinking water above 1.5 mg/l. Based on the results we find, the policy implications are that fluoride exposure through the drinking water either in the form of natural levels or artificial fluoridation is a good mean of improving dental health without risking negative side effects on cognitive development. Given our results, it is possible to do a cost-benefit analysis whether artificial fluoridation is cost-effective, without worrying about negative side effects.

Future studies should try to establish where the dangerous level of fluoride begins. Since we know that fluoride is lethal and dangerous in very high

dosages, it is crucial to find the safe limit for fluoride in the drinking water. Our results indicate that the dangerous level is not below 1.5 mg/l.

References

Aizer, Anna, Janet Currie, Peter Simon, and Patrick Viver. 2016. *Do low levels of blood lead reduce children's future test scores?* Working Paper 22558. National Bureau of Economic Research (NBER).

Augustsson, Anna, and Tobias Berger. 2014. "Assessing the risk of an excess fluoride intake among Swedish children in households with private wells – expanding static single-source methods to a probabilistic multi-exposure-pathway approach." *Environment International* 68:192–199.

Böhlmark, Anders, and Matthew J Lindquist. 2005. *Life-cycle variations in the association between current and lifetime income: Country, cohort and gender comparisons.* Working Paper 4. Swedish Institute for Social Research (SOFI).

Broadbent, Jonathan M, W Murray Thomson, Sandhya Ramrakha, Terrie E Moffitt, Jiaxu Zeng, Lyndie A Foster Page, and Richie Poulton. 2015. "Community water fluoridation and intelligence: Prospective study in New Zealand." *American Journal of Public Health* 105 (1): 72–76.

Card, David. 1999. "The causal effect of education on earnings." In *Handbook of Labor Economics,* edited by Orley Ashenfelter and David Card, 3:1801–1863. Amsterdam, North Holland: Elsevier.

Case, Anne, Angela Fertig, and Christina Paxson. 2005. "The lasting impact of childhood health and circumstance." *Journal of Health Economics* 24 (2): 365–389.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. "Economic status and health in childhood: The origins of the gradient." *The American Economic Review* 92 (5): 1308–1334.

Chioca, Lea R, Inara M Raupp, Cláudio Da Cunha, Estela M Losso, and Roberto Andreatini. 2008. "Subchronic fluoride intake induces impairment in habituation and active avoidance tasks in rats." *European Journal of Pharmacology* 579 (1–3): 196–201.

Choi, Anna L, Guifan Sun, Ying Zhang, and Philippe Grandjean. 2012. "Developmental fluoride neurotoxicity: A systematic review and meta-analysis." *Environmental Health Perspectives* 120 (10): 1362–1368.

Choi, Anna L, Ying Zhang, Guifan Sun, David C Bellinger, Kanglin Wang, Xiao Jing Yang, Jin Shu Li, Quanmei Zheng, Yuanli Fu, and Philippe Grandjean. 2015. "Association of lifetime exposure to fluoride and cognitive functions in Chinese children: A pilot study." *Neurotoxicology and Teratology* 47:96–101.

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92 (1): 1–45.

Cunha, Flavio, and James J Heckman. 2007. "The technology of skill formation." *The American Economic Review: Papers & Proceedings* 97 (2): 31–47.

———. 2009. "The economics and psychology of inequality and human development." *Journal of the European Economic Association* 7 (2-3): 320–364.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78 (3): 883–931.

Currie, Janet. 2009. "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development." *Journal of Economic Literature* 47 (1): 87–122.

———. 2011. "Inequality at birth: Some causes and consequences." *The American Economic Review: Papers & Proceedings* 101 (3): 1–22.

Currie, Janet, Joshua Graff Zivin, Katherine Meckel, Matthew Neidell, and Wolfram Schlenker. 2013. "Something in the water: Contaminated drinking water and infant health." *The Canadian Journal of Economics/Revue Canadienne d'Economique* 46 (3): 791–810.

Currie, Janet, and Mark Stabile. 2003. "Socioeconomic status and child health: Why is the relationship stronger for older children?" *The American Economic Review* 93 (5): 1813–1823.

Das, Kousik, and Naba Kumar Mondal. 2016. "Dental fluorosis and urinary fluoride concentration as a reflection of fluoride exposure and its impact on IQ level and BMI of children of Laxmisagar, Simlapal Block of Bankura District, WB, India." *Environmental Monitoring and Assessment* 188 (4): 1–14.

Dey, Sananda, and Biplab Giri. 2016. "Fluoride fact on human health and health problems: A review." *Medical & Clinical Reviews* 2 (1:11): 1–6.

Ding, Yunpeng, Gao Yanhui, Huixin Sun, Hepeng Han, Wei Wang, Xiaohong Ji, Xuehui Liu, and Dianjun Sun. 2011. "The relationships between low levels of urine fluoride on children's intelligence, dental fluorosis in endemic fluorosis areas in Hulunbuir, Inner Mongolia, China." *Journal of Hazardous Materials* 186 (2): 1942–1946.

Edmunds, W Mike, and Pauline L Smedley. 2013. "Fluoride in natural waters." In *Essentials of Medical Geology,* edited by Olle Selinus, 311–336. New York NY and London: Springer.

Engström, Per, and Johannes Hagen. 2015. *Income underreporting among the self-employed: A permanent income approach.* Working Paper 2. Center for Fiscal Studies, Department of Economics, Uppsala University.

Fagrell, Bengt. 2009. "Drick när du är törstig – för mycket vatten kan vara livsfarligt." *Läkartidningen* 106 (7): 448–449.

Fawell, John, Kenneth Bailey, J Chilton, E Dahi, L Fewtrell, and Y Magara. 2006. *Fluoride in Drinking-Water.* IWA Publishing, London: World Health Organization.

Feigenbaum, James J, and Christopher Muller. 2016. "Lead exposure and violent crime in the early twentieth century." Unpublished manuscript. Accessed October 24, 2016. `http://scholar.harvard.edu/jfeigenbaum/publications`.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky. 2005. "Water for life: The impact of the privatization of water services on child mortality." *Journal of Political Economy* 113 (1): 83–120.

Gessner, Bradford D, Michael Beller, John P Middaugh, and Gary M Whitford. 1994. "Acute fluoride poisoning from a public water system." *New England Journal of Medicine* 330 (2): 95–99.

Glied, Sherry, and Matthew Neidell. 2010. "The economic value of teeth." *The Journal of Human Resources* 45 (2): 468–496.

Grandjean, Philippe, and Philip J Landrigan. 2014. "Neurobehavioural effects of developmental toxicity." *The Lancet Neurology* 13 (3): 330–338.

Griffin, S O, E Regnier, P M Griffin, and V Huntley. 2007. "Effectiveness of fluoride in preventing caries in adults." *Journal of Dental Research* 86 (5): 410–415.

Grönqvist, Hans, Peter J Nilsson, and Per-Olof Robling. 2014. *Childhood lead exposure and criminal behavior: Lessons from the Swedish phase-out of leaded gasoline.* Working Paper 9. Swedish Institute for Social Research (SOFI).

Grossman, Michael. 1972. "On the concept of health capital and the demand for health." *Journal of Political Economy* 80 (2): 223–255.

Harrison, Paul T C. 2005. "Fluoride in water: A UK perspective." *Journal of Fluorine Chemistry* 126 (11): 1448–1456.

Heck, Brandon. 2016. "The effect of community water fluoridation on health and education." Unpublished manuscript. Accessed October 24, 2016. http://brandonheck.com/#research.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics* 24 (3): 411–482.

Hu, Yu-Huan, and Si-Shung Wu. 1988. "Fluoride in cerebrospinal fluid of patients with fluorosis." *Journal of Neurology, Neurosurgery and Psychiatry* 51 (12): 1591–1593.

Jans, Jenny, Per Johansson, and Peter Nilsson. 2014. *Economic status, air quality, and child health: Evidence from inversion episodes.* Discussion Paper 7929. IZA.

John, Jeyanthi H. 2002. "No adequate evidence to assess differences between natural and artificial water fluoridation." *Evidence-Based Dentistry* 3 (2): 50–50.

Johnston, Philip. 2014. "Fluoride: Just when you thought it was safe to drink the water..." Accessed January 19, 2016. http://www.telegraph.co.uk/lifestyle/wellbeing/healthadvice/10722701/Fluoride-Just-when-you-thought-it-was-safe-to-drink-the-water....html.

Khan, Suleman Abbas, Saumya Navit Rahul Kumar Singh, Dheera Chadha, Nikita Johri, Pragati Navit, Anshul Sharma, and Rachana Bahuguna. 2015. "Relationship between dental fluorosis and intelligence quotient of school going children in and around Lucknow District: A cross-sectional study." *Journal of Clinical and Diagnostic Research* 9 (11): ZC10–ZC15.

Kundu, Hansa, P Basavaraj, Ashish Singla, Ritu Gupta, Khushboo Singh, and Swati Jain. 2015. "Effect of fluoride in drinking water on children's intelligence in high and low fluoride areas of Delhi." *Journal of Indian Association of Public Health Dentistry* 13 (2): 116.

Lamberg, Matti, Hannu Hausen, and Terttu Vartiainen. 1997. "Symptoms experienced during periods of actual and supposed water fluoridation." *Community Dentistry and Oral Epidemiology* 25 (4): 291–295.

Li, Mang, Yanhui Gao, Jing Cui, Yuanyuan Li, Bingyun Li, Yang Liu, Jing Sun, et al. 2016. "Cognitive impairment and risk factors in elderly people living in fluorosis areas in China." *Biological Trace Element Research* 172 (1): 1–8.

Lindqvist, Erik, and Roine Vestman. 2011. "The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment." *American Economic Journal: Applied Economics* 3 (1): 101–128.

Liteplo, R, R Gomes, P Howe, and H Malcolm. 2002. *Fluorides.* Environmental Health Criteria Series 227. Geneva: World Health Organization.

Livsmedelsverket. 2015. "Egen brunn." Accessed June 16, 2016. `http://www.livsmedelsverket.se/matvanor-halsa--miljo/egen-brunn/`.

Malin, Ashley J, and Christine Till. 2015. "Exposure to fluoridated water and attention deficit hyperactivity disorder prevalence among children and adolescents in the United States: An ecological association." *Environmental Health* 14 (17): 1–10.

Mercola, Joseph. 2013. "Harvard study confirms fluoride reduces children's IQ." Accessed January 19, 2016. `http://www.huffingtonpost.com/dr-mercola/fluoride_b_2479833.html`.

Mullen, Joe. 2005. "History of water fluoridation." *British Dental Journal* 199:1–4.

Mullenix, Phyllis J, Pamela K Denbesten, Ann Schunior, and William J Kernan. 1995. "Neurotoxicity of sodium fluoride in rats." *Neurotoxicology and Teratology* 17 (2): 169–177.

Nagarajappa, Ramesh, Piyush Pujara, Archana J Sharda, Kailash Asawa, Mridula Tak, Pankaj Aapaliya, and Nikhil Bhanushali. 2013. "Comparative assessment of intelligence quotient among children living in high and low fluoride areas of Kutch, India – a pilot study." *Iranian Journal of Public Health* 42 (8): 813–818.

Näsman, Peggy, J Ekstrand, F Granath, A Ekbom, and C M Fored. 2013. "Estimated drinking water fluoride exposure and risk of hip fracture: A cohort study." *Journal of Dental Research* 92 (11): 1029–1034.

Näsman, Peggy, Fredrik Granath, Jan Ekstrand, Anders Ekbom, Gunilla Sandborgh-Englund, and C Michael Fored. 2016. "Natural fluoride in drinking water and myocardial infarction: A cohort study in Sweden." *Science of the Total Environment* 562:305–311.

Neidell, Matthew. 2009. "Information, avoidance behavior, and health: The effect of ozone on asthma hospitalizations." *Journal of Human Resources* 44 (2): 450–478.

Nilsson, Peter J. 2009. "The long-term effects of early childhood lead exposure: Evidence from the phase-out of leaded gasoline." Unpublished manuscript. Accessed June 16, 2016. https://sites.google.com/site/nilssonjanpeter/.

Öhman, Mattias. 2015. *Be smart, live long: the relationship between cognitive and non-cognitive abilities and mortality.* Working Paper 2015:21. Institute for Evaluation of Labour Market and Education Policy (IFAU).

Peckham, Stephen, and Niyi Awofeso. 2014. "Water fluoridation: A critical review of the physiological effects of ingested fluoride as a public health intervention." *The Scientific World Journal* 2014:1–10.

Persson, Kenneth M, and Stefan Billqvist. 2004. *Pilotförsök med nanofiltrering av fluoridhaltigt dricksvatten.* VA-Forsk rapport 2004-13. Svenskt Vatten.

Ramesh, Maya, Rita Mary Aruna, Narasimhan Malathi, and Ramesh Krishnan. 2014. "A review of fluoride and its diverse effects." *SRM Journal of Research in Dental Sciences* 5 (1): 42–45.

Richards, Derek, ed. 2002. "Water fluoridation – controversy or not?" *Evidence-Based Dentistry* 3 (2). http://www.nature.com/ebd/journal/v3/n2/.

Saxena, Sudhanshu, Anjali Sahay, and Pankaj Goel. 2012. "Effect of fluoride exposure on the intelligence of school children in Madhya Pradesh, India." *Journal of Neurosciences In Rural Practice* 3 (2): 144–149.

Schlenker, Wolfram, and W Reed Walker. 2015. "Airports, air pollution, and contemporaneous health." *The Review of Economic Studies* 83 (2): 768–809.

Sebastian, Shibu Thomas, and S Sunitha. 2015. "A cross-sectional study to assess the intelligence quotient (IQ) of school going children aged 10-12 years in villages of Mysore district, India with different fluoride levels." *Journal of Indian Society of Pedodontics and Preventive Dentistry* 33 (4): 307–311.

Seraj, B, M Shahrabi, M Shadfar, R Ahmadi, M Fallahzadeh, H Farrokh Eslamlu, and M J Kharazifard. 2012. "Effect of high water fluoride concentration on the intellectual development of children in Makoo/Iran." *Journal of Dentistry* 9 (3): 221–229.

SGU. 2013. *Bedömningsgrunder för grundvatten.* Rapport 2013:01. Sveriges Geologiska Undersökning (SGU).

Smith, James P. 2009. "The impact of childhood health on adult labor market outcomes." *The Review of Economics and Statistics* 91 (3): 478–489.

Svenskt Vatten. 2016. "Fluorid." Accessed June 8, 2016. http://wwwold.svensktvatten.se/Vattentjanster/Dricksvatten/Takt-till-kran/Kemiska-amnen/Fluorid/.

Tang, Qin-qing, Jun Du, Heng-hui Ma, Shao-jun Jiang, and Xiao-jun Zhou. 2008. "Fluoride and children's intelligence: A meta-analysis." *Biological Trace Element Research* 126 (1): 115–120.

Twetman, Svante, Susanna Axelsson, Helena Dahlgren, Anna-Karin Holm, Carina Källestål, Folke Lagerlöf, Peter Lingström, et al. 2003. "Caries-preventive effect of fluoride toothpaste: A systematic review." *Acta Odontologica Scandinavia* 61 (6): 347–355.

WHO. 2001. "Water sanitation and health: Dental fluorosis. World water day: Oral health." Accessed June 16, 2016. http://www.who.int/water_sanitation_health/oralhealth/en/index3.html.

———. 2004. *Fluoride in drinking-water, background document for development of WHO guidelines for drinking-water quality (03.04.96).* World Health Organization.

———. 2011. *Guidelines for Drinking-Water Quality.* 4th ed. Geneva: World Health Organization.

———. 2015. "Naturally occurring hazard. Water sanitation and health: Fluoride." Accessed June 16, 2016. http://www.who.int/water_sanitation_health/naturalhazards/en/index2.html.

Zhang, Jing. 2012. "The impact of water quality on health: Evidence from the drinking water infrastructure program in rural China." *Journal of Health Economics* 31 (1): 122–134.

Zivin, Joshua Graff, Matthew Neidell, and Wolfram Schlenker. 2011. "Water quality violations and avoidance behavior: Evidence from bottled water consumption." *The American Economic Review* 101 (3): 448–53.

Appendix

A Exogenous variation in fluoride: Geological background

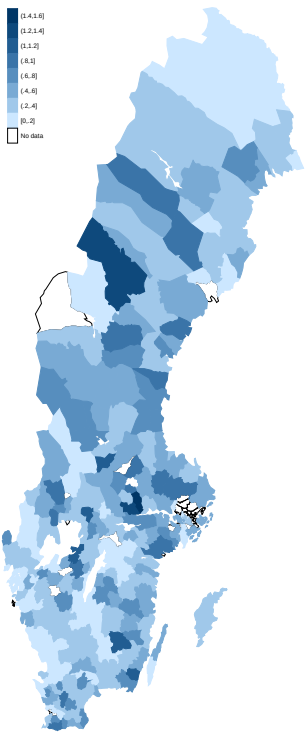
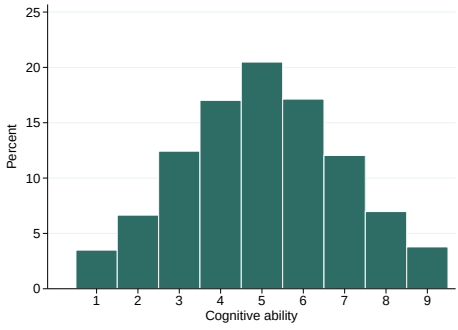
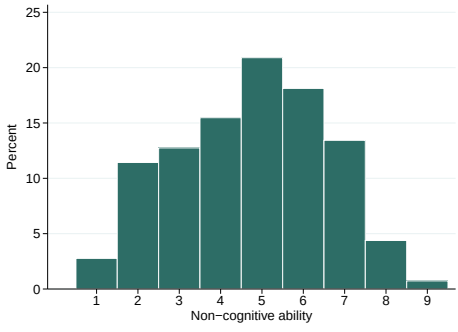


Figure A1. Fluoride levels in Sweden: Variation between municipalities after mapping.

B Data: Individual level



(a)



(b)

Figure A2. Distribution of cognitive and non-cognitive abilities.

Table A1. *Descriptive statistics of dental outcomes*

	Mean	SD
ADHD medicine	0.01	0.11
Antidepressants	0.06	0.24
Antipsychotics	0.01	0.10
Mental retardation in childhood	0.12	0.32
Neurological diseases	0.04	0.19
Musculoskeletal diseases	0.13	0.34

C Data: SAMS and cohort level

Table A2. *Descriptive statistics of dental outcomes*

	Mean	SD	Max	Min
Visits dental clinic	66.31	24.31	100.00	0.00
Basic check-ups	59.42	25.92	100.00	0.00
Risk evaluation, health improvement measures	64.78	24.64	100.00	0.00
Disease prevention	12.82	18.97	100.00	0.00
Disease treatment	31.31	23.21	100.00	0.00
Dental surgical measures	6.33	11.66	100.00	0.00
Root canal treatment	2.75	7.67	100.00	0.00
Orthognathic treatment	1.37	5.50	100.00	0.00
Dental repair	18.85	19.22	100.00	0.00
Prosthesis treatment	0.72	4.04	100.00	0.00
Orthodontics and replacement measures	0.18	2.06	100.00	0.00
Diagnosis: Check-ups and evalutions	64.77	24.64	100.00	0.00
Diagnosis: Dental health improvement measures	9.44	15.31	100.00	0.00
Diagnosis: Treatment of illness and pain	34.93	24.00	100.00	0.00
Diagnosis: Dental repair	22.86	20.67	100.00	0.00
Diagnosis: Habilitation and rehabilitation	0.76	4.05	100.00	0.00
Median remaining teeth	29.52	1.36	32.00	1.00
Median intact teeth	25.87	2.89	32.00	0.00

D Empirical framework: Balance tests

Our identifying variation stems from a geological variation in fluoride and from individuals' moving patterns. It is important that we verify that people are not moving from and to different SAMS because of the fluoride level. If people were, we would have self-selection into the intensity of treatment meaning that we cannot separate treatment from the outcomes. In the following balance test we investigate if the moving patterns are related to the fluoride level between birth and age 16 (the first year for our outcome variables).

Table A3 display balance tests for moving patterns where each row is a separate regression. Overall, the moving pattern is on average not depending on the individual fluoride treatment level. We run specific balance tests using dummy variables taking the value 1 if an individual has moved between SAMS within a municipality, if the individual has moved between municipalities, and if the individual has moved between counties. We also run balance tests for the number of moves between SAMS, municipalities and counties, and the average number of years within a SAMS, municipality or county. The point estimates are always small and statistically insignificant. If the individual fluoride treatment increases by 0.1 mg/l, the probability that the individual has moved between SAMS within a municipality is 0.49 percentage points lower according to row 1 in Table A3. We have also conducted a comparison in difference in means for first time movers. The mean fluoride level prior of moving was approximately 0.33 mg/l and after moving the mean was 0.34 mg/l. Hence, there is no evidence that people move from high fluoride areas.

Table A3. *Balance test. Moving pattern, individual fluoride treatment level*

	F. (0.1 mg/l)
Move within municipality	-0.00487 (0.00408)
Municipal Move	0.0000883 (0.00263)
County Move	0.00139 (0.00158)
# moves within municipality	-0.00371 (0.00807)
# moves between municipalities	0.00133 (0.00428)
# moves between counties	0.00240 (0.00223)
Average years SAMS	0.0184 (0.0354)
Average years municipality	-0.0329 (0.0365)
Average year county	-0.0367 (0.0229)
Observations	731,888

Notes: Standard errors clustered at the birth municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each row is a separate regression, where the dependent variable is displayed on the row. The number of observations refers to the maximum number of observations. For row 1 and 4, we restrict the sample to those who have moved within a municipality, but between SAMS. The number of observations are thus smaller for these two specification (566,631 observations).

In Table A4 we investigate whether the municipality provided water is endogenously rerouted to specific groups. We investigate this by running balance tests on predetermined characteristics on the SAMS level for where the individual was born. Municipalities may potentially know that fluoride is dangerous, and therefore give such water to groups with lower socioeconomic status. We also investigate whether other characteristics are dependent on the fluoride level, such as the size of SAMS or the distance to the water plant. These balance tests address the question whether fluoride is correlated with population

density, since less populated areas have larger SAMS. We have also run a test for those municipalities for which we do not have full information about their drinking water from 1985.

Table A5 and A6 displays a similar analysis for the years of immigration for the parents. This variable is also predetermined, where we run the balance test for various dummy variables for mothers and fathers respectively. We focus on where the individual was born and calculate the share of immigrants that arrived for each year. All shares are then included into a single regression.

We do not find support for the concerns discussed above. We have statistically significant results on the 10 percent level for the share (expressed between 0 and 1) of immigrants outside the Nordic countries (although not outside Europe), but the estimates are negatively related to the fluoride level. We have one statistically significant result for the number of water plants within a SAMS. Those SAMS without a water plant have on average lower fluoride. This is because the three largest cities in Sweden has few and large water plants and generally low fluoride levels. These areas also consist of many SAMS because of large populations. The point estimate is however very small. If the fluoride level within a SAMS increased by 0.1 mg/l, the number of water plants would increase by 0.02 water plants. In practice, this is a zero-effect. With regards to Table A5 and Table A6, there is no evidence that municipalities reroute fluoride to certain immigration cohorts. The share in this case is expressed between 0 and 100. Some results are statistically significant, but all point estimates are small in magnitude (below 0.1 mg/l), with the exception of one coefficient. Let us take the first row in Table A6 as an example. If the share of immigrant fathers that arrived to Sweden in 1945 increases by 1 percentage point of the SAMS population (a large increase), the fluoride level to that SAMS would be 0.08 mg/l lower. The reader should note when interpreting statistically significant results that the precision of fluoride measurement is 0.1 mg/l. The reader should also note that some of these immigration cohorts consist of very few people.

Table A4. *Balance test. Predetermined characteristics. Fluoride for each SAMS*

	F. (0.1 mg/l)
SAMS area	3.550
	(2.523)
Distance WP	0.0803
	(0.182)
Not full info	0.000580
	(0.0115)
Number WP, SAMS	0.0203***
	(0.00710)
Father immigrant	-0.00159
	(0.00171)
Mother immigrant	-0.00215
	(0.00169)
Both parents immigrants	-0.00119
	(0.000971)
Father immigrant outside Nordic	-0.00238*
	(0.00143)
Mother immigrant outside Nordic	-0.00237*
	(0.00129)
Both parents immigrant outside Nordic	-0.00136*
	(0.000807)
Father immigrant outside Europe	-0.00130
	(0.000892)
Mother immigrant outside Europe	-0.00120
	(0.000823)
Both parent immigrant outside Europe	-0.000762
	(0.000541)
Mother's age at birth	-0.0320
	(0.0317)
Father's age at birth	-0.0260
	(0.0245)
Gender	0.000304
	(0.000303)
Adopted	0.000101
	(0.000109)

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each row is a separate regression, where the dependent variable is displayed on the row. The number of observations ranges between 8,023 and 8,597.

Table A5. *Fathers*

	Fluoride (0.1 mg/l)
1945	-0.8420***
1946	-0.3145***
1947	-0.6139*
1948	0.2294
1949	0.0332
1950	0.5998*
1951	0.5872***
1952	0.0959
1953	-0.4260***
1954	0.0065
1955	0.3217**
1956	0.1253
1957	0.1388*
1958	-0.0244
1959	0.0870
1960	0.0484
1961	0.0525
1962	-0.0331
1963	0.0387
1964	0.0231
1965	0.1123
1966	0.0762
1967	-0.0096
1968	-0.0192
1969	0.0018
1970	0.0057
1971	-0.1015**
1972	-0.0200**
1973	-0.0412**
1974	-0.0116
1975	-0.0167
1976	-0.0326
1977	-0.0390
1978	-0.0127
1979	-0.0267
1980	-0.0143
1981	-0.0285
1982	-0.0304
1983	-0.0273
1984	-0.0451*
1985	-0.0379
1986	-0.0803**
1987	-0.0303*
1988	-0.0204
1989	0.0130
1990	-0.0747*
1991	-0.0365***
1992	0.0721

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The number of observations are 8,017. Fluoride is dependent variable.

Table A6. *Mothers*

	Fluoride (0.1 mg/l)
1944	-1.1273***
1945	-2.3393
1946	-0.1197
1947	-0.9070**
1948	-0.1104
1949	1.1819*
1950	-0.0141
1951	0.3395
1952	-0.0574
1953	0.1247
1954	0.2745*
1955	0.0103
1956	-0.0077
1957	0.0382*
1958	-0.1383
1959	-0.0401
1960	0.0325
1961	0.0068
1962	-0.0398
1963	0.0547
1964	0.0487
1965	0.0940
1966	0.0017
1967	-0.0463
1968	-0.0189
1969	0.0537
1970	-0.0108
1971	0.0334
1972	-0.0424
1973	-0.0388
1974	0.0173
1975	-0.0745***
1976	-0.0401*
1977	-0.0323**
1978	-0.0561***
1979	-0.0673
1980	-0.0070
1981	-0.0142
1982	-0.0123
1983	-0.0607**
1984	0.0030
1985	-0.0296*
1986	-0.0271
1987	-0.0267
1988	-0.0110
1989	-0.0186*
1990	-0.0692**
1991	-0.0735**
1992	-0.0375

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The number of observations are 8,029. Fluoride is dependent variable.

A third category of predetermined characteristics concerns cohorts. Assume that people suddenly become very concerned about fluoride, and moves from high fluoride areas. If that is the case, later cohorts would have a lower fluoride level than older cohorts. We test this in Table A7, with cohort 1985 as benchmark. We also include sibling order for those with at least one sibling (twins removed). We have three statistically significant results, but the point estimates are very small. Those born in 1992 received on average 0.007 mg/l

lower fluoride than those born in 1985. In terms of economic significance, this is a zero-effect and below the measurable precision level of fluoride.

Table A7. *Balance test. Cohorts and sibling order*

	F. (0.1 mg/l)
Cohort 1986	0.00691
	(0.0119)
Cohort 1987	-0.00783
	(0.0146)
Cohort 1988	0.00542
	(0.0161)
Cohort 1989	-0.00657
	(0.0154)
Cohort 1990	-0.0360**
	(0.0165)
Cohort 1991	-0.0208
	(0.0180)
Cohort 1992	-0.0744***
	(0.0201)
Sibling order	0.0415*
	(0.0215)

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The number of observation is 731,888 for the cohorts and 419,558 for the sibling order regression. Fluoride is dependent variable.

Another concern would be that high cognitive ability individuals, who were exposed to lower dosages of fluoride, were able to avoid enlistment, meaning that when we run the analysis we only estimate the effect for a biased sample. Therefore we run balance tests to see if the fluoride treatment level for men without cognitive and non-cognitive ability scores differs from those who enlisted. We also run the test for taking the math test in ninth grade (for both males and females). In conclusion, there is no evidence of such sorting.

Table A8. *Balance test. Missing test scores*

	F. (0.1 mg/l)
No Cog. ab.	0.000742
	(0.000797)
No Non-Cog. ab.	-0.000155
	(0.000307)
No math test	-0.000168
	(0.000911)

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each row is a separate regression, where the dependent variable is displayed at the row. The number of observations for the two first outcomes are 376,402 and for the last outcome 569,648.

In Table A10, we have regressed the search intensity (data from Google Trends) on the fluoride level on the county level. The reader should note that Google does not provide data if the number of searches has been too low in an area. We have downloaded data for various search words in Swedish between 2004 and August 2016. More specifically we have run the analysis for *Fluor, Fluor - kemiskt ämne, Dricksvatten* and *Fluorid*. *Fluor* is the Swedish everyday word used for the chemical compound fluoride. *Dricksvatten* is Swedish for Drinking Water.

We only find one statistically significant result. People living in areas with higher fluoride seems use the word for drinking water more in their searches. We do not however find any evidence that they search more for fluoride, which is reassuring. The reader should note that we have no information about the number of searches, meaning that relative search intensity may still be based on very few actual searches.

Table A9 of the sales of bottled water discussed in the empirical framework section is also presented here.

Table A9. *Bottled water sales*

	Bott. wat. l./inh.
1994	12.13
1995	13.16
1996	13.00
1997	14.31
1998	14.25
1999	16.18
2000	16.95
2001	18.06
2002	19.52
2003	20.76
2004	22.03
2005	25.02
2006	29.34
2007	27.95
2008	23.90
2009	21.91
2010	22.01
2011	22.27
2012	22.43
2013	23.35
2014	24.38
2015	23.50

Notes: This data comes from the Swedish Brewers Association, *Sveriges Bryggerier*.

Table A10. *Google searches*

	F. (0.1 mg/l)
Drinking water	0.814**
	(0.338)
Fluor, chemical	0.719
	(0.699)
Fluor, search	0.720
	(0.468)
Fluoride	1.329
	(0.805)

Notes: Data from Google trends. Number of observations depends on whether Google Trends display searches for each county. The number of observations ranges between 752 and 8,370. Each outcome has a maximum of 100 and displays the relative search intensity on the county level in Sweden. 50 means that the word was half as popular and 1 means that the search word was 1 percent as popular in comparison to where it was the most popular.

E Results: Effects of fluoride on dental health

Table A11. *Unweightened regressions dental outcomes*

	CheckUps	DentalSurgery	Orthognathic	Prosthesis	OrthodontReplace	DiCheckUpsEval	DiDentHealth	DiDiseasePain	DiRepairs	DiRehabHab	MedianRemaining	MedianIntact
2013	-0.745**	0.0215	-0.0509*	-0.00810	-0.00641	-0.688**	-0.371*	-0.614**	-0.531***	-0.0208	-0.0127	0.0135
	(0.330)	(0.0451)	(0.0292)	(0.00902)	(0.0280)	(0.302)	(0.205)	(0.262)	(0.193)	(0.0290)	(0.0101)	(0.0194)
2008	-0.714**	-0.0856***	-0.0323*	0.0141	-0.00386	-0.677**	-0.229	-0.120	-0.279***	-0.0116	-0.0718**	-0.0186
	(0.345)	(0.0308)	(0.0169)	(0.0167)	(0.00312)	(0.320)	(0.194)	(0.117)	(0.0722)	(0.0154)	(0.0329)	(0.0449)

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The number of observations ranges between 7,386 and 7,622 for 2013 and between 7,352 and 7,606 for 2008.

Table A12. *Dental outcomes 2013. Additional specifications. Weighted regressions*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CheckUps	-0.3635*	-0.0626	-0.0101	-0.0159	0.0227	0.0139	0.0202
	(0.2016)	(0.0550)	(0.0512)	(0.0503)	(0.0388)	(0.0397)	(0.0403)
DentalSurgery	0.0093	-0.0160	-0.0046	-0.0039	-0.0206	-0.0202	-0.0230
	(0.0307)	(0.0125)	(0.0163)	(0.0161)	(0.0151)	(0.0158)	(0.0149)
Orthognathic	-0.0250**	-0.0069*	-0.0075	-0.0076*	-0.0028	-0.0012	-0.0012
	(0.0098)	(0.0038)	(0.0047)	(0.0046)	(0.0043)	(0.0055)	(0.0055)
Prosthesis	-0.0176***	-0.0108***	-0.0161***	-0.0156***	-0.0114***	-0.0094***	-0.0096***
	(0.0043)	(0.0022)	(0.0030)	(0.0030)	(0.0028)	(0.0030)	(0.0030)
OrthodontReplace	-0.0051**	-0.0021*	-0.0031**	-0.0031**	-0.0018	-0.0012	-0.0011
	(0.0024)	(0.0011)	(0.0015)	(0.0015)	(0.0015)	(0.0017)	(0.0017)
DiCheckUpsEval	-0.3032*	-0.0671	-0.0126	-0.0174	0.0062	-0.0042	0.0002
	(0.1685)	(0.0478)	(0.0444)	(0.0438)	(0.0345)	(0.0360)	(0.0364)
DiDentHealth	-0.1990	-0.0252	0.0026	0.0005	0.0017	0.0095	0.0100
	(0.1325)	(0.0305)	(0.0294)	(0.0295)	(0.0232)	(0.0260)	(0.0261)
DiDiseasePain	-0.2500*	-0.0829*	-0.0642	-0.0633	-0.0557*	-0.0605*	-0.0614*
	(0.1396)	(0.0439)	(0.0394)	(0.0396)	(0.0337)	(0.0347)	(0.0348)
DiRepairs	-0.1770*	-0.1034***	-0.1049**	-0.1028**	-0.0973***	-0.0831**	-0.0884**
	(0.0929)	(0.0375)	(0.0449)	(0.0450)	(0.0370)	(0.0391)	(0.0374)
DiRehabHab	-0.0121**	-0.0095***	-0.0114***	-0.0114***	-0.0095***	-0.0082**	-0.0084**
	(0.0050)	(0.0026)	(0.0035)	(0.0035)	(0.0033)	(0.0034)	(0.0034)
MedianRemaining	-0.0172**	-0.0085***	-0.0133***	-0.0128***	-0.0078***	-0.0066***	-0.0065***
	(0.0069)	(0.0021)	(0.0023)	(0.0026)	(0.0018)	(0.0018)	(0.0018)
MedianIntact	-0.0165	-0.0038	-0.0125*	-0.0131*	-0.0049	-0.0058	-0.0045
	(0.0196)	(0.0066)	(0.0076)	(0.0075)	(0.0056)	(0.0055)	(0.0050)
Small set covariates	No	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	No	Yes
Fe. birth muni.	No	No	Yes	Yes	Yes	Yes	Yes
Fe. cohort	No	No	No	Yes	Yes	Yes	Yes
Fe. muni. 2014	No	Yes	No	No	Yes	Yes	Yes
Sample	All	All	All	All	All	Col 7	All
Observations	720,401	720,401	720,401	720,401	720,401	469,207	469,207

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Outcomes on each row. The number of observations ranges between 469,207 and 725,004.

Table A13. *Dental outcomes 2008. Main outcomes. Weighted regressions*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Visit	-2.3819**	-0.0094	-0.0544	-0.1228	0.3412	0.2654	0.3253
	(0.9978)	(0.2545)	(0.3992)	(0.3900)	(0.3377)	(0.3446)	(0.3417)
Repair	-0.4461	-0.3960*	-0.3079	-0.2778	-0.3676	-0.4719	-0.4972
	(0.4539)	(0.2015)	(0.3277)	(0.3278)	(0.2970)	(0.3178)	(0.3098)
RiskEvaluation	-2.5889**	-0.0158	-0.0938	-0.1646	0.3230	0.2402	0.3040
	(1.0831)	(0.2649)	(0.4114)	(0.4011)	(0.3465)	(0.3556)	(0.3562)
DiseasePrevention	-2.7806*	0.2148	0.2625	0.2434	0.1689	0.1820	0.2176
	(1.5433)	(0.2577)	(0.5424)	(0.5425)	(0.3500)	(0.3721)	(0.3665)
DiseaseTreatment	0.7981	0.0019	-0.2339	-0.1992	-0.3082	-0.4745*	-0.4807*
	(0.6791)	(0.1626)	(0.2517)	(0.2506)	(0.2360)	(0.2761)	(0.2755)
RootCanal	-0.1575	-0.0721	-0.1270	-0.1114	-0.0525	-0.0334	-0.0432
	(0.1006)	(0.0481)	(0.0796)	(0.0803)	(0.0720)	(0.0808)	(0.0804)
Small set covariates	No	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	No	Yes
Fe. birth muni.	No	No	Yes	Yes	Yes	Yes	Yes
Fe. cohort	No	No	No	Yes	Yes	Yes	Yes
Fe. muni. 2014	No	Yes	No	No	Yes	Yes	Yes
Sample	All	All	All	All	All	Col 7	All

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Outcomes on each row. The number of observations ranges between 209,468 and 335,687.

Table A14. *Dental outcomes 2008. Additional specifications. Weighted regressions*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CheckUps	-2.8652**	0.1945	0.0302	-0.0574	0.5416	0.4332	0.5130
	(1.2202)	(0.2930)	(0.4519)	(0.4403)	(0.3832)	(0.3935)	(0.3935)
DentalSurgery	-0.2571	-0.2090***	-0.3171***	-0.2915***	-0.3022***	-0.3260***	-0.3415***
	(0.1753)	(0.0784)	(0.1079)	(0.1080)	(0.1062)	(0.1226)	(0.1216)
Orthognathic	-0.1309**	0.0207	-0.0661	-0.0649	0.0040	-0.0086	-0.0060
	(0.0548)	(0.0311)	(0.0403)	(0.0405)	(0.0420)	(0.0503)	(0.0501)
Prosthesis	-0.0251	0.0066	-0.0278	-0.0237	0.0011	0.0232	0.0227
	(0.0379)	(0.0253)	(0.0348)	(0.0349)	(0.0339)	(0.0414)	(0.0413)
OrthodontReplace	-0.0294*	-0.0308***	-0.0392***	-0.0396***	-0.0375***	-0.0388***	-0.0385***
	(0.0162)	(0.0081)	(0.0112)	(0.0112)	(0.0121)	(0.0147)	(0.0147)
DiCheckUpsEval	-2.5889**	-0.0158	-0.0938	-0.1646	0.3230	0.2402	0.3040
	(1.0831)	(0.2649)	(0.4114)	(0.4011)	(0.3465)	(0.3556)	(0.3562)
DiDentHealth	-1.3861	0.3730	0.5994	0.5900	0.2934	0.3275	0.3626
	(1.2635)	(0.2265)	(0.4893)	(0.4889)	(0.2995)	(0.3302)	(0.3269)
DiDiseasePain	-0.7863	-0.1631	-0.5904**	-0.5555*	-0.3587	-0.5330**	-0.5378**
	(0.5878)	(0.1776)	(0.2912)	(0.2902)	(0.2449)	(0.2692)	(0.2688)
DiRepairs	-0.5358	-0.4949**	-0.4261	-0.3908	-0.5116	-0.6089*	-0.6391*
	(0.4692)	(0.2129)	(0.3458)	(0.3460)	(0.3164)	(0.3412)	(0.3311)
DiRehabHab	-0.0636	-0.0266	-0.0427	-0.0426	-0.0289	-0.0059	-0.0067
	(0.0479)	(0.0273)	(0.0386)	(0.0386)	(0.0377)	(0.0466)	(0.0468)
MedianRemaining	-0.4245***	-0.0497***	-0.2175***	-0.2136***	-0.0365**	-0.0283	-0.0295
	(0.1457)	(0.0149)	(0.0590)	(0.0596)	(0.0183)	(0.0209)	(0.0209)
MedianIntact	-0.0759	0.1321***	0.0627	0.0551	0.0901*	0.1057*	0.1168**
	(0.2200)	(0.0369)	(0.0684)	(0.0688)	(0.0517)	(0.0550)	(0.0539)
Small set covariates	No	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	No	Yes
Fe. birth muni.	No	No	Yes	Yes	Yes	Yes	Yes
Fe. cohort	No	No	No	Yes	Yes	Yes	Yes
Fe. muni. 2014	No	Yes	No	No	Yes	Yes	Yes
Sample	All	All	All	All	All	Col 7	All

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Outcomes on each row. The number of observations ranges between 208,245 and 335,687.

In Table A15, we run the dental regressions for older cohorts to investigate further the effect on the median of remaining teeth and the median of intact teeth.[33] In our main analysis, we found effects that sometimes pointed in the opposite direction that we expected. In the analysis below, we use data for older cohorts. This data is only available to us on the municipal level because it is not part of our main dental dataset which only includes cohorts born 1985-1992. The analysis is based on the assumption that those people living in a municipality in 2013 have also lived there for a longer period of time. The results from the analysis should thus be interpreted with caution. We find that the median of intact teeth now points in the expected direction, namely that increased fluoride increases the median of intact teeth in a municipality. This is reassuring given that intact teeth should be more closely related to dental health status that could be affected by fluoride. For remaining teeth we still have results that points in an opposite direction than expected. However, no

[33]The data originates from the open data published at the website of The National Board Board of Health and Welfare.

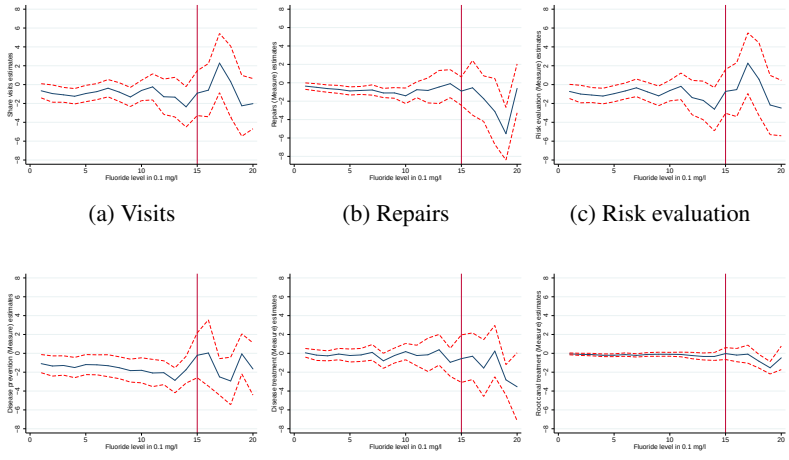
point estimates are statistically significant with the exception of one that is significant at the 10 percent level.

Table A15. *Dental outcomes. Older cohorts. Aggregated data*

	Remaning teeth	Intact teeth
F. (0.1 mg/l)	-0.0450*	0.0304
	(0.0269)	(0.0247)
F. (0.1 mg/l)	-0.0609	0.0319
	(0.0397)	(0.0234)

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1. First row is for people age 40-90 years old. The second row is for individuals aged 60-90 years old. The dependent variable is displayed at the top of each column. The number of observations are 8,597. The outcome is aggregated and measured at the municipal level.

F Results: Non-linear effects. Dental health



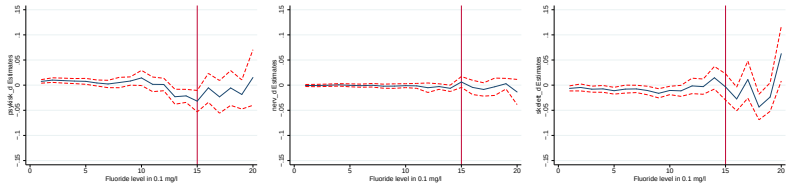
(a) Visits (b) Repairs (c) Risk evaluation

(d) Disease prevention (e) Disease treatment (f) Root canal

Figure A3. Non-linear effects: Dental health estimates.

G Results: Non-linear effects. Additional health outcomes

(a) ADHD medicine (b) Depression medicine (c) Psychiatric medicine



(d) Psychiatric diagnosis (e) Neurological diagnosis (f) Skeletal and muscular diagnosis

Figure A4. Non-linear effects: Additional health outcomes estimates.

H Results: Non-linear effects, regression tables. Main outcomes

Table A16. *Cognitive ability*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride 2nd quartile	0.1360**	0.0532	0.0505	0.0084	0.0528*	0.0161	0.0402
	(0.0662)	(0.0416)	(0.0421)	(0.0437)	(0.0282)	(0.0510)	(0.0470)
Fluoride 3nd quartile	-0.1649**	-0.0542	-0.0526	-0.0465	-0.0184	-0.0091	-0.0385
	(0.0712)	(0.0341)	(0.0339)	(0.0350)	(0.0256)	(0.0466)	(0.0553)
Fluoride 4nd quartile	0.0099	0.0197	0.0194	-0.0069	0.0042	0.0547	0.1086
	(0.0516)	(0.0262)	(0.0261)	(0.0335)	(0.0263)	(0.0433)	(0.0677)
Mean	5.006726	5.006726	5.006726	5.022206	5.022206	5.089748	4.924601
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	Col 5	All	SAMS stayers	SAMS movers
Observations	81,776	81,776	81,776	51,203	51,203	20,513	19,178

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A17. *Non-cognitive ability*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride 2nd quartile	-0.0188	-0.0542	-0.0546	-0.0749*	-0.0422	-0.0376	-0.0127
	(0.0656)	(0.0341)	(0.0340)	(0.0388)	(0.0344)	(0.0619)	(0.0623)
Fluoride 3nd quartile	-0.0687	0.0182	0.0186	0.0313	0.0539*	0.0913*	0.0866
	(0.0663)	(0.0313)	(0.0311)	(0.0354)	(0.0304)	(0.0522)	(0.0777)
Fluoride 4nd quartile	0.0608	0.0267	0.0270	0.0273	0.0367	0.0419	0.1574**
	(0.0428)	(0.0255)	(0.0255)	(0.0357)	(0.0331)	(0.0559)	(0.0634)
Mean	4.733996	4.733996	4.733996	4.775411	4.775411	4.921403	4.6953
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	Col 5	All	SAMS stayers	SAMS movers
Observations	66,375	66,375	66,375	41,636	41,636	16,731	15,425

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A18. *Math points*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride 2nd quartile	-0.0314	-0.2692**	-0.2558*	-0.2558*	-0.3340**	-0.1886*	-0.0878	-0.2538*
	(0.2729)	(0.1348)	(0.1374)	(0.1374)	(0.1328)	(0.0989)	(0.1487)	(0.1513)
Fluoride 3nd quartile	-0.9200***	-0.3043**	-0.3031**	-0.3029**	-0.2915**	-0.1373	0.0764	-0.1384
	(0.3260)	(0.1202)	(0.1187)	(0.1186)	(0.1311)	(0.1045)	(0.1347)	(0.1261)
Fluoride 4nd quartile	0.0789	0.1104	0.1186	0.1186	0.0015	0.0967	-0.0059	0.1525
	(0.2537)	(0.0949)	(0.0965)	(0.0965)	(0.0934)	(0.0929)	(0.1060)	(0.1246)
Mean	26.20586	26.20586	26.20586	26.20586	26.48997	26.48997	27.22212	26.04409
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	499,892	499,892	499,892	499,892	336,827	336,827	139,149	127,062

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A19. *Annual log income in SEK*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride 2nd quartile	-0.0224	0.0074	-0.0210**	-0.0138	-0.0162	-0.0128	0.0073	0.0268
	(0.0290)	(0.0107)	(0.0106)	(0.0100)	(0.0104)	(0.0099)	(0.0196)	(0.0166)
Fluoride 3nd quartile	0.0394	0.0112	0.0065	0.0130	0.0098	0.0122	0.0194	0.0247*
	(0.0255)	(0.0081)	(0.0064)	(0.0119)	(0.0123)	(0.0125)	(0.0197)	(0.0133)
Fluoride 4nd quartile	0.0194	0.0127**	0.0207***	0.0214***	0.0195***	0.0184***	0.0167	0.0022
	(0.0150)	(0.0059)	(0.0057)	(0.0055)	(0.0060)	(0.0059)	(0.0168)	(0.0119)
Mean	11.91243	11.91243	11.91243	11.91243	11.92288	11.92288	11.84519	11.9544
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	634,793	634,793	634,793	634,793	419,162	419,162	72,089	150,458

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A20. *Employment status*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride 2nd quartile	-0.0052	0.0038	-0.0047	-0.0024	-0.0032	-0.0016	0.0004	0.0104
	(0.0121)	(0.0045)	(0.0044)	(0.0040)	(0.0043)	(0.0040)	(0.0077)	(0.0074)
Fluoride 3nd quartile	0.0107	0.0020	0.0005	0.0027	0.0023	0.0034	-0.0006	0.0119**
	(0.0109)	(0.0034)	(0.0030)	(0.0046)	(0.0045)	(0.0045)	(0.0080)	(0.0056)
Fluoride 4nd quartile	0.0107	0.0074***	0.0098***	0.0113***	0.0104***	0.0098***	0.0121*	0.0072
	(0.0074)	(0.0027)	(0.0028)	(0.0027)	(0.0028)	(0.0027)	(0.0073)	(0.0057)
Mean	.7346382	.7346382	.7346382	.7346382	.7458825	.7458825	.7129002	.7582255
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	728,074	728,074	728,074	728,074	474,556	474,556	81,867	170,142

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

I Robustness analysis: Analysis with adoptees only

Table A21. *Cognitive ability, adopted*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride up until age 18 (0.1 mg/l)	-0.0207	-0.0451	-0.0472	0.0317	0.0436	-0.1027	-0.2074
	(0.0218)	(0.0645)	(0.0651)	(0.0692)	(0.0782)	(0.3207)	(0.2184)
Mean	4.294677	4.294677	4.294677	4.328671	4.328671	4.160714	4.456522
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	Col 5	All	SAMS stayers	SAMS movers
Observations	526	526	526	286	286	112	92

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A22. *Non-cognitive ability, adopted*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride up until age 18 (0.1 mg/l)	-0.0271	0.0302	0.0236	-0.0359	-0.0405	-0.1255	-0.0914
	(0.0206)	(0.0648)	(0.0645)	(0.0890)	(0.0878)	(0.2728)	(0.1546)
Mean	4.4914	4.4914	4.4914	4.671233	4.671233	4.592593	4.685714
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	Col 5	All	SAMS stayers	SAMS movers
Observations	407	407	407	219	219	81	70

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A23. *Math points, adopted*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until age 16 (0.1 mg/l)	-0.0387	-0.1384	-0.1467	-0.1488	-0.0992	-0.0913	-0.1310	0.0019
	(0.0934)	(0.1325)	(0.1308)	(0.1310)	(0.1614)	(0.1550)	(0.2505)	(0.3810)
Mean	23.74629	23.74629	23.74629	23.74629	24.07754	24.07754	24.70705	23.52427
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	2,089	2,089	2,089	2,089	1,251	1,251	553	412

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A24. *Annual log income, adopted*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	0.0138**	0.0045	0.0043	-0.0027	0.0008	-0.0008	0.0720	-0.0115
	(0.0070)	(0.0092)	(0.0090)	(0.0104)	(0.0136)	(0.0139)	(0.0554)	(0.0411)
Mean	11.86561	11.86561	11.86561	11.86561	11.85763	11.85763	11.69303	11.8584
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	3,176	3,176	3,176	3,176	1,714	1,714	306	565

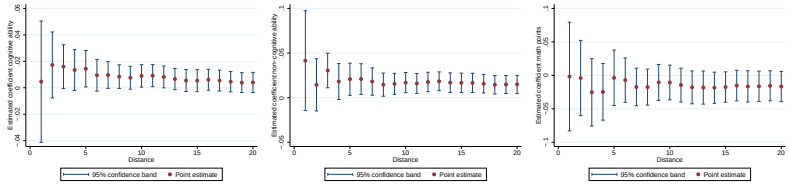
Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A25. *Employment status, adopted*

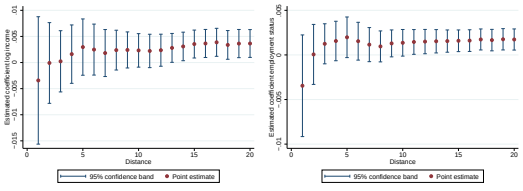
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	0.0013	-0.0005	-0.0008	-0.0004	0.0059	0.0061	0.0110	0.0116
	(0.0026)	(0.0045)	(0.0044)	(0.0046)	(0.0062)	(0.0064)	(0.0206)	(0.0087)
Mean	.7005768	.7005768	.7005768	.7005768	.696837	.696837	.6005435	.7016248
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	3,814	3,814	3,814	3,814	2,055	2,055	368	677

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

J Robustness analysis: Distance of SAMS



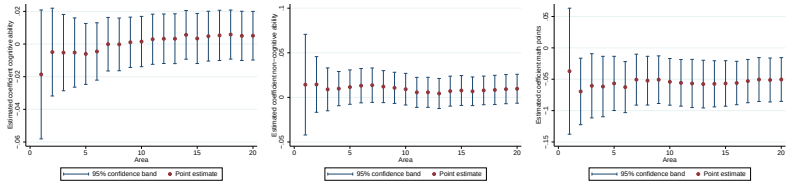
(a) Cognitive ability (b) Non-Cognitive ability (c) Math points



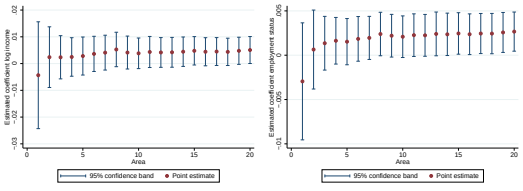
(d) Annual log income (e) Employment status

Figure A5. Estimates for different geographical distances from water plant. The X-axis corresponds to distances in kilometers between water plant and the center point of the SAMS.

K Robustness analysis: Area of SAMS



(a) Cognitive ability (b) Non-Cognitive ability (c) Math points



(d) Annual log income (e) Employment status

Figure A6. Estimates for different geographical areas SAMS. The X-axis corresponds to areas in square kilometers.

L Robustness analysis: Confirmed water source

Table A26. *Cognitive ability, confirmed water source since 1985*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride up until age 18 (0.1 mg/l)	-0.0187*	0.0091	0.0087	0.0122	0.0176**	0.0025	0.0375**
	(0.0109)	(0.0081)	(0.0080)	(0.0077)	(0.0084)	(0.0087)	(0.0187)
Mean	4.974421	4.974421	4.974421	4.972386	4.972386	5.078782	4.862705
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	Col 5	All	SAMS stayers	SAMS movers
Observations	18,922	18,922	18,922	12,204	12,204	6,042	5,317

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A27. *Non-cognitive ability, confirmed water source since 1985*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fluoride up until age 18 (0.1 mg/l)	-0.0038	0.0086	0.0086	0.0165	0.0248	0.0234*	0.0192
	(0.0096)	(0.0121)	(0.0121)	(0.0147)	(0.0154)	(0.0123)	(0.0276)
Mean	4.77522	4.77522	4.77522	4.817776	4.817776	4.951318	4.670572
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	Col 5	All	SAMS stayers	SAMS movers
Observations	15,246	15,246	15,246	9,856	9,856	4,930	4,268

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A28. *Math points, confirmed water source since 1985*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until age 16 (0.1 mg/l)	-0.2401***	-0.0423	-0.0436	-0.0437	-0.0629**	-0.0182	0.0027	-0.0480
	(0.0558)	(0.0288)	(0.0270)	(0.0270)	(0.0282)	(0.0261)	(0.0249)	(0.0366)
Mean	26.35896	26.35896	26.35896	26.35896	26.53781	26.53781	27.26578	25.83514
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	113,378	113,378	113,378	113,378	79,497	79,497	40,402	34,618

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A29. *Annual log income, confirmed water source since 1985*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	0.0057	0.0012	0.0028*	0.0027*	0.0011	0.0010	0.0047	0.0037
	(0.0042)	(0.0018)	(0.0015)	(0.0017)	(0.0020)	(0.0020)	(0.0036)	(0.0029)
Mean	11.94695	11.94695	11.94695	11.94695	11.95188	11.95188	11.84664	11.97675
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	145,385	145,385	145,385	145,385	99,557	99,557	20,511	40,975

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A30. *Employment status, confirmed water source since 1985*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	0.0020	0.0008	0.0012*	0.0013	0.0007	0.0007	0.0013	0.0029*
	(0.0019)	(0.0007)	(0.0007)	(0.0009)	(0.0011)	(0.0011)	(0.0012)	(0.0016)
Mean	.7524632	.7524632	.7524632	.7524632	.7609301	.7609301	.712957	.7686438
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	164,626	164,626	164,626	164,626	111,641	111,641	23,223	46,262

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

M Robustness analysis: Only those born in 1985

Table A31. *Annual log income, cohort 1985*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	-0.0014	-0.0027	-0.0027	0.0027	0.0020	0.0029	-0.0030	-0.0018
	(0.0015)	(0.0019)	(0.0019)	(0.0021)	(0.0025)	(0.0025)	(0.0150)	(0.0078)
Mean	12.22359	12.22359	12.22359	12.22359	12.23666	12.23666	12.25366	12.24548
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	70,114	70,114	70,114	70,114	41,544	41,544	1,977	13,083

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A32. *Employment status, cohort 1985*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	0.0007	0.0001	0.0001	0.0018**	0.0013	0.0016	-0.0007	0.0047**
	(0.0009)	(0.0008)	(0.0008)	(0.0009)	(0.0010)	(0.0010)	(0.0041)	(0.0021)
Mean	.8374533	.8374533	.8374533	.8374533	.8529284	.8529284	.8105082	.8553713
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	79,005	79,005	79,005	79,005	46,168	46,168	2,322	14,596

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

N Robustness analysis: Confirmed water source and only one water plant within SAMS, non-movers

Table A33. *Cognitive ability*

	(1)	(2)	(3)	(4)	(5)
Fluoride up until age 18 (0.1 mg/l)	-0.0188	0.0123	0.0120	0.0091	0.0091
	(0.0111)*	(0.0168)	(0.0165)	(0.0180)	(0.0180)
Mean	4.9905	4.9905	4.9905	4.9144	4.9144
Birth cohort FE	No	No	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	Yes	Yes
Sample	All	All	All	Col 5	All
Observations	1992	1992	1992	1285	1285

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A34. *Non-cognitive ability*

	(1)	(2)	(3)	(4)	(5)
Fluoride up until age 18 (0.1 mg/l)	-0.0134	0.0071	0.0073	0.0137	0.0137
	(0.0136)	(0.0134)	(0.0134)	(0.0182)	(0.0182)
Mean	4.8369	4.8369	4.8369	4.8711	4.8711
Birth cohort FE	No	No	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	Yes	Yes
Sample	All	All	All	Col 5	All
Observations	1,625	1,625	1,625	1,055	1,055

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A35. *Math points*

	(1)	(2)	(3)	(4)	(5)	(6)
Fluoride up until age 16 (0.1 mg/l)	-0.0457	0.0463	0.0412	0.0408	0.0104	0.0036
	(0.0192)**	(0.0273)*	(0.0270)	(0.0270)	(0.0298)	(0.0247)
Mean	26.6661	26.6661	26.6661	26.6661	26.8053	26.8053
Birth cohort FE	No	No	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes
Sample	All	All	All	All	Col 6	All
Observations	12,661	12,661	12,661	12,661	9,164	9,164

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A36. *Annual log income*

	(1)	(2)	(3)	(4)	(5)	(6)
Fluoride up until year 2014 (0.1 mg/l)	-0.0042	0.0022	0.0026	0.0024	0.0020	0.0029
	(0.0048)	(0.0045)	(0.0044)	(0.0039)	(0.0060)	(0.0060)
Mean	11.9282	11.9282	11.9282	11.9282	11.9345	11.9345
Birth cohort FE	No	No	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes
Sample	All	All	All	All	Col 6	All
Observations	6,955	6,955	6,955	6,955	5,035	5,035

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A37. *Employment status*

	(1)	(2)	(3)	(4)	(5)	(6)
Fluoride up until year 2014 (0.1 mg/l)	-0.0013	-0.0009	-0.0009	-0.0010	-0.0008	-0.0007
	(0.0012)	(0.0018)	(0.0019)	(0.0018)	(0.0019)	(0.0018)
Mean	0.7474	0.7474	0.7474	0.7474	0.7502	0.7502
Birth cohort FE	No	No	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes	Yes
Sample	All	All	All	All	Col 6	All
Observations	7,802	7,802	7,802	7,802	5,616	5,616

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

O Robustness analysis: Alternative income measure

Table A38. *Log income, "förvärvsinkomst"*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fluoride up until year 2014 (0.1 mg/l)	0.0063*	0.0040**	0.0046***	0.0045***	0.0034**	0.0034**	0.0034*	0.0013
	(0.0035)	(0.0017)	(0.0016)	(0.0017)	(0.0015)	(0.0013)	(0.0021)	(0.0042)
Mean	11.99991	11.99991	11.99991	11.99991	12.01073	12.01073	11.88782	12.04571
Birth cohort FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes	Yes	Yes
Sample	All	All	All	All	Col 6	All	SAMS stayers	SAMS movers
Observations	641,629	641,629	641,629	641,629	423,411	423,411	72,861	151,885

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

P Robustness analysis: Analysis with sibling fixed effects

Table A39. *Cognitive ability*

	(1)	(2)	(3)	(4)	(5)
Fluoride up until age 18 (0.1 mg/l)	-0.2302	-0.2354	-0.2074	-0.3170	-0.2894
	(0.6207)	(0.7068)	(0.6598)	(0.8508)	(0.8524)
Mean	5.049126	5.049126	5.049126	5.096304	5.096304
Birth cohort FE	No	No	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes
Sample	All	All	All	Col 5	All
Observations	46,208	46,208	46,208	32,439	32,439

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A40. *Non-cognitive ability*

	(1)	(2)	(3)	(4)	(5)
Fluoride up until age 18 (0.1 mg/l)	-0.3620	-0.2547	-0.2314	-0.2583	-0.2316
	(0.9665)	(1.0682)	(1.0435)	(1.4663)	(1.3804)
Mean	4.775179	4.775179	4.775179	4.826302	4.826302
Birth cohort FE	No	No	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes
Large set covariates	No	No	No	No	Yes
Sample	All	All	All	Col 5	All
Observations	37,492	37,492	37,492	26,454	26,454

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A41. *Math points*

	(1)	(2)	(3)	(4)	(5)	(6)
Fluoride up until age 16 (0.1 mg/l)	0.1369	0.0802	0.0554	0.0553	0.0912	0.1062
	(0.1527)	(0.1656)	(0.1688)	(0.1689)	(0.2073)	(0.2019)
Mean	26.23297	26.23297	26.23297	26.23297	26.50438	26.50438
Birth cohort FE	No	No	Yes	Yes	Yes	Yes
Municipal FE, age 0-16	No	Yes	Yes	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes
Sample	All	All	All	All	Col 6	All
Observations	306,834	306,834	306,834	306,834	216,311	216,311

Notes: Standard errors clustered at the municipal level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A42. *Annual log income*

	(1)	(2)	(3)	(4)	(5)	(6)
Fluoride up until year 2014 (0.1 mg/l)	-0.0421***	-0.0393***	-0.0130**	-0.0088	-0.0098	-0.0100
	(0.0075)	(0.0071)	(0.0065)	(0.0093)	(0.0088)	(0.0088)
Mean	11.92662	11.92662	11.92662	11.92662	11.94066	11.94066
Birth cohort FE	No	No	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes
Sample	All	All	All	All	Col 6	All
Observations	380,077	380,077	380,077	380,077	267,436	267,436

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A43. *Employment status*

	(1)	(2)	(3)	(4)	(5)	(6)
Fluoride up until year 2014 (0.1 mg/l)	-0.0171***	-0.0161***	-0.0081***	-0.0039	-0.0029	-0.0029
	(0.0029)	(0.0029)	(0.0026)	(0.0033)	(0.0033)	(0.0033)
Mean	.7415351	.7415351	.7415351	.7415351	.7523387	.7523387
Birth cohort FE	No	No	Yes	Yes	Yes	Yes
Birth municipal FE	No	Yes	Yes	Yes	Yes	Yes
Municipal FE, year 2014	No	No	No	Yes	Yes	Yes
Small set covariates	No	No	No	Yes	Yes	Yes
Large set covariates	No	No	No	No	No	Yes
Sample	All	All	All	All	Col 6	All
Observations	433,587	433,587	433,587	433,587	301,666	301,666

Notes: Standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Q ATC-codes and diagnostic codes

Table A44 and A45 This is a list for the ATC-codes and the diagnostic codes (on the chapter level) we have used for our health outcomes.

Table A44. *ATC codes for medicines*

Medicine	**ATC**
ADHD	N06B
Antidepressants	N06A
Neuroleptics	N05A

Table A45. *ICD codes for diagnoses*

Diagnosis	**ICD10**
Psychiatric	F
Neurological	G
Skeleton and muscular	M

II. Be Smart, Live Long: The Relationships between Cognitive and Non-Cognitive Abilities and Mortality

Acknowledgments: I would like to thank Damon Clark, Michael Grossman, Per-Anders Edin, Olof Åslund, Gawain Heckley and the participants at the NHESG 2014 meeting in Reykjavík, the HEFUU the U-CARE seminars at Uppsala university, as well as Erik Grönqvist, Matz Dahlberg and Linuz Aggeborn for helpful discussions, comments and suggestions.

1 Introduction

For most of us, questions on how to improve our health are of great personal interest. We want to know how to live long and healthy. However, these questions are of wider importance than the narrow self-interest. The political debate in many countries revolves around the public health in one way or another; governments want to know how to improve the health of the population. There are, of course, many good reasons for this focus on health. One common argument is that a healthy population is more productive and is less of a burden on the health care. The implications of knowing how to improve the general health of the population are far-reaching.

Longevity is one of the most important measures of the public health, and in the recent decade researchers have put a lot of effort trying to find the determinants of longevity, and its counterpart, mortality. The three main channels discussed in the literature on the determinants of health and mortality are income, education and relative socioeconomic position (e.g. Marmot 2002; Deaton 2003; Cutler et al. 2006; Batty, Deary, et al. 2007; Vogl et al. 2011).[1] While no one is denying that income, or more general, wealth, is strongly positively correlated with health and longevity, the causal relationship is still debated (e.g. Lindahl 2005; Frijters et al. 2005; Cesarini et al. 2016).[2] The same is true for education. Some authors claim to find a negative causal effect of education on the risk of mortality (e.g. Lleras-Muney 2005; Kippersluis et al. 2011; Buckles et al. 2013), while others find a negligible effect (e.g. Albouy and Lequien 2009; Clark and Roayer 2013). But perhaps these channels are not the fundamental factors. We know from earlier literature that both income and education are related to underlying skills such as cognitive and non-cognitive ability (Lindqvist and Vestman 2011). Income and education could be just mediating factors for the innate abilities rather than causal factors in themselves. It is, for example, possible that cognitive and non-cognitive abilities promotes health behaviors which prolong life, such as exercise and non-smoking, commonly ascribed to education (Conti, Heckman, and Urzua 2010; Chiteji 2010).

Cognitive ability is usually defined and measured as IQ or the *g* factor.[3] IQ is considered to be an innate capacity to solve abstract problems, and is a well-established measure of intelligence. It should, however, be noted that it is possible to improve *at least* the measured IQ, which is why it is common

[1] Other important channels discussed are, among others, nutrition, public health (better water supplies, sanitation systems, etc.), vaccinations and other medical treatments (e.g. Cutler et al. 2006; Batty, Deary, et al. 2007).

[2] The life expectancy has greatly increased during the last hundred years in wealthy countries, and there is a strong association between the life expectancy in a country and GDP, see Cutler et al. (2006).

[3] The *g* factor is a concept introduced by Charles Spearman in the early 20th century reflecting the fact that an individual's performances in different cognitive tasks often are highly positively correlated.

in the literature to use cognitive skill rather than ability (Kautz et al. 2014). Non-cognitive ability, on the other hand, is not as well defined. What is often meant is personality and social and emotional traits (Heckman, Stixrud, et al. 2006; Cunha et al. 2010; Lindqvist and Vestman 2011), and this is how I define it in this paper. In line with the earlier literature I consider non-cognitive ability as something distinct from what is measured by cognitive ability. As with cognitive ability, non-cognitive ability is partly an innate trait, but can be improved by training.[4]

The epidemiological literature suggests a negative association between cognitive ability and mortality (Hemmingsson, Melin, et al. 2006; Deary and Batty 2007; Batty, Deary, et al. 2007; Batty, Wennerstad, et al. 2009; Batty, Gale, et al. 2009; Hemmingsson, Melin, et al. 2009; Lager et al. 2009; Calvin et al. 2011). Epidemiologists have also found that cognitive ability is associated with less severe health outcomes than death, such as schizophrenia and psychosis (David et al. 1997), but not with cancer (Batty, Wennerstad, et al. 2007) or coronary heart disease (Hemmingsson, Essen, et al. 2007). An emerging literature in economics has studied the relationship with cognitive and non-cognitive ability for various outcomes, such as success in the labor market (Bowles et al. 2001; Nyhus and Pons 2005; Heckman, Stixrud, et al. 2006; Lindqvist and Vestman 2011), how teachers' abilities can explain student performance (Grönqvist and Vlachos 2008), the intergenerational transmission of the abilities (Grönqvist, Öckert, et al. 2010), and how cognitive ability is related to risk aversion and impatience (Burks et al. 2009; Dohmen et al. 2010; Benjamin et al. 2013; Andersson et al. 2013).

Psychologists have naturally been interested in the effects of both cognitive and non-cognitive ability. Roberts et al. (2007) conducted a meta-analysis and reviewed the evidence in the psychology literature of the associations between personality traits (non-cognitive ability) on three outcomes, mortality, education and marital status. They find that the effect of the personality traits was equivalent with that of cognitive ability (IQ).[5]

Economists are becoming increasingly interested in the relationships between cognitive and non-cognitive ability and health. For example, Conti and Hansman (2013) find that non-cognitive ability is nearly as important as cognitive ability for explaining the education-health gradient. Heckman, Humphries, et al. (2014) find that non-cognitive ability is important not only for education choice and labor market outcomes, but also for health. Savelyev and Tan (2014) and Savelyev (2014) study socioemotional skills and longevity

[4] I will not distinguish between "ability" and "skill". It is nearly impossible to measure an innate capacity since the ability to solve more or less any task is affected by training and experience. The terms "abilities" and "skills" are often used interchangeably in the literature. Non-cognitive skill is sometimes called "socioemotional skills". Note that in all "non-cognitive" problems naturally some form of cognition must be involved.

[5] However, in comparison with the studies reviewed, I have a much large sample and arguably a better and more complete measure of personality.

for high IQ (above 140) individuals, with a focus on the causal effect of education on longevity and health behaviors. They find strong effects of personality skills on health and longevity for men but not for women. Baker et al. (2015) find that children who experience a "negative shock" on the development of non-cognitive skills experienced worse health and higher crime rates later in life. There are also evidence suggesting that there are heterogeneous effects. Basu et al. (2014) find that children with low non-cognitive abilities were affected negatively on later health behavior (for example smoking) when exposed to mixed-ability schools.

As the growing literature in epidemiology and economics suggests, the relationships between cognitive ability, non-cognitive ability and mortality has interested researchers in itself. The relationships tell us something about what is possible to do with policy, and where we should focus on health improving policies. But there is another reason to study these relationships as well. A common practice in economics is to use income and education as proxy measures for individual ability. This paper contributes to the literature in two ways. To my knowledge, this is by far the largest study using credible measures on both cognitive and non-cognitive ability studying these relationships. Compared to the earlier epidemiological literature which looks at the association between cognitive ability and mortality, I am also able to include measures of non-cognitive ability. Second, I can see how these relationships are affected when I include measures on income and education. This is a test of how well income and education capture individual ability, which is rare, as we seldom have access to good skill measures.

I use Swedish military enlistment data for measures of cognitive and non-cognitive ability and link this register with demographic variables and the year of death. The data consists of 692,303 men born between 1950 and 1965, enlisted between 1969 and 1983. This is almost the full male population during the sample period. There are 28,570 deaths in the sample between the years 1969 and 2009. The sample period ends when the oldest individuals are 59 years old and the youngest individuals 44 years old. In that sense I estimate the associations between cognitive and non-cognitive ability and *premature* mortality, as even the oldest possible age for an individual in the dataset is an early age of death in Sweden.

At the time, military enlistment in Sweden was mandatory for all young men. Enlistment usually took place in the year when the individual turned 18 years old and spanned over two days with tests of health status and, most important for this study, cognitive and non-cognitive ability. The cognitive ability test consisted of a non-standard IQ test, aiming at measuring the g factor, while the non-cognitive ability was measured by a psychologist during an interview. The aim of the interview was to assess the individual's ability to cope with stress and fulfill military service, and included assessment of, among other things, social skills, emotional stability and persistence.

The main results support the literature on the negative association between cognitive ability and mortality. However, the results suggest that non-cognitive ability is of even greater importance; the Cox proportional hazard models indicate that the association between the risk of mortality and non-cognitive ability is more than two times the association with cognitive ability when controlling for income and education. The abilities are related to both income and education; cognitive ability mainly with education, and non-cognitive ability mainly with income. In addition to these independent associations, results suggest that income and education act as mediators for the relation with mortality. Lastly, cognitive and non-cognitive abilities are important in the relation with mortality for individuals with low income or non-college education. Using income and education as proxy measures for individual ability may therefore miss the large variation within these groups. The results are mainly driven by the bottom of the distributions.

The paper is organized as follows. In the next section I provide a conceptual framework of the causal chains of interest. In section 3, I describe the data, present descriptive statistics and discuss study limitations. I then turn to the results in section 4. I discuss the findings in section 5, and section 6 concludes. In the appendix I present OLS results, and additional figures and tables describing the data.

2 Theoretical framework

In this section I introduce a conceptual framework and review the literature on the relationships of interest: cognitive and non-cognitive abilities, health, income and education.

Figure 1 presents a simple framework of the causal chains discussed in the literature, including cognitive and non-cognitive abilities.[6] In addition to the two abilities and mortality (our measure of health), it consists of two mediating paths: income and education. In this figure the relationships are assumed to go from education and income to "health". However, it should be noted that, theoretically, health matters for income and education (Grossman 1972; Deaton 2003). Individuals with very bad health cannot work or go to school. Health could also, possibly, matter for the skills. A brain damage, for example, clearly affects both the cognitive ability and the non-cognitive ability.

Empirically, however, causal effects are not easy to estimate, due to two-way causality, mediators and confounders. It is difficult to find exogenous variation to single out an effect. However, both cognitive and non-cognitive skills have been shown to be causally linked with income and education

[6]Relative socioeconomic position is excluded in the framework. It is a relatively common explanation for differences in health, but it is not obvious how to operationalize it. What is the individual socioeconomic position relative to?

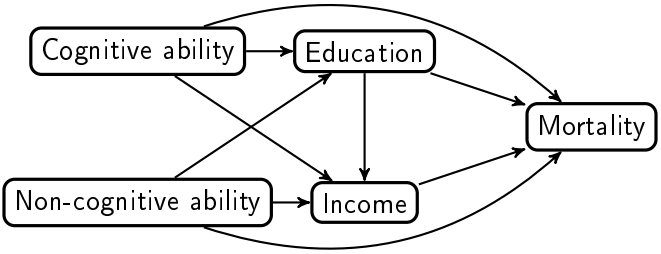


Figure 1. Relationships between cognitive and non-cognitive ability, mediators, and mortality.

(Bowles et al. 2001; Heckman, Stixrud, et al. 2006). Savelyev and Tan (2014) and Savelyev (2014) find that non-cognitive skill is linked with health and longevity for men with very high cognitive skill. If we interpret the epidemiological literature causally, cognitive skill has been shown to be linked with health and mortality (e.g. Batty, Wennerstad, et al. 2007, 2009).[7] In addition, Nyhus and Pons (2005) and Lindqvist and Vestman (2011) have shown that cognitive and non-cognitive skills are associated with success in the labor market.

As shown in Figure 1, cognitive and non-cognitive abilities are not only linked with health and mortality directly, but also through the mediators. If the mediators in fact are causally linked with health, at least a part of this effect stems from the abilities. As noted in the introduction the causal relationship between income, education and health are widely discussed and these questions are far from settled (Lindahl 2005; Frijters et al. 2005; Lleras-Muney 2005; Clark and Royer 2013; Fischer et al. 2013; Savelyev and Tan 2014; Savelyev 2014; Cesarini et al. 2016).

Empirical studies have shown that health and income affect cognitive skill, that is, the reverse of what is shown in Figure 1 (Currie 2009; Mani et al. 2013). However, these studies focus on very severe outcomes such as brain damage or extreme poverty, not common in Sweden, so I do not address this channel. Additionally, there is evidence of an association between birth-weight and nutrition in the childhood and cognitive functions in adult life (Sørensen et al. 1997; Gomez-Pinilla 2008; Bharadwaj et al. 2013). Studies have also found that socioeconomic factors or injuries such as head trauma are associated with lower cognitive ability (Batty, Deary, et al. 2007; Calvin et al. 2011). A number of studies present evidence that health affects both education and income (Contoyannis and Rice 2001; Currie 2009; Ding et al. 2009). It is well-established in the literature that higher education is causally linked with

[7]The epidemiological literature usually does not discuss the identification problem. To study causal effects epidemiologists often settle with controlling for potential mediators and confounders, a practice normally not endorsed in the econometric literature.

higher income (Angrist and Krueger 1991; Card 1999; Acemoglu and Angrist 2001).

Obviously, there are no direct effects of the skills on mortality. Rather, the links go through behavior which affects health. It is likely that an individual with a high cognitive ability in general is more prone to act in a way that promotes health, i.e., invests in health capital (cf. the Grossman model). We should also expect that an individual with a high non-cognitive ability (have a good social life, is calm, can cope with stress etc.) is more prone to engage in behavior which promotes health, or at least is more likely to avoid circumstances that are related with behaviors associated with bad health (Heckman, Pinto, et al. 2013; Conti, Heckman, and Pinto 2015).

3 Data

The data consists of several Swedish population-wide registers, linked by using unique individual identification numbers. The Swedish military enlistment data includes information on cognitive and non-cognitive abilities for all individuals in the sample, described in section 3.1. This register is linked with information on the year of death, mean yearly income at 31-35 years of age and education (from 1985 up till the year 1999 for education, and up till 2000 for income).[8] The income variable is inflation-adjusted with the year 2000 as base.

Individuals in the sample is born between year 1950 and 1965, and were enlisted between 1969 and 1983 in the year they turned 18-20.[9] Military service was mandatory only for men, therefore the small fraction of women who enlisted for military service are excluded from the data. With these restrictions the sample consists of 692,303 men with records from the military service. However, I do not have full information for income (missing 13,035 observations) and education (missing 8,943 observations). One reason is that about 16 percent of the deaths in the sample (4,562 observations) occurred before

[8] The choice of the 31-35 age bracket was guided by Böhlmark and Lindquist (2005), who found that for men this age bracket is a good proxy for lifetime income. Education is measured at age 30. If education is missing for that specific age, then the education level at age 29 or 31 is used. If information on education is missing for all these ages, the last record of education is used. For the cohorts born between 1950-54, the income variable for the oldest cohorts is a mean of the available years during the age bracket 31-35, and the education is measured at 31-36 or the last record of education.

[9] Some individuals were older than 20 years at enlistment. These individuals were excluded due to the possible unobserved factors affecting the timing.

1985, the first year of the demographic variables.[10] In total the dataset includes 28,570 deaths, about 4 percent of the individuals in the sample.[11]

The data on year of death ranges between the years 1969 and 2009, implying that the oldest individuals in the data, born 1950, is at most 59 years old when censored, and the youngest individuals, born 1965, at most 44 years old. The focus in this paper is all-cause premature mortality, which is used as a proxy for health. Data on cause of death is not available. However, the five most common causes of death for men aged 20-59 between the years 1969-2006 in Sweden are, in order: ischemic heart disease; suicide; malignant tumor; "other" accidents; traffic accidents.

3.1 Enlistment data

During this time period military enlistment was mandatory for all men in Sweden, with exemptions only for institutionalized individuals, prisoners, individuals living abroad and individuals with a severe medical condition or disability.[12] Otherwise, practically all men between 18-20 years old were enlisted. Individuals who refused to enlist were punished with a fine or, eventually, imprisonment. Almost 72 percent of the sample enlisted in the year they turned 18, and about 25 percent in the year they turned 19.

Enlistment usually spanned over two days and involved tests of the individual's health status, physical fitness, cognitive ability and non-cognitive ability. There was no incentive to underperform since it was not possible to avoid military service by scoring low on these tests.

The Swedish military has conducted tests of cognitive and non-cognitive abilities since the mid-1940s to help determine the military service of the enlisted. Cognitive ability was measured by a non-standard IQ test, aiming at measuring the g factor.[13] The test consisted by four sub-tests, representing logical, spatial, verbal and technical comprehension. The result at each subtest was standardized to give a score between 1 and 9, a so called Stanine distribution.[14] The sum of these four Stanine scores (ranging from 4-36) was,

[10] This likely results in a downward bias of the estimates for cognitive and non-cognitive abilities when income and education are included. Individuals who died before 1985 have, on average, a lower cognitive and non-cognitive ability (4.6 and 4.3 respectively) compared with individuals who died 1985 or later (5.2 and 5.1 respectively).

[11] Before exclusions, the number of observations in the data are 724,748 individuals, so I use about 94 percent of the total number of observations. In the full dataset there were 32,255 deaths, so I use about 87 percent of all deaths.

[12] This could lead to biased estimates, but since almost everyone enlisted the bias should be small.

[13] Carlstedt (2000) provides evidence that the cognitive ability test is a good measure of "general intelligence", in contrast with the US military Armed Forces Qualification Test, AFQT, which focuses on "crystallized" intelligence (Lindqvist and Vestman 2011).

[14] A Stanine ("STAndard NINE") distribution is calculated such that the mean value is 5 and the standard deviation is 2, with 1 as the lowest value and 9 as the highest value. 20 percent

Table 1. *Cognitive ability score and IQ*

Stanine	1	2	3	4	5	6	7	8	9
IQ	<74	74-81	82-89	90-95	96-104	105-110	111-118	119-126	>126

Notes: Stanine score and corresponding IQ with a mean of 100 and standard deviation of 15 (David et al. 1997).

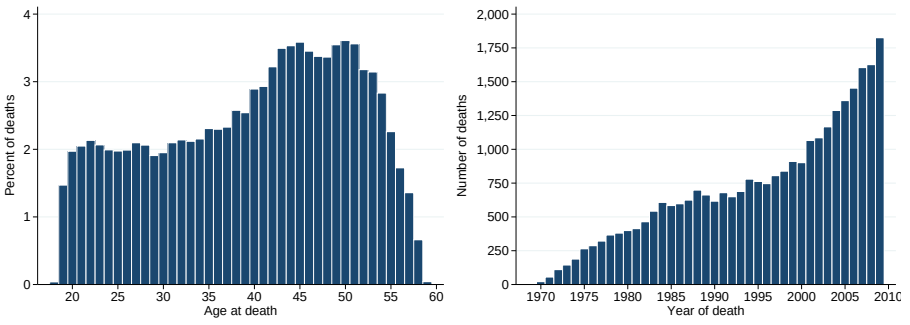
in turn, standardized into a Stanine variable of cognitive ability. Each Stanine score represents a range in IQ in accordance with Table 1.[15] As described by Batty, Wennerstad, et al. (2007) and others, the logical test measured how well the individual could understand written instructions and apply them to solve problems. In the spatial test the task was to identify the correct 2D plan drawing from a series of drawings of fully assembled 3D objects. The verbal test measured the individual's knowledge of synonyms. The individual was given a word and four alternatives of synonyms, and the task was to choose the correct synonym. Lastly, the technical abilities test measured the individual's knowledge of physics and chemistry. This test can be considered as a measure of general knowledge.

The non-cognitive ability was measured according to a procedure which remained unchanged during the sample period (Lindqvist and Vestman 2011). The conscripts were interviewed by a certified psychologist for about 25 minutes. The interviewer had information on the results at the cognitive ability test, physical fitness test, the grades in school and answers to about 80 questions about friends and family etc. that the individual had answered before the interview. The interview followed semi-structured rules. The psychologist followed a manual that stated the topics to discuss during the meeting, but no question was specified beforehand. The objective of the interview and the non-cognitive measure was to capture the general ability rather than a specific personal trait. The psychologist had to evaluate the individual's capability to function and fulfill the requirements in a demanding environment, i.e., military duty and armed combat. Motivation for doing military service was not judged. A high score was given if the individual was considered to be emotionally stable, willing to assume responsibility (in general), able to cope with stress, and take initiatives etc. (Grönqvist, Öckert, et al. 2010).[16] The final Stanine score of non-cognitive ability was determined, partly, by four different sub-scores which ranged from 1 to 5. These sub-scores only functioned as a guide for the psychologist; two individuals with exactly the same scores could receive

of the distribution is centered at 5, and 4 percent at 1 or 9 respectively. Each interval has a 0.5 standard deviation width except the first and the last, which contains the remainder of the distribution.

[15] Note that the Stanine scores and IQ scores are not exactly comparable, since the Stanine scores represents the generalized intelligence.

[16] As shown by Lindqvist and Vestman (2011), individuals who score high in this measure are more likely to succeed in the labor market.



(a) Age at death (b) Year of death

Figure 2. Mortality.

different final scores. The details of how the final assessment was done are classified.

3.2 Descriptive statistics

Figure 2 show age at death and the number of deaths each year, and descriptive statistics are presented in Table 2.

If we start by studying age at death in Figure 2a, we can see the age of death in the sample and the percentage each age represents of all deaths. The share of deaths is quite stable around age 20-35, but is then starting to increase. It is important to remember that the youngest cohort is censored at age 44. Not surprisingly, deaths are more common the later years, which can be seen in (b).

Mean age of enlistment in the sample is 18.3 years, as shown in Table 2. The cognitive and non-cognitive abilities are centered on a score of 5 with a standard deviation close to 2, as they follow the Stanine distribution (see footnote 14). Individuals with a mean labor income below the 1st quartile earn about 55,000 SEK a year on average, with a maximum of 130,000 SEK, while individuals above the 3rd quartile earn 301,000 SEK a year on average, and at least 232,000 SEK a year. About 4 percent of the sample (26,338 individuals) have no income, i.e., did not receive any income during their early 30's. 24 percent of the individuals have at most 9 years of education, and about the same percent have at least 13 years of education (i.e., college education).

Table 2. *Descriptive statistics*

	Observations	Mean	Standard dev.
Cognitive ability	692,303	5.18	1.95
Non-cognitive ability	692,303	5.08	1.79
Enlistment age	692,303	18.31	0.52
Deaths[1]	28,570	0.04	
Income	679,268	182.16	106.62
Low income (<Q1)	169,817	54.93	44.45
Middle income	339,634	186.11	26.19
High income (>Q3)	169,817	301.48	108.03
Years of education	683,360	11.68	2.50
At most 9 years of education[1]	169,455	0.24	
At most high school education[1]	344,256	0.50	
College education[1]	169,649	0.25	
N	692,303		

Notes: Data is missing on income and education for 13,035 and 8,943 individuals respectively. Income is measured as the mean yearly labor income at 31-35 years of age, inflation-adjusted with the year 2000 as base, in 1,000's SEK. Years of schooling is measured around age 30, and ranges from 7.1 to 19.9 years, where *At most high school education* is defined as at most 12 years of schooling.

[1] Mean in the full sample.

3.3 Study limitations

The measures of cognitive and non-cognitive ability that is used at the enlistment are only proxies for the true abilities, as pointed out by Grönqvist, Öckert, et al. (2010). They identify at least two potential sources of measurement errors. First, the evaluation instruments only test a subset of the ability, and individual ability can differ in these specific traits. For example, the subtest of synonyms for cognitive ability can of course only cover a few words. Second, individual ability can differ in the respect of taking tests (high/low motivation, illness or nervousness). The measurement error for non-cognitive ability is probably more severe than for cognitive ability, since the evaluation instruments for measuring cognitive ability are more developed.

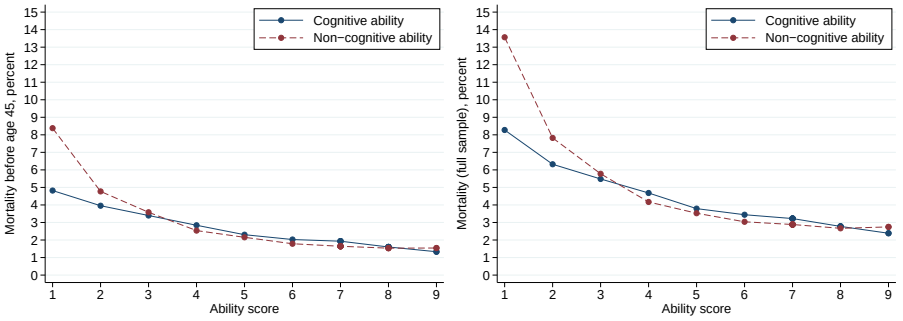
Another problem is that the ability measures possibly measure health in themselves. This is especially true for individuals with low non-cognitive ability, since that could be an indication of psychological ill-health, for example depression. However, an individual can have a low non-cognitive ability without any psychological health problems.[17] It should also be noted that the psychologist know the result on the cognitive ability test.

The relationships between the variables (Figure 1) are demanding. Cognitive and non-cognitive abilities are related to both income and education. Ideally, a control variable should be fixed when the independent variable of interest is determined. Therefore, in a regression analysis on mortality with cognitive and non-cognitive abilities as independent variables, inclusion of income and education as control variables may result in a bad controls problem (Angrist and Pischke 2009), as they themselves are outcomes. This may introduce selection bias and biased estimates. However, income and education are potentially channels for the association between the abilities and mortality, and to include them in the regressions can give us an indication if this indeed is the case. When interpreting the estimates one has to be aware of the potential bias, however. The consequence is likely a downward bias. For example, an individual with a high income but low ability is unusual, and is perhaps able to compensate the lack of ability with something unobserved in the data.

4 Results

I first study the relationships between the skills, mortality, income and education graphically, and then continue to a more formal statistical analysis later in this section. In the appendix additional results are presented.

[17] There is no information on, for example, psychiatric diagnoses in the data, but it should be noted that all individuals where healthy enough not to be exempted from the enlistment.



(a) Mortality before age 45 (b) Mortality in full sample

Figure 3. Associations between cognitive and non-cognitive ability and mortality.

4.1 Graphical analysis

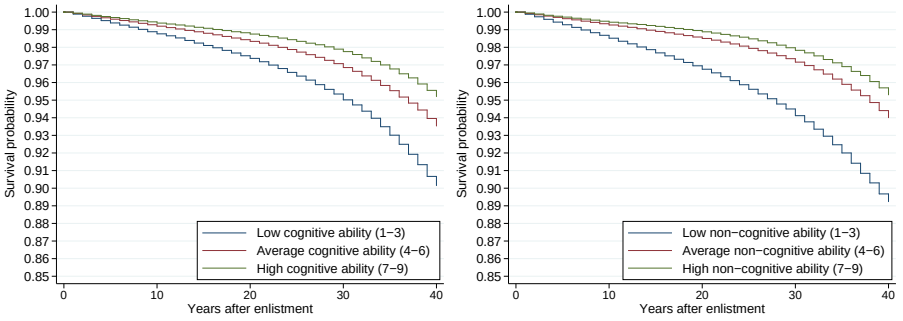
Let us turn to the graphical analysis.

Figure 3 presents the relationships between cognitive ability, non-cognitive ability and mortality. Up till age 45, about five percent of the individuals with a cognitive skill of 1 have died, compared to about one percent of those with the highest skill (a). This is a large difference in mortality rate, considering the relatively young age. However, the difference is even more pronounced if we look at non-cognitive ability. This is mainly driven by the bottom of the distribution, while the association for the cognitive ability is near-linear. Individuals with a very low non-cognitive ability have a much higher risk of mortality, and about 8 percent have died. The overall picture is the same in the full sample (b).

Figure 4 presents the survival probability each year after enlistment.[18] The skills are divided into three groups for the respective measure: low, average and high. First, we look at cognitive ability (a). At year 0, the probability of survival is 1. As time goes on the probability of survival shrinks. After a few years a distinct pattern emerges for each respective group. At the last period the probability of survival for the high ability individuals is above 0.95, but only about 0.90 for the low ability individuals. There is also a relatively large difference between the average skill individuals and the high skill individuals; the survival probability for the average skill individuals is below 0.94. The same pattern can be seen for non-cognitive ability (b). The low skill individuals have a survival probability of about 0.89 in the last period.

As this figure shows, there are not only differences in risk of mortality in relation to the abilities (cf. Figure 3); the differences grow over time. Low skill individuals have a lower survival probability than the average skill individuals, and the average skill individuals have a lower survival probability than the high skill individuals.

[18] Using the full sample, as Kaplan-Meier can handle censoring of the data.



(a) Cognitive ability (b) Non-cognitive ability

Figure 4. Kaplan-Meier survival curves for the relationships between cognitive and non-cognitive ability and mortality.

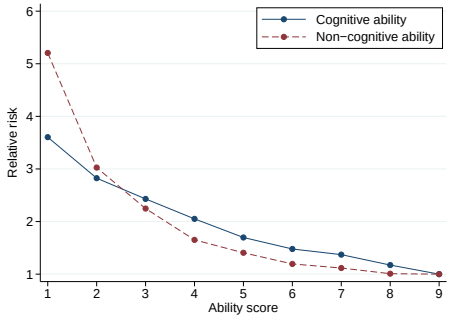
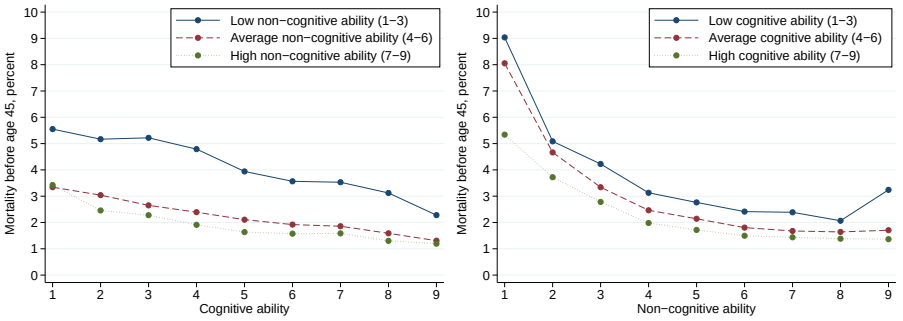


Figure 5. Relative risk.

The relative risk of mortality is shown in Figure 5. Using the highest score in the respective skill as reference, the figure plot the estimated the risk of mortality compared to the reference point. The patterns are much like Figure 3. The risk of mortality is much higher among individuals with lower scores. An individual with a cognitive skill of 1 has a risk of mortality about 3.6 times that of an individual with a score of 9, while the risk of mortality for individuals with a non-cognitive score of 1 is more than five times that of an individual with a score of 9. In comparison, the risk of mortality is considerably lower for an individual with a skill of 2, about three times that of an individual with a skill of 9.

In Figure 6, the skills are plotted against mortality before age of 45.[19] The individuals are divided into three groups in the *other* skill. Hence, in (a), individuals are grouped in non-cognitive ability while plotted at the respective cognitive ability score against mortality. In (b), the individuals are grouped

[19]This is to avoid the problem of right censoring. However, as seen in Figure 3, this does not change the pattern.



(a) Mortality and cognitive ability (b) Mortality and non-cognitive ability

Figure 6. Mortality by ability.

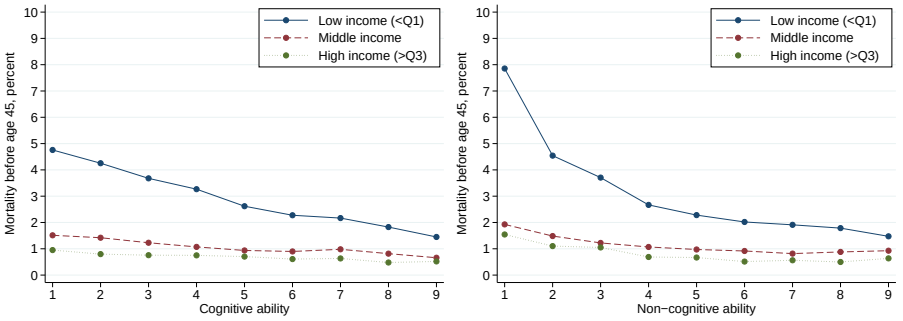
in cognitive ability and plotted at the respective non-cognitive ability score against mortality.

Earlier we have seen that there is an almost linear relationship between mortality and cognitive ability. The linear relationship is still apparent in (a), but there are large differences in levels between the groups. Individuals with a low non-cognitive ability have a much higher risk of premature mortality compared with individuals with at least an average score. The difference between individuals with an average or high non-cognitive ability is relatively small, however. With non-cognitive ability as base (b), the non-linear relationship seen earlier is clear. With exception for the individuals with the lowest and highest non-cognitive ability, the differences between the groups are more or less constant. Individuals with a low non-cognitive ability have a much higher risk of mortality than those with at least an average ability.

Figure 6 strengthens the conclusion that low non-cognitive ability is the more important predictor of premature mortality of the two abilities. The risk is considerably higher regardless of the score in cognitive ability. This suggests that it is not possible to fully compensate a low non-cognitive ability with high cognitive ability.

How are these associations affected if we take education and income into account? After all, we know that the abilities have strong relationships with these variables (e.g. Heckman, Stixrud, et al. 2006; Lindqvist and Vestman 2011).

We start by looking at the interaction with income (Figure 7). (a) presents the relationship between mortality and cognitive ability when the individuals are divided into income groups. There is still a linear relationship between cognitive ability and mortality, but we can see that it is mainly driven by individuals with low income. In the low income group, about 5 percent of those with a cognitive skill of 1 have died, compared to about 2 percent of the individuals with the highest skill. The middle and high income groups have much weaker associations. About 0.5-1.5 percent in these groups have died, regard-



(a) Mortality, income and cog. ability (b) Mortality, income and non-cog. ability

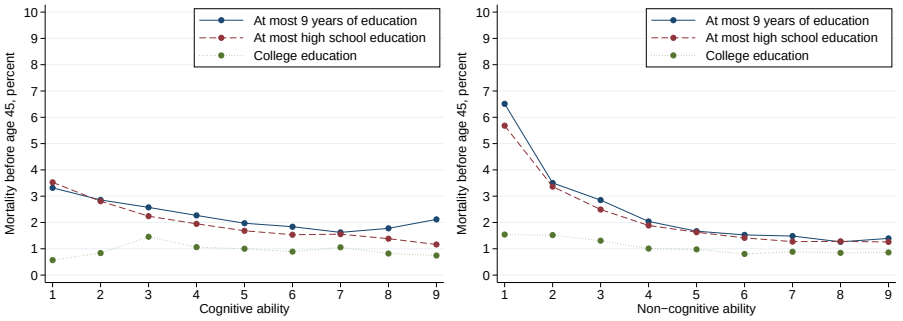
Figure 7. Mortality by income group.

less of skill. The difference in regards to income and mortality is even more pronounced if we look at non-cognitive ability, as shown in (b). Almost 8 percent of the individuals with low income and the lowest non-cognitive ability have died, but only 1.5 percent of the individuals with the highest ability. Again, the association are much weaker for the middle and the high income groups.

How is education interacted with skills and mortality? This is shown in Figure 8. (a) divides the sample into education level and show the relationship between cognitive ability and mortality. It is clear that the non-college educated groups follow the same trend, except for a rise in mortality for the low educated high ability individuals. The college educated group has a lower risk of mortality than the less educated groups, and show no association between the skill and mortality. There is a decline in mortality for the college educated low ability individuals.[20] The general picture is that there is a linear relationship between cognitive ability and mortality, but it seems that education plays the more important role. (b) presents the corresponding association between mortality and non-cognitive ability. There is a large difference between the college educated individuals and the non-college educated in the bottom of the distribution. Non-cognitive skill seems to be an important predictor of mortality for individuals without college education, but not for those with a college education. About 1 percent of the college educated individuals have died regardless of non-cognitive ability. In contrast, about 6-7 percent of the individuals in the lowest skill group without college have died.

The graphical analysis show, first, that there are strong associations between the skills and mortality. Second, that these associations are heterogeneous in regards to income and education. This suggest that using income and education as proxy variables for skill potentially miss a large within-group variation. For example, individuals with low education but high ability do not differ

[20]The somewhat surprising changes in the patterns described are likely driven by the relatively few number of observations in these groups.



(a) Mortality, education and cog. ability (b) Mortality, education and non-cog. ability

Figure 8. Mortality by education level.

much in regards of the risk of mortality from individuals with high school education.

4.2 Regression analysis

We now turn to the formal regression analysis, using Cox proportional hazard models in Table 3. OLS results are presented in the appendix.

The estimated hazard ratio for cognitive ability is 0.856 (i.e., a one point increase in cognitive ability results in a 14 percent lower risk of mortality). The corresponding result for non-cognitive ability is even stronger; a one point increase results in a 19 percent lower risk of mortality, indicating that both measures are important but that non-cognitive ability is a stronger predictor. These results support the graphical analysis earlier. In column 3, where both measures are included, cognitive ability is associated with a 9 percent lower risk of mortality, and non-cognitive about 16 percent. The attenuation is greater for cognitive ability, indicating that non-cognitive ability captures more variation in the data.

When income is included in the model (column 4), non-cognitive ability is more attenuated than cognitive ability, indicating that income and non-cognitive ability partly captures the same variation in data, and that income is a mediating factor mainly for non-cognitive ability (cf. Figure 1). The reverse is true when years of schooling is included instead of income (column 5); cognitive ability is attenuated and non-cognitive ability is stable, suggesting that education is a mediator for cognitive ability. The full model reveals that income and education have independent associations with mortality, but that non-cognitive ability is a stronger predictor of mortality than cognitive ability (about two times the strength).

The analyses conducted in this section reveals strong associations between cognitive and non-cognitive abilities and mortality. The analyses also suggest

Table 3. *Associations with mortality*

	(1)	(2)	(3)	(4)	(5)	(6)
Cog. ability	0.856***		0.913***	0.925***	0.946***	0.948***
	(0.003)		(0.003)	(0.003)	(0.004)	(0.004)
Non-cog. ability		0.806***	0.837***	0.865***	0.853***	0.870***
		(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Income				0.790***		0.797***
				(0.003)		(0.003)
Years of schooling					0.927***	0.952***
					(0.003)	(0.003)
Observations	692,235	692,235	692,235	679,268	683,360	677,983

Notes: Hazard ratios. Mortality as event. Log of mean inflation-adjusted yearly income between the age of 31-35. To adjust for zero income individuals, the variable is ln(Income+1). Years of schooling around age 30. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

that income and education are mediators; cognitive ability mainly through education, and non-cognitive ability mainly through income.

5 Discussion

This paper aimed to answer two questions. First, is there a relationship between non-cognitive ability and mortality, and how is that relationship compared with the relationship between cognitive ability and mortality? Second, are income and education good proxies for individual ability?

What can we say about the first question? While the abilities do not have a direct causal link with mortality, it seems that they create possibilities to live a better, or at least longer, life. First, cognitive and non-cognitive skills are both negatively related with mortality. Individuals with high cognitive and non-cognitive ability live, on average, longer. However, non-cognitive ability is a stronger predictor of mortality than cognitive ability. The association between mortality and non-cognitive ability is more than two times the association between cognitive ability and mortality. The results suggest that education is a mediator for cognitive ability, while non-cognitive ability is associated with mortality through income in the same way. The relationships are not linear. The graphical analysis show that the relationships are driven mainly by the bottom of the distribution, especially in respect to the non-cognitive ability. Possibly, this is because low non-cognitive ability can be a measure of psychological ill-health.[21]

[21] As mentioned earlier, however, it is possible to have low non-cognitive ability without any psychological health problems, and the individuals were at least healthy enough not to be exempted from the enlistment.

The relative strengths of the associations for cognitive and non-cognitive abilities are in line with what Heckman, Stixrud, et al. (2006) finds for labor market outcomes; namely that non-cognitive ability is generally more or at least as important as cognitive ability. As shown, this seems to be true also for mortality. The earlier epidemiological literature, which only looked at the relationship between cognitive ability and mortality, lacks the important dimension of non-cognitive skills.

How good are income and education as proxies for skill? The results suggest that cognitive and non-cognitive abilities are important for individuals with low income or without a college education. Thus, if we use income and education as proxies, we will miss a large within-group variation. Individuals with low income but high ability do not differ much from individuals with a higher income regarding the risk of mortality. This variation would be hidden and possibly lead to wrong conclusions. Not everyone with low income or below college education have a higher risk of mortality, but that is what we would conclude using these variables. On average, however, they are not misleading; lower income or education are associated with a higher risk of mortality (see the appendix and earlier references).

With the data at hand it is not possible to answer why non-cognitive ability is the stronger predictor of the risk of mortality, or why cognitive and non-cognitive abilities are important for low income and non-college educated individuals. It could be that premature mortality is more about avoiding "failure" (die) rather than achieving "success" (live on). In Sweden, relatively few individuals die prematurely, and it requires special conditions. These conditions are not common among individuals who are college educated, have a "decent" income or is "socially functional". Individuals who lack social skills, on the other hand, may have a harder time. This suggests that policies improving social skills, especially early in life and for individuals unlikely to continue to college, may be beneficial not only for the individual, but for the public health.

6 Conclusions

Using a dataset from the Swedish military enlistment of 692,303 men born between 1950 and 1963, I have in this paper shown that cognitive and non-cognitive abilities are associated with mortality. The results suggest that non-cognitive ability is a stronger predictor than cognitive ability, indicating that the literature on the relationship between cognitive ability and mortality have lacked an important dimension. The associations remain when controlling for income and education for low income and non-college educated individuals. Using income and education as proxy measures for skill therefore miss the large variation within these groups. The associations are mainly driven by the bottom of the distributions.

References

Acemoglu, Daron, and Joshua D Angrist. 2001. "How large are human-capital externalities? Evidence from compulsory schooling laws." In *NBER Macroeconomics Annual 2000,* edited by Ben S Bernanke and Kenneth Rogoff, 15:9–59. Cambridge and London: MIT Press.

Albouy, Valerie, and Laurent Lequien. 2009. "Does compulsory education lower mortality?" *Journal of Health Economics* 28 (1): 155–168.

Andersson, Ola, Jean-Robert Tyran, Erik Wengström, and Håkan Holm. 2013. *Risk aversion relates to cognitive ability: Fact or fiction?* Working Paper 964. Research Institute of Industrial Economics (IFN).

Angrist, Joshua D, and Alan B Krueger. 1991. "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics* 106 (4): 979–1014.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, New Jersey: Princeton University Press. ISBN: 9780691120348.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2015. *Non-cognitive deficits and young adult outcomes: The long-run impacts of a universal child care program.* Working Paper 21571. National Bureau of Economic Research.

Basu, Anirban, Andrew M Jones, and Pedro Rosa Dias. 2014. *The roles of cognitive and non-cognitive skills in moderating the effects of mixed-ability schools on long-term health.* Working Paper 20811. National Bureau of Economic Research.

Batty, G David, Ian J Deary, and Linda S Gottfredson. 2007. "Premorbid (early life) IQ and later mortality risk: Systematic review." *Annals of Epidemiology* 17 (4): 278–288.

Batty, G David, Catharine R Gale, Per Tynelius, Ian J Deary, and Finn Rasmussen. 2009. "IQ in early adulthood, socioeconomic position, and unintentional injury mortality by middle age: A cohort study of more than 1 million swedish men." *American Journal of Epidemiology* 169 (5): 606–615.

Batty, G David, Karin Modig Wennerstad, George Davey Smith, David Gunnell, Ian J Deary, Per Tynelius, and Finn Rasmussen. 2007. "IQ in early adulthood and later cancer risk: Cohort study of one million Swedish men." *Annals of Oncology* 18 (1): 21–28.

———. 2009. "IQ in early adulthood and mortality by middle age: Cohort study of 1 million swedish men." *Epidemiology* 20 (1): 100–109.

Benjamin, Daniel J, Sebastian A Brown, and Jesse M Shapiro. 2013. "Who is 'behavioral'? Cognitive ability and anomalous preferences." *Journal of the European Economic Association* 11 (6): 1231–1255.

Bharadwaj, Prashant, Katrine Vellesen Løken, and Christopher Neilson. 2013. "Early life health interventions and academic achievement." *The American Economic Review* 103 (5): 1862–1891.

Böhlmark, Anders, and Matthew J Lindquist. 2005. *Life-cycle variations in the association between current and lifetime income: Country, cohort and gender comparisons.* Working Paper 4. Swedish Institute for Social Research (SOFI).

Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "The determinants of earnings: A behavioral approach." *Journal of Economic Literature* 39 (4): 1137–1176.

Buckles, Kasey, Andreas Hagemann, Ofer Malamud, Melinda S Morrill, and Abigail K Wozniak. 2013. *The effect of college education on health.* Working Paper 19222. National Bureau of Economic Research (NBER).

Burks, Stephen V, Jeffrey P Carpenter, Lorenz Goette, and Aldo Rustichini. 2009. "Cognitive skills affect economic preferences, strategic behavior, and job attachment." *Proceedings of the National Academy of Sciences* 106 (19): 7745–7750.

Calvin, Catherine M, Ian J Deary, Candida Fenton, Beverly A Roberts, Geoff Der, Nicola Leckenby, and G David Batty. 2011. "Intelligence in youth and all-cause-mortality: Systematic review with meta-analysis." *International Journal of Epidemiology* 40 (3): 626–644.

Card, David. 1999. "The causal effect of education on earnings." In *Handbook of Labor Economics,* edited by Orley Ashenfelter and David Card, 3:1801–1863. Amsterdam, North Holland: Elsevier.

Carlstedt, Berit. 2000. "Cognitive abilities-aspects of structure, process and measurement." PhD diss., University of Gothenburg.

Cesarini, David, Erik Lindqvist, Robert Östling, and Björn Wallace. 2016. "Wealth, health, and child development: Evidence from administrative data on Swedish lottery players." *The Quarterly Journal of Economics* 131 (2): 687–738.

Chiteji, Ngina. 2010. "Time-preference, non-cognitive skills and well-being across the life course: Do non-cognitive skills encourage healthy behavior?" *The American Economic Review* 100 (2): 200–204.

Clark, Damon, and Heather Royer. 2013. "The effect of education on adult mortality and health: Evidence from Britain." *The American Economic Review* 103 (6): 2087–2120.

Conti, Gabriella, and Christopher Hansman. 2013. "Personality and the education-health gradient: A note on "Understanding differences in health behaviors by education"." *Journal of Health Economics* 32 (2): 480–485.

Conti, Gabriella, James J Heckman, and Rodrigo Pinto. 2015. *The effects of two influential early childhood interventions on health and healthy behaviors.* Working Paper 21454. National Bureau of Economic Research.

Conti, Gabriella, James J Heckman, and Sergio Urzua. 2010. "The education-health gradient." *The American Economic Review* 100 (2): 234–238.

Contoyannis, Paul, and Nigel Rice. 2001. "The impact of health on wages: Evidence from the British household panel survey." *Empirical Economics* 26 (4): 599–622.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78 (3): 883–931.

Currie, Janet. 2009. "Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development." *Journal of Economic Literature* 47 (1): 87–122.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney. 2006. "The determinants of mortality." *Journal of Economic Perspectives* 20 (3): 97–120.

David, Anthony S, A Malmberg, L Brandt, P Allebeck, and G Lewis. 1997. "IQ and risk for schizophrenia: A population-based cohort study." *Psychological Medicine* 27 (06): 1311–1323.

Deary, Ian J, and G David Batty. 2007. "Cognitive epidemiology." *Journal of Epidemiology and Community Health* 61 (5): 378–384.

Deaton, Angus. 2003. "Health, inequality, and economic development." *Journal of Economic Literature* 41 (1): 113–158.

Ding, Weili, Steven F Lehrer, J Niels Rosenquist, and Janet Audrain-McGovern. 2009. "The impact of poor health on academic performance: New evidence using genetic markers." *Journal of Health Economics* 28 (3): 578–597.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. "Are risk aversion and impatience related to cognitive ability?" *The American Economic Review* 100 (3): 1238–1260.

Fischer, Martin, Martin Karlsson, and Therese Nilsson. 2013. "Effects of compulsory schooling on mortality: Evidence from Sweden." *International Journal of Environmental Research and Public Health* 10 (8): 3596–3618.

Frijters, Paul, John P Haisken-DeNew, and Michael A Shields. 2005. "The causal effect of income on health: Evidence from German reunification." *Journal of Health Economics* 24 (5): 997–1017.

Gomez-Pinilla, Fernando. 2008. "The influences of diet and exercise on mental health through hormesis." *Ageing Research Reviews* 7 (1): 49–62.

Grönqvist, Erik, Björn Öckert, and Jonas Vlachos. 2010. *The intergenerational transmission of cognitive and non-cognitive abilities.* Working Paper 2010:12. Institute for Labour Market Policy Evaluation (IFAU).

Grönqvist, Erik, and Jonas Vlachos. 2008. *One size fits all? The effects of teacher cognitive and non-cognitive abilities on student achievement.* Working Paper 2008:25. Institute for Labour Market Policy Evaluation (IFAU).

Grossman, Michael. 1972. "On the concept of health capital and the demand for health." *Journal of Political Economy* 80 (2): 223–255.

Heckman, James J, John Eric Humphries, Greg Veramendi, and Sergio Urzua. 2014. *Education, health and wages.* Working Paper 19971. National Bureau of Economic Research.

Heckman, James J, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *The American Economic Review* 103 (6): 2052–2086.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics* 24 (3): 411–482.

Hemmingsson, Tomas, Jan von Essen, Bo Melin, Peter Allebeck, and Ingvar Lundberg. 2007. "The association between cognitive ability measured at ages 18-20 and coronary heart disease in middle age among men: A prospective study using the Swedish 1969 conscription cohort." *Social Science & Medicine* 65 (7): 1410–1419.

Hemmingsson, Tomas, Bo Melin, Peter Allebeck, and Ingvar Lundberg. 2006. "The association between cognitive ability measured at ages 18-20 and mortality during 30 years of follow-up – a prospective observational study among Swedish males born 1949–51." *International Journal of Epidemiology* 35 (3): 665–670.

———. 2009. "Cognitive ability in adolescence and mortality in middle age: A prospective life course study." *Journal of Epidemiology & Community Health* 63 (9): 697–702.

Kautz, Tim, James J Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. 2014. *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success.* Working Paper 20749. National Bureau of Economic Research.

Kippersluis, Hans van, Owen O'Donnell, and Eddy van Doorslaer. 2011. "Long-run returns to education does schooling lead to an extended old age?" *Journal of Human Resources* 46 (4): 695–721.

Lager, Anton, Sven Bremberg, and Denny Vågerö. 2009. "The association of early IQ and education with mortality: 65 year longitudinal study in Malmö, Sweden." *British Medical Journal* 339 (b5282).

Lindahl, Mikael. 2005. "Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income." *Journal of Human Resources* 40 (1): 144–168.

Lindqvist, Erik, and Roine Vestman. 2011. "The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment." *American Economic Journal: Applied Economics* 3 (1): 101–128.

Lleras-Muney, Adriana. 2005. "The relationship between education and adult mortality in the United States." *The Review of Economic Studies* 72 (1): 189–221.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. "Poverty impedes cognitive function." *Science* 341 (6149): 976–980.

Marmot, Michael. 2002. "The influence of income on health: Views of an epidemiologist." *Health Affairs* 21 (2): 31–46.

Nyhus, Ellen K, and Empar Pons. 2005. "The effects of personality on earnings." *Journal of Economic Psychology* 26 (3): 363–384.

Roberts, Brent W, Nathan R Kuncel, Rebecca Shiner, Avshalom Caspi, and Lewis R Goldberg. 2007. "The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes." *Perspectives on Psychological Science* 2 (4): 313–345.

Savelyev, Peter A. 2014. *Socioemotional skills, education, and longevity of high-ability individuals.* Working Paper. Vanderbilt University, Department of Economics.

Savelyev, Peter A, and Kegon Teng Kok Tan. 2014. *Personality, education, and health-related outcomes of high-ability individuals.* Working Paper. Vanderbilt University, Department of Economics.

Sørensen, Henrik Toft, Svend Sabroe, Jørn Olsen, Kenneth J Rothman, Matthew W Gillman, and Peer Fischer. 1997. "Birth weight and cognitive function in young adult life: Historical cohort study." *British Medical Journal* 315 (7105): 401–403.

Vogl, Tom, David M Cutler, and Adriana Lleras-Muney. 2011. "Socioeconomic status and health: Dimensions and mechanisms." In *The Oxford Handbook of Health Economics,* edited by Sherry Glied and Peter C Smith, 124–163. Oxford: Oxford University Press.

Appendix

A OLS results

Table A1 presents OLS regression estimates corresponding to the Cox proportional hazard models in Table 3. To handle right censoring of the data the mortality outcome variable is defined as death before age 45.

The results are in line with the results presented in the main section. Non-cognitive ability is a stronger predictor of mortality, and years of schooling is mainly a mediator for cognitive ability and income a mediator for non-cognitive ability.

B Additional figures

In this section I present additional figures of the relationships between cognitive and non-cognitive ability, income, education and mortality.

To begin with, we look at the association between cognitive and non-cognitive abilities in Figure A1a. The correlation is 0.38, and the relationship is almost linear. This means that an individual with a high cognitive ability also, on average, have a high non-cognitive ability. Next, I look at the relationship between income and education (b). Up till about 17 years of schooling there is a strong and linear positive relationship between education and income. Individuals with at least 17 years of schooling earns, on average, about 100,000 SEK more per year than individuals with only about 7 years.

We now turn to the relationships between income, education and mortality. The relationship between mortality and income is non-linear below the 50th percentile (c). More than 4 percent of the individuals below the 10th percentile have died, but only about 0.5 percent of the individuals above the 90th percentile. Similar differences can be seen for the relation between education and mortality (d). About 2.5 percent of the individuals with at most 9 years of education have died, but less than 1 percent of the individuals with at least a few years in college. Both income and education are important predictors of mortality, but it is mainly the bottom of the distributions that are driving these relationships.

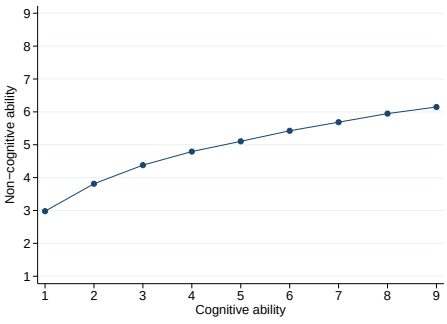
The next question is how income and education are related to the abilities (Figure A2).[22] There are strong and more or less linear relationships between

[22] The relationships between skills, income and education have been studied earlier, for example in Heckman, Stixrud, et al. (2006) and Lindqvist and Vestman (2011).

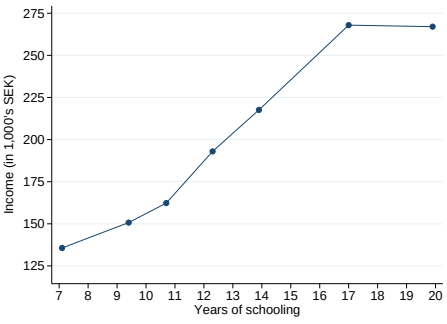
Table A1. *Associations with mortality*

	(1)	(2)	(3)	(4)	(5)	(6)
Cog. ability	-0.004***		-0.002***	-0.001***	-0.001***	-0.001***
	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
Non-cog. ability		-0.006***	-0.005***	-0.002***	-0.003***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Income				-0.007***		-0.007***
				(0.000)		(0.000)
Years of schooling					-0.001***	-0.000
					(0.000)	(0.000)
Birth cohort	Yes	Yes	Yes	Yes	Yes	Yes
Observations	692,303	692,303	692,303	679,268	683,360	677,983

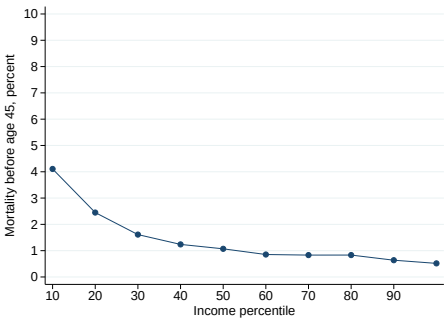
Notes: OLS. Death before age 45 as the dependent variable. Log of mean inflation-adjusted yearly income between the age of 31-35. To adjust for zero income individuals, the variable is ln(Income+1). Years of schooling around age 30. Robust standard errors in parentheses. *** p $<$ 0.01, ** p $<$ 0.05, * p $<$ 0.1.



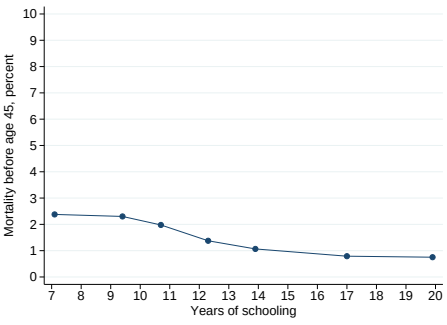
(a) Cognitive and non-cognitive ability



(b) Income and education

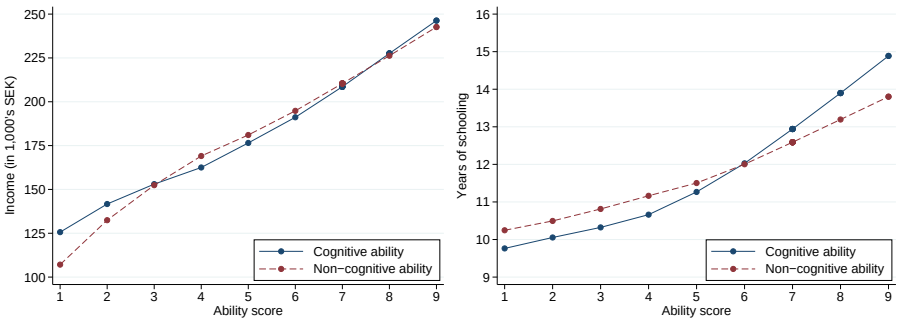


(c) Mortality and income



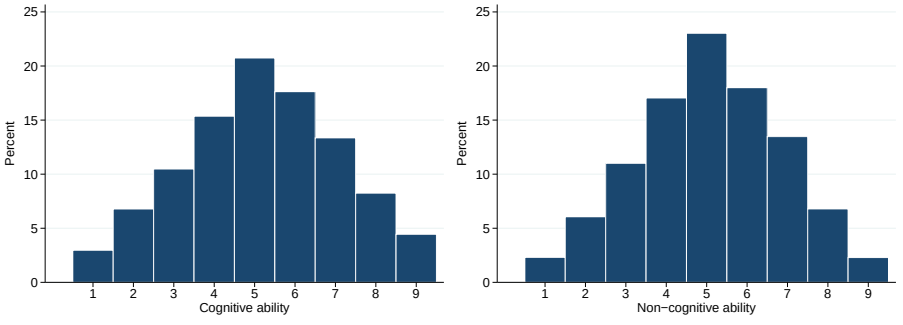
(d) Mortality and education

Figure A1. Associations.



(a) Income and abilities (b) Education and abilities

Figure A2. Associations between income, education and cognitive and non-cognitive ability.



(a) Cognitive ability (b) Non-cognitive ability

Figure A3. Distribution of abilities.

income and the abilities, as shown in (a). Low skill individuals earn, on average, much less than high skill individuals. The relationships between abilities and education are presented in (b). As with income, the relationships are positive and strong. Noticeably, above an average skill, cognitive ability becomes a stronger predictor of education than non-cognitive ability. Individuals with a score of 5 have, on average, a high school education, while individuals with higher skills continued to college.

Finally, there is possibly observations missing in the enlistment data, which would skew the Stanine distributions of the skills. As can be seen in Figure A3, both measures follow the expected distribution relatively closely, which suggests that there is no important missing data problem.

C Number of observations

The tables in this section presents the corresponding number of observations for Figure 6, Figure 7, and Figure 8.

Table A2. *Observations by cognitive ability*

	Non-cognitive ability		
Cog.	Low	Middle	High
1	13,745	6,493	292
2	20,919	23,555	2,483
3	22,185	42,844	7,507
4	23,322	66,314	16,686
5	22,684	92,485	28,432
6	14,891	74,461	32,690
7	9,151	52,431	30,893
8	4,836	29,205	23,043
9	2,413	14,210	14,133
N	134,146	401,998	156,159

Notes: Group frequency of individuals corresponding to Figure 6a.

Table A3. *Observations by non-cognitive ability*

	Cognitive ability		
Non-cog.	Low	Middle	High
1	8,863	5,836	1,292
2	20,175	17,580	4,219
3	27,811	37,481	10,889
4	29,459	66,040	22,511
5	28,190	94,761	36,447
6	15,243	72,459	36,888
7	7,624	50,018	35,639
8	2,226	21,870	22,912
9	432	5,920	9,518
N	140,023	371,965	180,315

Notes: Group frequency of individuals corresponding to Figure 6b.

Table A4. *Observations by cognitive ability*

	Income group		
Cog.	Low	Middle	High
1	8,681	10,446	946
2	16,480	25,831	3,761
3	22,184	40,815	8,198
4	29,542	58,848	16,129
5	35,472	74,629	31,138
6	26,156	59,208	34,494
7	17,345	39,766	33,499
8	9,474	20,848	25,481
9	4,483	9,243	16,171
N	169,817	339,634	169,817

Notes: Group frequency of individuals corresponding to Figure 7a.

Table A5. *Observations by non-cognitive ability*

	Income group		
Non-cog.	Low	Middle	High
1	8,632	5,967	777
2	17,458	20,003	3,447
3	24,521	40,561	9,516
4	30,915	63,774	21,349
5	36,169	84,025	36,678
6	24,849	60,984	36,748
7	17,161	40,640	33,741
8	7,670	18,398	19,847
9	2,442	5,282	7,714
N	169,817	339,634	169,817

Notes: Group frequency of individuals corresponding to Figure 7b.

Table A6. *Observations by cognitive ability*

	Education group		
Cog.	Low	Middle	High
1	11,724	8,195	176
2	22,472	22,718	955
3	29,482	38,747	3,223
4	35,523	60,369	9,030
5	34,391	83,381	24,235
6	20,993	64,143	35,502
7	10,220	40,503	40,679
8	3,658	19,067	33,676
9	992	7,133	22,173
N	169,455	344,256	169,649

Notes: Group frequency of individuals corresponding to Figure 8a.

Table A7. *Observations by non-cognitive ability*

	Education group		
Non-cog.	Low	Middle	High
1	7,708	6,794	908
2	18,124	19,558	3,356
3	28,028	38,022	8,868
4	34,723	62,765	19,125
5	37,551	86,194	33,963
6	23,498	63,285	36,569
7	13,545	43,040	35,710
8	5,059	19,046	22,318
9	1,219	5,552	8,832
N	169,455	344,256	169,649

Notes: Group frequency of individuals corresponding to Figure 8b.

III. Health Information and Well-Being: Evidence from an Asymptomatic Disease

Co-authored with Matz Dahlberg, Kevin Mani and Anders Wanhainen

Acknowledgments: We thank Lars Johansson, Damon Clark, Linda Lyttkens, Linuz Aggeborn, Pernilla Jernerén Maathz and Erik Grönqvist, as well as participants at the HEFUU seminar at Uppsala University and at the SHEA conference in Lund for helpful discussions, comments and suggestions. Mattias Öhman gratefully acknowledges financial support from U-CARE.

1 Introduction

It is well-known that information may affect individual behavior. For example, it has been shown that information to smoking, pregnant women about the dangers of smoking and on how to quit smoking significantly affects their children's weight and height at birth (Sexton and Hebel 1984). Likewise, some of the information experiments conducted by the Minnesota Department of Revenue in 1995 (the Minnesota Income Tax Compliance Experiment), such as information to tax payers about increased examination and auditing of tax returns, had a significant impact on reported income and taxes paid (see Slemrod et al. (2001) and the references cited therein). Zhao et al. (2013) show that individuals adjust behavior after receiving negative health information about hypertension.[1]

However, information might not only affect behavior, it might also affect individuals' well-being. This might of course be true for many types of (positive or negative) information, but it is not the least true in connection with medical examination and medical treatment where new information about one's own ill-health is often revealed. The information in itself might cause extra suffering (i.e. in addition to the suffering from the actual illness). So far we know very little of this extra patient suffering, both to what extent it exists and, if it exists, what the magnitude of it is.

The purpose of this paper is to examine how new health information affects individuals' physical and psychological well-being (or mental health).

The main methodological problem in answering this question is to ensure the exogeneity of the information of ill-health from the doctor to the patient. Symptoms of an illness might affect an individual's well-being already before a doctor inform the individual about the illness. To get around this problem, we will use data from a screening program for abdominal aortic aneurysm (AAA). Screening for AAA is well suited for answering the question at hand. First, AAA is well described as an asymptomatic disease, meaning that an individual do not know about it and, hence, the illness cannot affect the individual's well-being before information about it is provided by the doctor after examination. Second, based on certain pre-determined cut-offs, given by the measured size of the aneurysm, individuals are given different information about their health. More specifically, we will compare differential information at two cut-offs, 25 mm and 30 mm. At 25 mm, we will compare those who receive the information that they are healthy (below 25 mm) with those that receive the information that they have an enlarged aorta size and are in the risk zone for AAA (above 25 mm but below 30 mm). At 30 mm, we will compare those that receive the information that they have a small AAA, and therefore will be under increased surveillance by the health care system (between 30-34 mm), with those who receive the information provided between 25 and 29 mm,

[1] Other studies that show that information affects individual behavior include Strömberg (2004) and Engström et al. (2007).

i.e., enlarged aorta and increased risk for AAA, but only sparse surveillance by the health care system. Comparing individuals on either side of these cut-offs, we can use a regression discontinuity (RD) design to estimate the causal effect of new health information about ill-health on the individuals' well-being.

The literature on the psychological consequences of screening for AAA is limited and somewhat inconsistent (Lucarotti et al. 1997; The UK Small Aneurysm Trial Participants 1998; Lindholt, Vammen, et al. 2000; Scott and Group 2002; Spencer et al. 2004; Hansson et al. 2012). These papers do however all suffer from the same type of methodological problem (or use qualitative methods); since they are only comparing group averages, typically groups of individuals on either side of a cut-off for an aneurysm, they do not control properly for observed and unobserved confounders such as smoking history and/or an unhealthy lifestyle in general. By using the RD design, we are able to handle this methodological problem.

We use data from the Swedish screening program for AAA in Uppsala County.[2] The individuals' physical and psychological well-being is measured via the SF-36 and EQ-5D questionnaires as well as an AAA specific questionnaire, questionnaires that the patients fill out a couple of weeks after the information about their health has been revealed. We also have information about smoking and BMI. These are measured at the time of the screening.

We find no statistically significant effects at the 25 mm cut-off, but most outcomes have negative estimates. At the 30 mm cut-off, however, we find weak evidence of a positive effect of the information provided to the patients on the patients' psychological well-being. This result indicates that the information about increased surveillance (positive information) may outweigh the information about worse health (negative information).[3]

The rest of the paper is organized as follows: In the next section, we discuss AAA, the set-up of the screening program and the information provided within the program. In section 3, we provide a theoretical framework to organize the thinking on the question at hand. After describing the data (section 4) and setting up the econometric model (section 5), section 6 provide the baseline results and section 7 the sensitivity analyses. Section 8 discusses the results and section 9 concludes.

[2] In Sweden, there are two local governments, one at the municipal level and one at the county level. There are 290 municipalities and 21 counties, implying that each county constitute a fairly large geographical area. The county councils are mainly responsible for the health care system.

[3] Increased surveillance could possibly be seen as negative information, as the individual might interpret this as something really serious. However, we argue that, *ceteris paribus*, information about increased surveillance is positive.

2 Background

In this section we will briefly discuss AAA, the set-up of the screening program and the information provided within the program.

2.1 AAA

AAA is a common disease with potentially life-threatening consequences. Most AAAs are asymptomatic until rupture. The best therapy for AAA is pre-symptomatic elective surgical repair in appropriately selected individuals. However, most AAAs are undiagnosed and the large majority of patients with a ruptured AAA die if they do not receive immediate surgery.

There is no agreement on how to exactly define an AAA (Wanhainen 2008). The normal abdominal aortic diameter in elderly men varies between 15 mm and 24 mm (Sakalihasan et al. 2005). The most accepted definition of an AAA in clinical practice is a maximum infrarenal aortic diameter of at least 30 mm (McGregor et al. 1975; Moll et al. 2011).

In developed countries, AAA cause 1-3 percent of all deaths among men aged 65 to 85 years (Sakalihasan et al. 2005). It is estimated that about 600 men and 200 women die as a result of a ruptured AAA each year in Sweden (SBU 2008). Besides male sex and age, the most important risk factor for AAA is smoking; the prevalence of AAA among individuals with a history of smoking is more than four times that in non-smokers (Wanhainen, Bergqvist, et al. 2005; Sakalihasan et al. 2005; Svensjö et al. 2011). To reduce the high mortality, early detection by screening has been advocated.[4]

2.2 Screening for AAA in Sweden

Following the introduction of a general AAA screening program for 65-year-old men in the County of Uppsala in 2006, other counties in Sweden launched similar programs (Wanhainen and Björck 2011). Today all counties in Sweden have implemented an AAA screening program for 65-year-old men, which is also recommended by The National Board of Health and Welfare. Thereby, Sweden is the first country with a nationwide coverage. In this paper we use data from the screening program in Uppsala County.

Most counties have adopted a centralized hospital-based screening program to which all 65-year-old men, identified through the National Population Registry, are invited. The attendance rate is high, about 85 percent of the invited

[4] Several studies have demonstrated that screening for AAA cost-effectively reduce the AAA death rate by more than 50 percent (Ashton et al. 2007; Cosford et al. 2007; SBU 2008; Lindholt and Norman 2008; Lindholt, Sørensen, et al. 2010; Thompson et al. 2012), and nationwide screening programs have been launched in several countries (Lederle 2008; Wanhainen and Björck 2011; Davis et al. 2013). However, how the information from the screening programs affect the patients' physical and psychological well-being has not been considered in the cost-benefit calculations.

men participates in the screening. The baseline examination includes a single ultrasound scan where the maximum infrarenal anteroposterior diameter is measured according to the "leading edge to leading edge" principle with the ultrasound transducer longitudinally to the aorta. The ultrasound is estimated to have an error margin (variability) of about ± 4 mm (Gürtelschmid et al. 2014).

In addition to the screening of the 65-year-old men, from 2011 and onwards, all 70-year-old men are also invited to the screening in Uppsala County, which follows the same procedure as for the 65-year-old men. This older group also includes individuals who were on the earlier screening when they were 65-year-old.[5]

Most counties use 30 mm as the cut-off diameter. However, Uppsala County use 25 mm as the cut-off diameter based on the results from contemporary reports indicating that an aorta between 25 and 29 mm should be classified as a subaneurysmal aorta, or an "aneurysm in formation." Rescanning after five years has been recommended for this subgroup.

About 2,000 individuals annually are invited to an ultrasound examination of the abdominal aorta in Uppsala County. About 1.5 percent of those examined have an aorta between 25 and 29 mm and 1.5 percent an aorta equal to or larger than 30 mm.

However, screening for AAA is not uncontroversial. For example, Johansson et al. (2015) argue that the screening should be revisited since they find reduced benefits in modern populations. Relevant for this study is that they also argue that the psychological cost of burden may be too high for the individuals to bear.

2.3 The information provided within the screening program

The result from the ultrasound scan is instantly communicated to the participant by the ultrasound technician, and subjects with a screening-detected AAA or with an subaneurysmal aorta are scheduled for an appointment with a vascular surgeon or nurse, and are included in a surveillance program depending on size of the aneurysm. The information given to the patients is standardized and summarized in Table 1.

At the screening, from 2009 and onwards in Uppsala County, all individuals are asked to complete a health questionnaire with questions about height, weight, earlier/current illnesses, tobacco usage (i.e. if the individual is or has been a smoker) and current medication, as well as questions about marital status and country of birth. Individuals with an aorta of at least 25 mm are given or sent three questionnaires which measure the physical and psycholog-

[5] Hence, individuals with an aorta size less than 25 mm when they are 65 years old will have the same surveillance as 65-year-old individuals with an aorta size of 25-29 mm. The information of ill-health given at the screening will, however, differ.

Table 1. *Screening program procedure*

Size of aorta	Main information	Surveillance
$\leq$ 24 mm	Healthy	None
25-29 mm	Risk zone	Five years
30-39 mm	Small AAA	Two years
40-44 mm	Medium AAA	One year
45-49 mm	Large AAA	Six months
50-54 mm	Very large AAA	One month
$\geq$ 55 mm	AAA needing surgery	Immediately

Notes: Surveillance consists of a follow-up screening at the specified time.

ical well-being, the health related quality of life and AAA specific problems. In general, the individuals answer the questionnaires within two months of the screening (Table 5). When applicable, life style advices (e.g. "quit smoking") are given, and individuals with a very large AAA (at least 55 mm) are assessed for surgery.

Each year about 50-70 randomly picked individuals from the healthy group ($\leq$ 24 mm) from the cohort of 65-year-olds and 70-year-olds respectively are asked to answer the same questionnaires as the other groups with the exception of the AAA specific questionnaire. This group is considered to be healthy in the following study.

In the analysis we use individuals with a minimum aorta size of 20 mm and a maximum aorta size of 34 mm, implying that we will use the information provided at the 25 mm and the 30 mm cut-offs respectively.[6] This means that, at the 25 mm cut-off, we compare those individuals that receive the information that they are healthy and that there will be no further contact with the health care system with those individuals that receive the information that they are in the risk zone for AAA and that there will be a follow-up after five years. We consider this as being a comparison between one that is given information about being healthy with one that is given information about not being healthy.

At the 30 mm cut-off, we compare those individuals that receive the information that they are in the risk zone for AAA and that there will be a follow up after five years with those individuals that receive the information that they have a small AAA and that there will be a follow up after two years. We consider this as being a comparison between two individuals in which one receives the information about having worse health than the other, but in which the one with worse health also receives the information that he will have a quicker follow-up. It is worth stressing that the information provided to the

[6]Relatively few individuals have an aorta size above 34 mm, making the RD regressions problematic. We drop individuals below 20 mm both because a very small aorta can be a problem in itself, and to make the sample bandwidth symmetrical.

Table 2. *Effect on well-being*

	Action	Well-being
(1)	Negative health information	-
(2)	More active surveillance	+
(3)	Combination of (1) and (2)	?

two groups about AAA, its risk factors, and its natural development is very similar.

3 Theoretical framework

In this section we discuss the effects we should expect from the information provided in the AAA screening program on the individuals' well-being.

Suppose that individuals maximize $U(\phi\tilde{H}_t, B_t)$, where $\phi \in [0,1]$, $\tilde{H}_t$ is the individual's belief about his health capital, and B denotes home goods (enjoyments, obligations, etc.). $\tilde{H}_t$ consists of the true health and an error term, $\tilde{H}_t = H_t + \varepsilon$. The screening will reveal the true health for the individual. Since an abdominal aortic aneurysm is unknown for the individual, the screening is likely a negative information shock if it reveals that the individual has an aneurysm (i.e., $\varepsilon > 0$). On the other hand, for individuals with an aneurysm, the screening will offer a more active surveillance of the aneurysm and the health status. *Ceteris paribus*, this likely has a positive effect on the individuals well-being (i.e., increase B_t).

How shall we interpret the information provided in the screening program (Table 1)? For the information provided around the 25 mm cut-off we think it is fairly straightforward: since this can be considered as a comparison between one that is given information about being healthy with one that is given information about being less healthy, we would expect to see a negative effect on well-being at the 25 mm cut-off. However, for the information provided around the 30 mm cut-off, it is less clear: since this can be considered as being a comparison between two individuals who receives different information on health and surveillance (one individual is less healthy, but also receives information that he will be under better surveillance), it is not clear whether we should expect a positive or a negative effect at 30 mm. If the health information dominates the surveillance information, we would expect a negative effect at the 30 mm cut-off; if it is the other way around, we would expect a positive effect at the cut-off.[7] Table 2 summarizes the discussion in this section.

[7] This could be the case if the individual, for example, expect the surveillance to include additional health check-up.

4 Data

To measure the subjective health status of the individuals, we use the three different questionnaires that are provided to the patients: SF-36, EQ-5D-3L and an AAA specific questionnaire developed by Anders Wanhainen (which we call the Wanhainen questionnaire).

The three questionnaires are rich on information on several potential health outcomes. The richness enable us to measure the subjective physical as well as the psychological well-being. The downside of the richness is that it is hard to know beforehand which of the outcomes that are relevant for the case under study. We proceed by not making a selection of outcome variables ex ante, but rather to let data speak by itself (see further discussion on this in section 6.1).

As the main measure of physical and psychological well-being we use the Short Form-36 (SF-36) questionnaire (Ware, Kosinski, et al. 2000). SF-36 has repeatedly demonstrated high reliability and validity (Ware and Sherbourne 1992; McHorney et al. 1993). We construct the two standard summary measures, Physical health and Mental health, and eight suboutcome indexes, as explained in the SF-36 manual.[8] In all cases, a higher value represents a better health.

To assess health related quality of life, we use EQ-5D-3L (The EuroQol Group 1990). EQ-5D consists of two parts: A self-reported classification on five dimensions of health, and a self-rated global valuation of perceived health using a Visual Analogue Scale (VAS). There is evidence supporting the reliability and validity of the EQ-5D (Brooks and The EuroQol Group 1996). EQ-5D has two outcomes, a summary score measure and the VAS. The score measure is calculated according to the EQ-5D manual.[9] For both outcomes a higher value represents a better health.

The Wanhainen questionnaire is a questionnaire created specifically for the AAA screening in Uppsala. It consists of ten questions on a Likert scale from 0 to 10. We have constructed a summary index measure by calculating the average of all questions. We transformed the answers to the questions so that better subjective health (less anxiety, more knowledge etc.) gives a higher value in the index.[10] For a summary of all the outcomes in the three questionnaires used, see Table 3.

[8] The weights used for the summary measures are calibrated for the US. There are no calibrated measures available for Sweden. However, since we are only interested in comparing individuals in this study with each other and not the absolute levels, this poses no problem for us.

[9] The weights used for the summary measures are calibrated for the Great Britain. See footnote 8.

[10] More specifically, we recoded questions Q1, Q2, Q3, Q4, Q5 and Q10 so that an answer of 0 is 10 and vice versa when calculating the index measure.

Table 3. *Outcomes*

Variable	Range	Description
SF-36		
Physical health	0–100	Physical health (summary measure)
Mental health	0–100	Psychological well-being (summary measure)
Physical func	0–100	Physical functioning
Role lim phys	0–100	Role limitations due to physical health
Role lim emo	0–100	Role limitations due to emotional problems
Energy	0–100	Energy/fatigue
Emo well-being	0–100	Emotional well-being
Social func	0–100	Social functioning
Pain	0–100	Pain
General health	0–100	General health
EQ-5D-3L		
Score	-0.33–1	Score (summary measure)
VAS	0–100	Visual Analogue Scale
Wanhainen questionnaire		
Index	0–10	Wanhainen index (summary measure)
Q1	0–10	*"I wonder what caused my AAA"*
Q2	0–10	*"I feel fear/anxiety about my AAA"*
Q3	0–10	*"My relatives are concerned about my AAA"*
Q4	0–10	*"My relatives concern can be troublesome for me"*
Q5	0–10	*"My relatives sometimes treat me different because of my AAA"*
Q6	0–10	*"My doctor has informed me sufficiently"*
Q7	0–10	*"I try to learn more about AAA"*
Q8	0–10	*"I do not feel ill because of my AAA"*
Q9	0–10	*"I am glad to learn about having an AAA"*
Q10	0–10	*"Because of my AAA I have been forced to give up activities"*

Notes: All outcomes are constructed so that a higher value represents better health status. The Wanhainen questionnaire consists of ten different statements on a Likert scale.

We have questionnaires for in total 1,019 individuals between the years 2009-2014.[11] We only use the first set of questionnaires from each individual, i.e., we do not use follow-up questionnaires. We exclude all individuals with ages other than 64-66 and 69-71, i.e., individuals that is not part of the main screening program (see Table 4). We also exclude all individuals in the sample with an aorta size smaller than 20 mm (427 observations) or larger than 34 mm (81 observations). This is due to two reasons. First, we want all individuals in the sample to be roughly comparable to the individuals close the nearest cut-off. A very small aorta is a health risk in itself, which could bias our results. Second, in the screening program there are cut-offs at 40 mm, 45 mm, 50 mm and 55 mm, but there are too few individuals around each cut-off to make estimations meaningful. Some individuals answered the questionnaires after more than four months, or 120 days, and are excluded from the sample (see Table 5 for the cumulative response rate in the sample used).

As is often the case with questionnaires some individuals did not answer all questions. We exclude 12 individuals who did not fully respond to the EQ-5D and the SF-36 questionnaires. There are 15 individuals left who did not fully respond to the Wanhainen questionnaire.[12] Since this questionnaire is not our primary outcome, it would be too restrictive to also exclude these individuals.

The final sample consists of 407 males with an aorta size between 20 and 34 mm. 188 observations belong to the healthy group ($\leq$ 24 mm), 163 to the risk zone group and 56 to the small aorta size group (30-34 mm).

From the health related questionnaire we have information about smoking history (all observations) and BMI (390 observations). We use this subsample to check the RD identification assumption.

The age frequency and the response time for the sample is presented in Table 4 and 5 respectively. Means and standard deviations for each group and variables used are presented in Table 6. A histogram of the aorta size of the individuals is shown in Figure 1.

The distribution of the aorta size is not ideal for the RD design, with some spikes in the frequency. This can be due to a couple of reasons. We cannot expect the distribution to be smooth around the 25 mm cut-off since the healthy group consists of only a subsample of the individuals with an aorta size below 25 mm. Therefore, the absolute levels of the number of observations are not comparable around this cut-off. However, we may also suspect that the ultrasound technician, who knows about the cut-offs and the error margin of the ultrasound scan, have a tendency to register aorta sizes just below a cut-off as being on or slightly above the cut-off since it is often regarded better to be treated (or "overtreated") than non-treated. In addition, humans have a ten-

[11] All 65-year-old men in Uppsala County were eligible, and, additionally, from 2011 and onwards, all 70-year-old men.

[12] The results does not change in any important sense due to the exclusion/inclusion of these individuals. Only individuals with an aorta size of at least 25 mm answers the Wanhainen questionnaire, so this outcome is measured only for the 30 mm cut-off.

Table 4. *Age frequency*

Age	Frequency	Percent
64	36	8.85
65	231	56.76
66	1	0.25
69	38	9.34
70	100	24.57
71	1	0.25
N	407	

Notes: Frequency of individuals by age.

Table 5. *Follow-up time*

Days	Treatment	Control	All
0	0.00	86.70	40.05
7	0.00	95.21	43.98
14	4.11	98.40	47.67
30	40.64	99.47	67.81
60	83.11	100.00	90.91
90	94.98	100.00	97.30
120	100.00	100.00	100.00

Notes: Cumulative percent of the number of days between screening and questionnaire answers.

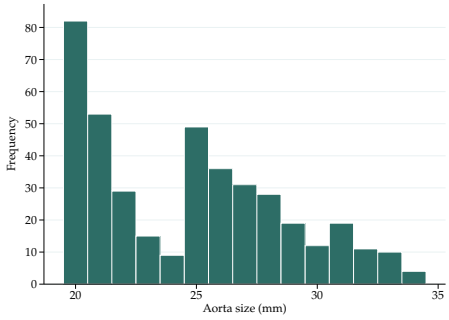


Figure 1. Below 25 mm the data consists of a random sample of the healthy individuals, while all individuals with an aorta size of 25 mm or larger are included.

dency to prefer "anchor numbers" (20, 25, 30, etc.). The RD design demands that the assignment variable, the aorta size, is smooth in the sense that the individuals cannot self-select into being in a specific group. In this case this should not be a problem, since it is the ultrasound technician, not the patient, who makes the decision without the patient's knowledge. If the ultrasound technician is systematically biased, however, it would still be a violation of the RD assumption. In the sensitivity analyses we will deal with this potential problem by adopting a donut estimator.

5 Empirical strategy

We apply the quasi-experimental regression discontinuity (RD) design. An RD design may be appropriate if we want to estimate a causal effect of a treatment, but randomization is unfeasible or not appropriate. If there is a rule which decide whether an individual is treated or not this could create discontinuities which an RD design can exploit (Angrist and Pischke 2009; Lee and Lemieux 2010).

Table 6. *Descriptive statistics*

	20-24 mm		25-29 mm		30-34 mm		All	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Aorta size	21.02	1.16	26.58	1.38	31.55	1.22	24.70	3.97
Covariates:								
Age	66.64	2.42	66.55	2.33	66.09	2.10	66.53	2.34
Height (cm)	180.17	6.54	180.58	6.52	181.61	5.67	180.52	6.43
Weight (kg)	89.69	13.68	88.74	13.14	89.80	11.20	89.33	13.15
BMI	27.70	4.25	27.19	3.64	27.16	3.07	27.43	3.88
Smoking history	0.59	0.49	0.77	0.42	0.88	0.33	0.70	0.46
Follow-up time	1.18	4.11	39.73	21.78	42.82	24.85	22.35	25.83
Aggregated:								
Physical health	50.90	7.92	48.98	9.68	46.95	10.85	49.59	9.17
Mental health	58.02	5.98	55.72	8.00	56.10	9.42	56.84	7.44
EQ-5D Score	0.88	0.17	0.86	0.18	0.81	0.23	0.86	0.18
EQ-5D VAS	84.69	14.01	80.95	16.53	76.84	21.38	82.11	16.40
Wan index	.	.	7.86	1.23	7.73	1.12	7.82	1.20
Disaggregated:								
Physical func	87.68	17.07	83.18	20.75	80.00	24.77	84.82	19.93
Role lim phys	89.89	25.02	81.90	33.02	79.91	36.75	85.32	30.41
Role lim emo	94.50	17.22	88.96	26.20	85.12	32.98	90.99	23.87
Energy	80.29	17.26	72.70	22.09	70.36	25.35	75.88	20.89
Emo well-being	89.32	13.48	84.66	17.00	86.64	17.04	87.09	15.59
Social func	93.22	15.37	90.41	17.98	90.40	15.99	91.71	16.56
Pain	81.54	22.46	79.63	25.29	75.07	28.72	79.88	24.57
General health	79.55	17.14	73.10	20.88	68.51	22.59	75.45	19.88
Q1	.	.	4.15	3.71	5.13	3.73	4.41	3.73
Q2	.	.	2.27	2.75	2.79	2.96	2.41	2.81
Q3	.	.	2.44	2.93	3.45	2.96	2.71	2.97
Q4	.	.	1.41	2.23	1.89	2.56	1.54	2.33
Q5	.	.	0.78	1.68	0.73	1.70	0.77	1.68
Q6	.	.	9.30	1.64	9.65	0.67	9.39	1.45
Q7	.	.	4.82	3.45	5.64	3.13	5.04	3.38
Q8	.	.	7.98	3.72	7.88	3.76	7.95	3.72
Q9	.	.	8.35	3.06	8.58	2.90	8.41	3.01
Q10	.	.	0.57	1.52	0.53	1.53	0.56	1.52
N	188		163		56		407	

Notes: Individuals with an aorta size less than 25 mm do not answer the Wanhainen questionnaire. There are missing data for BMI, smoking history and the Wanhainen questionnaire for some individuals.

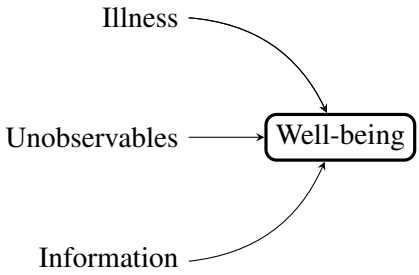


Figure 2. AAA is an asymptomatic illness and will not affect well-being, but unobservable background characteristics and the information of having the illness may.

The ideal experiment would be to randomize the individuals into different groups and give them different information about their health regarding AAA. Due to the randomization, the groups can be regarded as equal in unobservable characteristics that otherwise could bias the estimated effect of the treatment. A randomization of the individuals would therefore allow a causal interpretation of the information effect. The identification problem is illustrated in Figure 2.

However, in this case there has been no opportunity to randomize the individuals. Even if it had been possible it could be argued that it is unethical to randomize the individuals into groups which are given different information about an illness that is potentially fatal. Instead, we exploit the use of fixed boundaries in the screening program that determine the information the individuals are given. By applying the RD design we compare individuals in the sample just below and just above the boundary for a treatment, the so called cut-off point. We use a specific strategy called local linear regression.[13] For example, we can compare the individuals with a maximum infrarenal aortic diameter of 24 mm, who receive the information that he is healthy regarding AAA, with individuals with a diameter of 25 mm, who receive information that he is in the risk zone for AAA. Since the difference in diameter is so small and the choice of 25 mm as a cut-off is somewhat arbitrary, a group of individuals with 24 mm can be assumed being equal or very similar to a group of individuals with a diameter of 25 mm concerning both observable and unobservable characteristics, with the single exception of the treatment (i.e., the information given). Hence, the RD design will estimate the causal effect of the health information.

The RD design allows for the use of a larger bandwidth if we are willing to assume that the larger difference in aorta size is not important. It is also possible to use more than one cut-off point in the estimation. As discussed

[13] Another commonly used strategy in the RD context is to use higher order polynomials, which try to mimic the data when moving away from the cut-off. However, the drawback is that we then rely on points far away from the cut-off when estimating the causal effect of the treatment. We will therefore not use this strategy.

Table 7. *Group comparisons*

Information	Surveillance	Size		Size	Information	Surveillance
Healthy	None	20-24	vs	25-29	Risk zone	Five years
Risk zone	Five years	25-29	vs	30-34	Small AAA	Two years

Notes: The respective groups that will be compared, and the information and surveillance each group receives from the screening. Surveillance consists of a follow-up screening at the specified time. Size in millimeters (mm).

earlier, we use the two lower cut-off points used in the screening program. The relevant group comparisons and the respective information to the individuals are presented in Table 7. We will use a bandwidth of 5 mm and estimate both the 25 mm cut-off and the 30 mm cut-off in the same equation.[14]

The group with a diameter of 20-24 mm is healthy, and receives no treatment in the screening program. The healthy group will be compared with the group with a diameter of 25-29 mm, who are considered to be in the risk zone for AAA, and will have surveillance follow-up in five years and receives information about this. This will allow an estimation of the effect of the information of being in the risk zone for AAA with a sparse surveillance compared to information about being healthy. The risk zone group will then be compared with the group with a diameter of 30-34 mm. These two groups differ in the way that the latter group receives the information that they have a small AAA and a more frequent surveillance (follow-up in two years). Hence, this will allow an estimation of the effect of information of having a small sized AAA and receive an increased surveillance compared to information of being in the risk zone and sparse surveillance.

While the healthy group will function only as a control group and the small sized AAA group only will function as a treatment group, the risk zone group will function both as a control group and a treatment group, depending on comparison at hand. The equation to be estimated is given by the following two cut-off RD specification:

$$\begin{aligned} health_i = {} & \tau_1 \mathrm{I}(X_i \geq 25) + \beta_1 \left[(X_i - 25) * \mathrm{I}(X_i \geq 25)\right] + \\ & \tau_2 \mathrm{I}(X_i \geq 30) + \beta_2 \left[(X_i - 30) * \mathrm{I}(X_i \geq 30)\right] + \\ & \delta X_i + \alpha + \varepsilon_i, \end{aligned} \tag{1}$$

where X_i is the aorta diameter size in mm, $(X_i - 25)$ and $(X_i - 30)$ are the distances in millimeters from the respective cut-off, and

$$\mathrm{I}(X_i \geq 25) = \begin{cases} 1 & \text{if } X_i \geq 25 \text{ mm} \\ 0 & \text{otherwise,} \end{cases} \qquad \mathrm{I}(X_i \geq 30) = \begin{cases} 1 & \text{if } X_i \geq 30 \text{ mm} \\ 0 & \text{otherwise.} \end{cases}$$

[14]The point estimates are the same and the standard errors only insignificantly affected compared with single cut-off estimations.

The coefficients of interest are τ_1 and τ_2, which show the jump in health outcomes at the respective cut-off. The interaction terms allow the slopes to be different before and after the cut-offs.

When the assignment variable is discrete one must rely on an extrapolation of the data at the cut-off point. Lee and Card (2008) recommends clustering of the data on the discrete variable. However, since we would have too few clusters (15), clustered standard errors are unreliable (Angrist and Pischke 2009). Instead, we rely on robust standard errors in our regressions (Wooldridge 2010; Fredriksson and Öckert 2013). The robust standard errors are sometimes larger and sometimes smaller than the clustered standard errors.

6 Baseline results

In this section we present the main results. We do this in two ways; first we provide a graphical presentation of the baseline results in section 6.1, and then we turn to the regression analysis in section 6.2. Sensitivity and robustness analyses are provided in section 7.

6.1 Graphical analyses

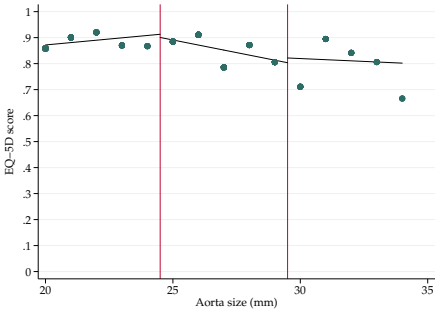
We start the graphical analyses with the aggregated measures of the individual's subjective health.

Beginning with the two measures from the EQ-5D questionnaire, we find no jumps at the cut-offs (Figures 3a and b). There is, however, indication of a broken trend at the 25 mm cut-off, but the binned values are quite scattered and do not lie closely connected to the regression lines. It seems difficult to draw any clear conclusion from the observed pattern. Turning to the two aggregate measures of the SF-36 questionnaire, the general pattern seems to be the same as for the EQ-5D measures for physical health (Figure 3c). There is an indication of a positive jump at 30 mm for psychological well-being (Figure 3d), but nothing at 25 mm. The bins are much less scattered and are relatively tight around the regression lines compared with the EQ-5D measures. The Wanhainen index (Figure 3e) does not deviate from the above pattern. The observations are relatively close to the regression lines, with some indication of a positive jump at the 30 mm cut-off.[15]

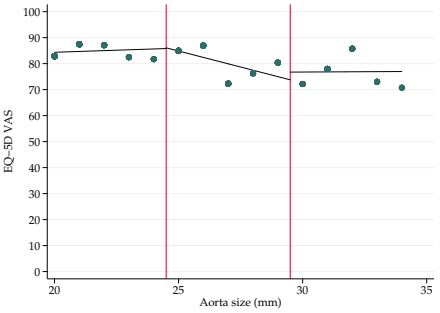
So far we have had a look at the more aggregated measures of subjective health. These measures can, however, hide important variation over the more disaggregated and specific measures of health. We now turn to look separately at the eight different measures within the SF-36 questionnaire.

It is clear from Figures 4a-h that we observe the same general pattern for all disaggregated outcomes as for the aggregated measures; that is, no effect at 25

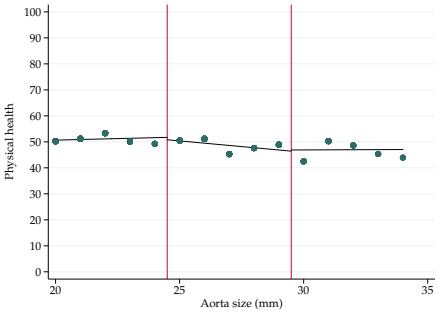
[15]There are no measures for those with an aorta size below 25 mm since the healthy group does not answer the AAA specific questionnaire.



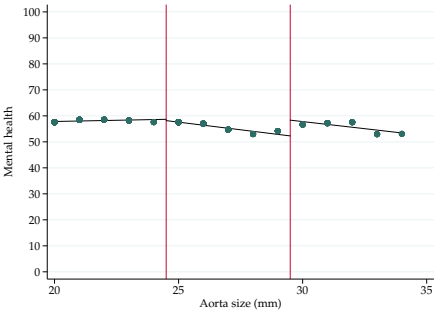
(a) EQ-5D Score



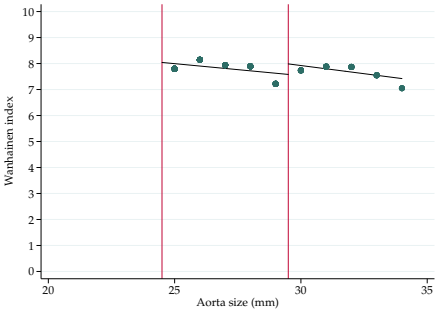
(b) EQ-5D VAS



(c) SF-36: Physical health

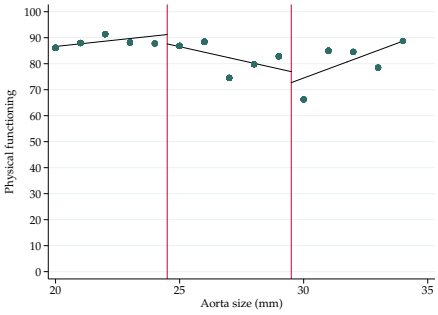


(d) SF-36: Mental health

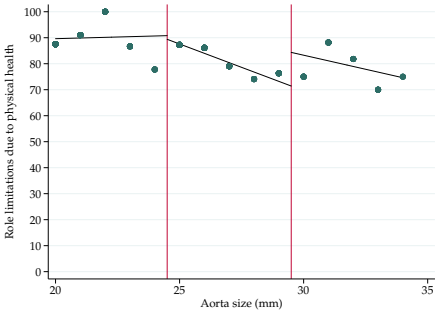


(e) Wanhainen index

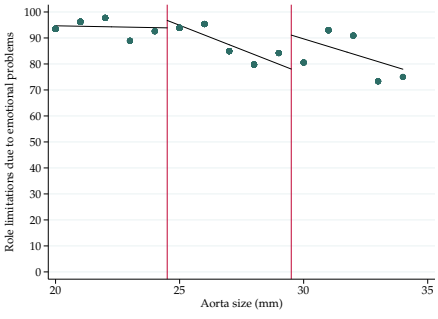
Figure 3. Aggregate health outcomes.



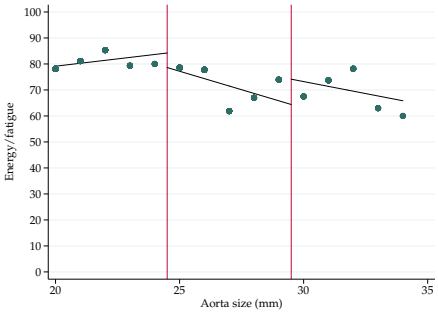
(a) Physical functioning



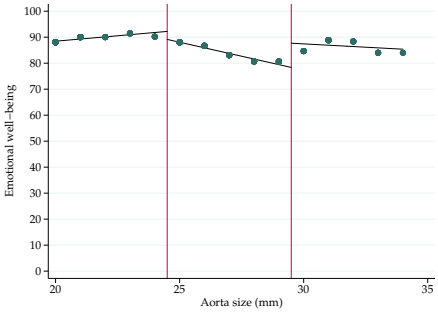
(b) Role lim. due to physical health



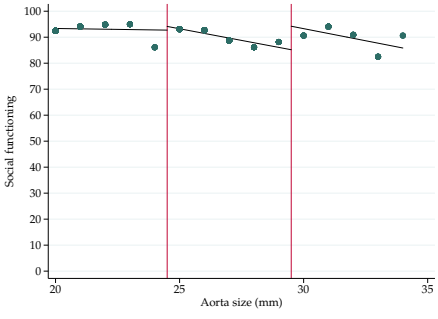
(c) Role lim. due to emotional problems



(d) Energy/fatigue



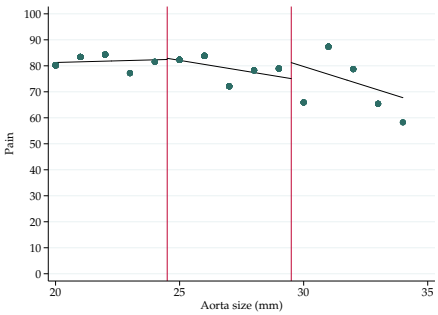
(e) Emotional well-being



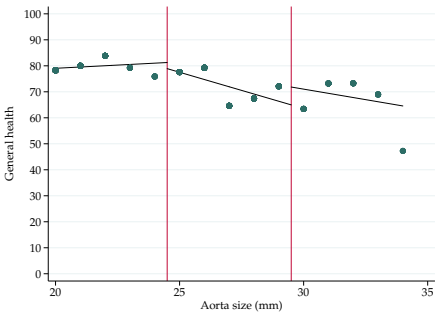
(f) Social functioning

Continued on the next page.

Figure 4. SF-36: Disaggregate health outcomes.



(g) Pain



(h) General health

Figure 4 (cont.). SF-36: Disaggregate health outcomes.

mm and possibly a positive effect at 30 mm. The magnitude of the effects, and the pattern of the binned data, are clearer in some cases, such as for emotional well-being (Figure 4e) and social functioning (Figure 4f).

To conclude, the overall pattern observed in the graphical analyses is that, if there is an effect, it seems to be positive at the 30 mm cut-off. No effect at the 25 mm cut-off can be seen. However, from an RD analysis point of view, it is difficult to see how a trustworthy analysis can be made and draw conclusions for some of the outcomes due to the scattered pattern of the binned data. For the outcomes where we observe a more stable pattern of the binned data (i.e., physical health, mental health, the Wanhainen index, emotional well-being, and social functioning), we will turn to more formal estimations of the potential effects, to examine the significance both statistically and clinically (i.e., the magnitude of the effects).[16] It is worth noting that the stable patterns observed are mainly on outcomes measuring psychological well-being.

6.2 Baseline estimations

We use the full aorta size interval of 20-34 mm, hence, the two cut-offs simultaneously, and estimate the effects of new health information on the individuals' well-being (as expressed in terms of physical health, mental health, the Wanhainen index, emotional well-being, and social functioning). The RD estimates, obtained from estimation of Equation 1, are provided in Table 8.

The signs of the point estimates follow the pattern observed in the graphical analyses; negative at the 25 mm cut-off (social functioning positive but close to zero) and positive at the 30 mm cut-off. None of the estimates at 25 mm are statistically significant, and are clinically relatively small. However, three

[16] RD estimates for the other outcomes are presented in the appendix.

Table 8. *RD analysis*

	Aggregated			Disaggregated	
	Physical health	Mental health	Wan index	Emo well-being	Social func
25 mm	-1.471	-1.182		-4.574	0.594
	(2.437)	(1.688)		(3.470)	(5.546)
30 mm	0.902	6.103**	0.370	10.140*	9.012*
	(3.152)	(2.777)	(0.368)	(5.312)	(4.958)
Mean	49.589	56.836	7.821	87.086	91.708
N	407	407	204	407	407

Notes: Individuals with an aorta size of 20-34 mm. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

of the estimates at the 30 mm cut-off are statistically significant, and arguably clinically significant.[17]

7 Sensitivity analyses

A crucial assumption in the RD design is that nothing else changes at the cut-off, i.e., that the observations around the cut-offs are equal in all other characteristics other than the treatment, so that we can consider the treatment as good as randomized. If this is not the case, for example, if the ultrasound technician's decision of the aorta size is a function of the individual's background characteristics (like weight or smoking history), and these covariates in turn affect the well-being of an individual, the estimations would be biased.

In this section we examine if our baseline results are sensitive to potential bunching. To do this, we will use two different sensitivity and robustness analyses. We start out by using covariates as outcomes in the estimations in section 7.1, followed by conducting donut estimations in section 7.2. In the appendix, we provide two additional sensitivity analyses. First, we redo the estimations in section 7.1 using a donut estimator, and second, we redo the main analysis but include the covariates in the estimations.

7.1 Estimations with covariates as outcomes

One way to check if the assumption of treatment assignment being as good as random is fulfilled is to use covariates as the outcome variable instead of

[17] In addition to the multiple cut-off estimations, we have also estimated the two cut-offs separately (for the sample in the interval 20-34 mm with varying bandwidth), but find few statistically significant effects. These additional results are available upon request.

Table 9. *RD analysis with covariates as outcomes*

	Age dummy	Smoking history	BMI	Height	Weight	Follow-up time
25 mm	-0.091	0.009	1.123	-0.667	3.138	37.379***
	(0.140)	(0.129)	(1.110)	(1.818)	(3.683)	(3.194)
30 mm	0.017	0.083	1.980*	-0.575	5.828	8.348
	(0.150)	(0.129)	(1.136)	(2.007)	(3.905)	(7.408)
Mean	0.344	0.692	27.432	180.490	89.327	21.387
N	390	390	390	390	390	390

Notes: Individuals with an aorta size of 20-34 mm, with information on both BMI and smoking. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

the health measures of interest (Lee and Lemieux 2010). If the individuals are as good as randomized, the groups below and above the cut-offs should be balanced, i.e., there should not be any statistically significant effects in the estimations. We have data for six different covariates for a subsample of our sample: Age, smoking history, height, weight, BMI, and the number of days before answering the questionnaires.[18]

The results, presented in Table 9, are fairly reassuring. Four of the outcomes (age, smoking history, height, and weight) have no statistically significant changes at the cut-off points. For the remaining individual outcome variable, BMI, there is one marginally statistically significant change (a positive jump at the 30 mm cut-off), and one insignificant change. Regarding the last covariate, the number of days before answering the questionnaires, individuals in the risk zone answer the questionnaires about a month later than individuals in the control group. However, this is not necessarily a problem. It is reasonable to assume that individuals in the control group see no effects on well-being, so the results for this group should be stable. We are mainly interested in the long-term effects on well-being for individuals who receive negative health information, which we are able to estimate using this data.

The results in Table 9 looks fairly good overall and indicate that the baseline results are reliable. It is a bit troublesome to get a significant estimate for BMI when using covariates as outcomes, but in the end, what matters is to what extent the estimates are sensitive to this. In the appendix we examine how sensitive the baseline estimates are to the inclusion of the covariates.

Table 10. *RD donut analysis*

	Aggregated			Disaggregated	
	Physical health	Mental health	Wan index	Emo well-being	Social func
25 mm	-1.661	-0.522		-4.330	-1.820
	(3.535)	(2.727)		(6.049)	(6.528)
30 mm	10.471*	10.480**	0.564	16.594*	17.526*
	(5.646)	(4.399)	(0.638)	(9.329)	(9.022)
Mean	49.765	56.873	7.908	87.333	91.903
N	318	318	133	318	318

Notes: Individuals with an aorta size of 20-34 mm except 24-25 mm and 29-30 mm. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

7.2 Donut estimations

Figure 1 suggests that some individuals just around the cut-offs may be incorrectly registered since there is some indication of bunching to the right of the cut-offs. Bunching violates the assumption of the treatment assignment being as good as random. One reason for the tendency to bunching could be that the ultrasound technician must make a decision whether the aorta size of an individual is below or above the cut-off, and that it is usually regarded as better to be "overtreated" than "undertreated". If the registration of the aorta size is correlated with characteristics of the patients that might affect their well-being, like weight, this might bias our results.

One way of dealing with the potential problem of sorting at the cut-offs is to estimate a "donut version" of the RD design (Barreca et al. 2011), dropping all observations just around the cut-offs – observations that may be incorrect – and only use observations farther away from the cut-off, leaving a "donut hole". Therefore, we proceed with estimating the baseline regressions but drop individuals with 24-25 or 29-30 mm aorta size. We estimate the same outcomes as in the baseline regressions.

The results are presented in Table 10. The effects estimated in the baseline regressions are stable. At the 30 mm cut-off, four of five outcomes are statistically significant (if only marginally) compared with three in the baseline estimation, and the effects are clinically larger. There is no statistically significant effect at the 25 mm cut-off. Overall, the donut estimations indicate that the baseline results are reliable and relatively robust.

[18] Age is a dummy variable which takes the value 0 if the individual is between 64-66, and 1 if the individual is between 69-71, c.f. Table 4.

8 Discussion

Looking at the general pattern, the results presented in the graphical analysis, the baseline analysis, and the sensitivity analysis, indicate that the effects of the information given in the screening might go in two different directions. At the low cut-off point at 25 mm, if anything, the effects seem to be negative (i.e., the individuals react negatively on the information), and at the cut-off point at 30 mm, the effects seem to be positive.

However, from a statistical point of view, all the action seem to be at the 30 mm cut-off. There are no significant effects at the 25 mm cut-off regardless of specification. In comparison, the estimated positive effects at the 30 mm cut-off are robust to different sensitivity analyses. This leads us to the conclusion that the differential information given to the patients around the 30 mm cut-off seem to have an effect on the patients' psychological well-being, while the differential information given at the 25 mm cut-off has not.

Are the magnitudes of the estimated positive effects at the 30 mm cut-off of any importance? To get a sense of that, we relate the statistically significant point estimates in the baseline analyses to the mean value of each outcome variable.[19] In doing this, it seems like the magnitude of the estimated effects are of clinical importance. Taking mental health as an example, the point estimate (6.10) constitute almost 11 % of the overall mean of 56.84 (see the next to last column in Table 6). The corresponding figures for the other significant point estimates are both about 10 % (emotional well-being and social functioning).[20]

How can we understand the estimated positive effect on the patients' well-being after receiving information about ill-health? The theoretical discussion suggest that an individual who receives information that he has an AAA (i.e., unexpected information about ill-health) would experience a decrease in well-being. However, the discussion in section 3 also suggest that it is unclear how to interpret the information provided around the 30 mm cut-off; the patient just above 30 mm receives information that he is less healthy, but also receives information that he will be under better surveillance. Our interpretation of the positive point estimate at 30 mm is that the information of better surveillance by the health care system (a positive effect on well-being) outweighs the information of being less healthy (a negative effect on well-being).

Finally, it can be informative to relate how the RD results obtained in this paper differ from the results we would have found had we conducted a traditional OLS analysis, similar to the type of analysis conducted earlier in the literature, e.g. Lindholt, Vammen, et al. (2000) and Spencer et al. (2004). Applying a more naïve estimation strategy, we estimate an OLS model which

[19] The point estimates in the baseline analyses are more conservative than the point estimates obtained when applying the donut estimator.

[20] If we relate the point estimates to the mean for the individuals in the 25-29 mm or 30-34 mm intervals, we get very similar order of magnitudes.

Table 11. *Mean comparisons*

	Aggregated			Disaggregated	
	Physical health	Mental health	Wan index	Emo well-being	Social func
25 mm	-1.921**	-2.306***		-4.657***	-2.804
	(0.954)	(0.764)		(1.656)	(1.801)
30 mm	-2.029	0.380	-0.127	1.980	-0.012
	(1.629)	(1.401)	(0.181)	(2.628)	(2.550)
Mean	49.589	56.836	7.821	87.086	91.708
N	407	407	204	407	407

Notes: Individuals with an aorta size of 20-34 mm. Mean comparison around cut-offs. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

only include dummy variables for whether the individual has an aorta size above each respective cut-off. Comparing the OLS estimates in Table 11 with the baseline RD estimates, it is clear that the OLS results overstate the negative importance at the 25 mm cut-off in terms of statistical significance, and miss the positive effect at 30 mm. In fact, none of the estimates at the 30 mm cut-off are statistically significant. The *negative* point estimates at 30 mm provided by the OLS estimator is hard to believe given the pattern observed in the figures presented earlier.

9 Conclusions

In this paper we have examined how new health information affects individuals' subjective physical and psychological well-being.

To solve the endogeneity problem, we apply a regression discontinuity estimator on data from a screening program for the asymptomatical disease AAA in Sweden. Since screened individuals receive different information about their health, as a function of the measured size of their aorta and its relation to pre-determined cut-off levels, we are able to estimate causal effects.

We find a robust and positive significant effect on the individuals' psychological well-being when comparing those that receive information that they have a small AAA, and therefore will be under increased surveillance by the health care system, with those who receive the information that they have an enlarged aorta and increased risk for AAA, but only sparse surveillance by the health care system. This indicates that the information about increased surveillance (positive information) outweighs the information about worse health (negative information). The magnitudes of the estimated effects also indicate that they are clinically important. We do not find any statistically sig-

nificant *negative* effects of the information about ill-health on the individuals' well-being.

The positive effects on well-being indicate that the benefit side in a traditional cost-benefit analysis of the AAA screening program would gain more than the cost side if the patients' subjective well-being were taken into account.

References

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, New Jersey: Princeton University Press. ISBN: 9780691120348.

Ashton, H A, L Gao, L G Kim, P S Druce, S G Thompson, and R A P Scott. 2007. "Fifteen-year follow-up of a randomized clinical trial of ultrasonographic screening for abdominal aortic aneurysms." *British Journal of Surgery* 94 (6): 696–701.

Barreca, Alan I, Melanie Guldi, Jason M Lindo, and Glen R Waddell. 2011. "Saving babies? Revisiting the effect of very low birth weight classification." *The Quarterly Journal of Economics* 126 (4): 2117–2123.

Brooks, Richard, and The EuroQol Group. 1996. "EuroQol: The current state of play." *Health Policy* 37 (1): 53–72.

Cosford, Paul A, Gillian C Leng, and Justyn Thomas. 2007. *Screening for abdominal aortic aneurysm.* Review 2. Cochrane Database of Systematic Reviews.

Davis, Meryl, Mike Harris, and Jonothan J Earnshaw. 2013. "Implementation of the national health service abdominal aortic aneurysm screening program in England." *Journal of Vascular Surgery* 57 (5): 1440–1445.

Engström, Per, Patrik Hesselius, and Malin Persson. 2007. *Excess use of temporary parental benefit.* Working Paper 2007:18. Institute for Labour Market Policy Evaluation (IFAU).

Fredriksson, Peter, and Björn Öckert. 2013. "Life-cycle effects of age at school start." *The Economic Journal* 124 (September): 977–1004.

Gürtelschmid, Mikael, Martin Björck, and Anders Wanhainen. 2014. "Comparison of three ultrasound methods of measuring the diameter of the abdominal aorta." *British Journal of Surgery* 101 (6): 633–636.

Hansson, Anders, John Brodersen, Susanne Reventlow, and Monica Pettersson. 2012. "Opening pandora's box: The experiences of having an asymptomatic aortic aneurysm under surveillance." *Health, Risk & Society* 14 (4): 341–359.

Imbens, Guido W, and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 142 (2): 615–635.

Johansson, Minna, Anders Hansson, and John Brodersen. 2015. "Estimating overdiagnosis in screening for abdominal aortic aneurysm: Could a change in smoking habits and lowered aortic diameter tip the balance of screening towards harm." *British Medical Journal* 350:h825.

Lederle, F A. 2008. "Screening for AAA in the USA." *Scandinavian Journal of Surgery* 97 (2): 139–141.

Lee, David S, and David Card. 2008. "Regression discontinuity inference with specification error." *Journal of Econometrics* 142 (2): 655–674.

Lee, David S, and Thomas Lemieux. 2010. "Regression discontinuity designs in economics." *Journal of Economic Literature* 48 (June 2010): 281–355.

Lindholt, J S, and P Norman. 2008. "Screening for abdominal aortic aneurysm reduces overall mortality in men. A meta-analysis of the mid-and long-term effects of screening for abdominal aortic aneurysms." *European Journal of Vascular and Endovascular Surgery* 36 (2): 167–171.

Lindholt, J S, J Sørensen, R Søgaard, and E W Henneberg. 2010. "Long-term benefit and cost-effectiveness analysis of screening for abdominal aortic aneurysms from a randomized controlled trial." *British Journal of Surgery* 97 (6): 826–834.

Lindholt, J S, S Vammen, H Fasting, and E W Henneberg. 2000. "Psychological consequences of screening for abdominal aortic aneurysm and conservative treatment of small abdominal aortic aneurysms." *European Journal of Vascular and Endovascular Surgery* 20 (1): 79–83.

Lucarotti, M E, B P Heather, E Shaw, and K R Poskitt. 1997. "Psychological morbidity associated with abdominal aortic aneurysm screening." *European Journal of Vascular and Endovascular Surgery* 14 (6): 499–501.

McGregor, J C, J G Pollock, and H C Anton. 1975. "The value of ultrasonography in the diagnosis of abdominal aortic aneurysm." *Scottish Medical Journal* 20 (3): 133–137.

McHorney, Colleen A, John E Ware Jr., and Anastasia E Raczek. 1993. "The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs." *Medical Care* 31 (3): 247–263.

Moll, F L, J T Powell, G Fraedrich, F Verzini, S Haulon, M Waltham, J A van Herwaarden, et al. 2011. "Management of abdominal aortic aneurysms clinical practice guidelines of the European society for vascular surgery." *European Journal of Vascular and Endovascular Surgery* 41:S1–S58.

Sakalihasan, N, R Limet, and O D Defawe. 2005. "Abdominal aortic aneurysm." *The Lancet* 365 (9470): 1577–1589.

SBU. 2008. *Screening för bukaortaaneurysm.* SBU Alert-Rapport 2008-04. SBU.

Scott, R A P, and The Multicentre Aneurysm Screening Study Group. 2002. "The multicentre aneurysm screening study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: A randomised controlled trial." *The Lancet* 360 (9345): 1531–1539.

Sexton, Mary, and J Richard Hebel. 1984. "A clinical trial of change in maternal smoking and its effect on birth weight." *JAMA* 251 (7): 911–915.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian. 2001. "Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota." *Journal of Public Economics* 79 (3): 455–483.

Spencer, Carole A, Paul E Norman, Konrad Jamrozik, Raywin Tuohy, and Michael Lawrence-Brown. 2004. "Is screening for abdominal aortic aneurysm bad for your health and well-being?" *ANZ Journal of Surgery* 74 (12): 1069–1075.

Strömberg, David. 2004. "Radio's impact on public spending." *The Quarterly Journal of Economics* 119 (1): 189–221.

Svensjö, Sverker, Martin Björck, Mikael Gürtelschmid, Khatereh Djavani Gidlund, Anders Hellberg, and Anders Wanhainen. 2011. "Low prevalence of abdominal aortic aneurysm among 65-year-old Swedish men indicates a change in the epidemiology of the disease." *Circulation* 124 (10): 1118–1123.

The EuroQol Group. 1990. "EuroQol – a new facility for the measurement of health-related quality of life." *Health Policy* 16 (3): 199–208.

The UK Small Aneurysm Trial Participants. 1998. "Health service costs and quality of life for early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms." *The Lancet* 352 (9141): 1656–1660.

Thompson, S G, H A Ashton, L Gao, M J Buxton, and R A P Scott. 2012. "Final follow-up of the multicentre aneurysm screening study (MASS) randomized trial of abdominal aortic aneurysm screening." *British Journal of Surgery* 99 (12): 1649–1656.

Wanhainen, A. 2008. "How to define an abdominal aortic aneurysm – influence on epidemiology and clinical practice." *Scandinavian Journal of Surgery* 97 (2): 105.

Wanhainen, Anders, David Bergqvist, Kurt Boman, Torbjörn K Nilsson, Jörgen Rutegård, and Martin Björck. 2005. "Risk factors associated with abdominal aortic aneurysm: A population-based study with historical and current data." *Journal of Vascular Surgery* 41 (3): 390–396.

Wanhainen, Anders, and Martin Björck. 2011. "The Swedish experience of screening for abdominal aortic aneurysm." *Journal of Vascular Surgery* 53 (4): 1164–1165.

Ware, John E, Mark Kosinski, James E Dewey, and Barbara Gandek. 2000. *SF-36 Health Survey: Manual and Interpretation Guide.* Boston, Massachusetts: Quality Metric Inc.

Ware, John E, and Cathy Donald Sherbourne. 1992. "The MOS 36-Item short-form health survey (SF-36): I. Conceptual framework and item selection." *Medical Care* 30 (6): 473–483.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* Cambridge and London: MIT press.

Zhao, Meng, Yoshifumi Konishi, and Paul Glewwe. 2013. "Does information on health status lead to a healthier lifestyle? Evidence from China on the effect of hypertension diagnosis on food consumption." *Journal of Health Economics* 32 (2): 367–385.

Appendix

The appendix consists of two parts. In section A the estimates for the outcomes not included in the main analysis are presented. In section B we provide two additional sensitivity analyses.

A Other outcomes

In Table A1 the estimates for the outcomes which were not included in the main analysis are presented. As the figures in section 6.1 suggest, the estimates are more noisy. There is no statistically significant result. However, it should be noted that the direction of the point estimates in general are in line with the outcomes in the main analysis.

B Additional sensitivity analyses

In Table A2 we redo the sensitivity analysis in section 7.1 but use the donut estimator. The results are fairly similar to those in Table 9, but with no statistically significant estimates, except for the follow-up time at the 25 mm cut-off, indicating that the donut estimator can be used if we are willing to assume that the individuals farther away from the cut-off are as good as randomly assigned.

Finally, we redo the main analysis but include age, BMI, smoking history and follow-up time as covariates. In the ideal case, inclusion of covariates should not affect the point estimates since the covariates are (assumed to be) independent of the treatment. In practice, however, inclusion of covariates can improve the precision, reduce small sample bias and reduce biases when

Table A1. *RD analysis*

	Aggregated		Disaggregated					
	EQ-5D Score	EQ-5D VAS	Physical func	Role lim phys	Role lim emo	Energy	Pain	General health
25 mm	-0.027	-1.111	-5.176	-3.336	1.081	-7.575	-0.464	-3.996
	(0.048)	(4.580)	(4.412)	(8.477)	(5.794)	(4.955)	(6.327)	(5.485)
30 mm	0.026	4.225	-1.392	13.655	13.478	10.243	5.437	7.458
	(0.063)	(5.510)	(7.791)	(10.910)	(9.623)	(7.060)	(8.572)	(6.534)
Mean	0.863	82.113	84.819	85.319	90.991	75.885	79.882	75.448
N	407	407	407	407	407	407	407	407

Notes: Individuals with an aorta size of 20-34 mm. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2. *RD donut analysis with covariates as outcomes*

	Age	Smoking history	BMI	Height	Weight	Follow-up time
25 mm	-0.038	0.060	0.729	1.424	4.615	45.102***
	(0.203)	(0.185)	(1.480)	(2.691)	(5.612)	(7.148)
30 mm	-0.061	0.052	0.852	1.685	5.353	12.599
	(0.253)	(0.192)	(1.938)	(3.294)	(7.093)	(10.924)
Mean	0.352	0.681	27.457	180.570	89.530	16.919
N	307	307	307	307	307	307

Notes: Individuals with an aorta size of 20-34 mm except 24-25 mm and 29-30 mm, with information on both BMI and smoking history. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

observations further away from the cut-off are included (Imbens and Lemieux 2008; Lee and Lemieux 2010). Since we are only able to do these estimations on a subsample (due to missing data on some of the covariates), the results are not fully comparable with the baseline estimations presented in Table 8.[21]

Comparing the results with the main analysis, it is clear that we get similar results. The same outcomes at the same cut-off are significant whether we control for covariates or not. In addition, the significant coefficients are very similar. These results indicate, first, that the baseline results are not sensitive to the inclusion of covariates, and, second, that the significant estimates for BMI and follow-up time that we found in the sensitivity analysis is not of qualitative importance for the conclusions since the estimates in general are much the same.

[21] However, the estimates are close to what we get if we exclude individuals with missing data in the main analysis.

Table A3. *RD analysis with covariates*

	Aggregated			Disaggregated	
	Physical health	Mental health	Wan index	Emo well-being	Social func
25 mm	0.626	-1.340		-2.755	-1.326
	(2.434)	(1.832)		(3.754)	(5.758)
30 mm	3.653	6.019**	0.491	10.476*	9.420*
	(3.197)	(3.036)	(0.400)	(5.704)	(5.263)
Mean	49.806	56.840	7.847	87.108	92.083
N	390	390	189	390	390

Notes: Individuals with an aorta size of 20-34 mm. Using age, BMI and smoking history as covariates. Robust standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

IV. Myocardial Infarction, Antidepressants and Mortality

Acknowledgments: I would like to thank Claes Held, Erik Olsson, Ingeborg Waernbaum, Ulrika Wikman, Claudia Lissåker, Ronnie Pingel, Erik Grönqvist, Matz Dahlberg, Bodil Svennblad and Linuz Aggeborn for helpful discussions, comments and suggestions. I also thank Nikolaj Siersbæk and the participants at the NHESG 2015 conference in Uppsala who commented on an early version of this paper.

1 Introduction

Myocardial infarction (MI), more commonly known as heart attack, occurs when the blood cannot flow to a part of the heart muscle. MI patients have an increased risk of developing depression, and individuals who have experienced a MI have a higher mortality rate than the general population (e.g. Taylor et al. 2005; Thombs et al. 2006; Aso et al. 2011). Between 11-25 percent of the outpatients and as many as 35-70 percent of the inpatients with a heart failure meet the criteria for depression (Thomas et al. 2001; Joynt, Whellan, et al. 2004). Depression is associated with an increased risk of cardiac morbidity and mortality for individuals with an established coronary heart disease. Thus it is important to treat patients with a depression after an MI to alleviate the risk of a subsequent MI (Joynt, Whellan, et al. 2004). Antidepressants are a very common treatment for depression and anxiety today (Olfson and Marcus 2009).

However, studies suggest that treatment of depression has not decreased the risk of mortality (ENRICHD 2003; Joynt and O'Connor 2005). Some studies find that antidepressants lower the risk of mortality, while other studies find that both the old tricyclic antidepressants (TCA) and the newer Selective serotonin reuptake inhibitor (SSRI) antidepressants may increase the risk (Tata et al. 2005; Hamer et al. 2010; Noordam et al. 2016). The current consensus is that TCA should be avoided and that SSRI is relatively safe, but there is a need for large-scale studies using credible identification strategies.

In this study I estimate the effect of SSRI antidepressants on the risk of two-year mortality after the first MI, for patients who were prescribed SSRI within six months of the infarction. The identification strategy relies on nearest-neighbor matching on the propensity score on a rich set of covariates, such as patient history and MI severity. By matching on these variables, the aim is to create statistical twins who differs only in treatment status. The main contribution of this study is the use of a large and rich dataset of almost the complete population of MI patients in Sweden, which to my knowledge has not been utilized before, together with an identification strategy that allows for a causal interpretation of the results.

I use data from several Swedish population wide registers. The Swedish quality register for cardiac intensive care (RIKS-HIA and SEPHIA) from Swedeheart, the National Patient Register (in- and outpatient care), the prescribed drug register and the cause of death registry from The National Board of Health and Welfare. Individuals included in the data had their first MI between 2007 and 2011.

Results indicate no increased risk of mortality. The results are robust for various specifications. There are, however, no indication of a protective effect of using antidepressants.

The paper is organized as follows. In the next section I shortly discuss the medical background. Section 3 describes the data, followed by a discussion

of the empirical strategy and methodological considerations in section 4. Descriptive statistics is presented in section 5, and the main results in section 6. I discuss the results in section 7, and section 8 concludes the paper. Additional results, descriptive statistics, and covariate balance are presented in the appendix.

2 Medical background

As depression has been on the rise during the last decades in the western world, so has the use of antidepressants (Olfson and Marcus 2009; Reid and Barbui 2010). Depression is a mental disorder characterized by a persistent low mood. The individual often have low self-esteem, feelings of worthlessness, and have lost of interest in activities that he or she normally enjoys. Individuals who are depressed often have unusual loss or gain of weight and experience insomnia (American Psychiatric Association 2013). Estimates suggest that about nine out of ten individuals who commit suicide suffered from depression (Hawton et al. 2013). Antidepressants are, together with psychological therapy, the most common treatment for depression (Olfson and Marcus 2009). According to statistics from The National Board of Health and Welfare (Socialstyrelsen), almost 10 percent of the population in Sweden in 2015 was prescribed some kind of antidepressant.[1] However, the effects of antidepressants are a heavily debated topic. For example, some studies find that antidepressants lower the risk of mortality, while others find the opposite, or no effect at all (Dahlberg and Lundin 2005; Cipriani et al. 2005; Ludwig et al. 2009; Ghassemi et al. 2014).

Myocardial infarction occurs when the blood cannot flow to a part of the heart muscle. Common symptoms are chest pain, sweating, and dizziness (National Institutes of Health 2015). The mortality rate for MI patients is high (Aso et al. 2011). Depression and anxiety is common among patients recovering from a myocardial infarction (Ziegelstein 2001; Thombs et al. 2006; Williams 2011). 11-25 percent of the outpatients and 35-70 percent of the inpatients with heart failure meet the criteria for depression (Thomas et al. 2001; Joynt, Whellan, et al. 2004). It is well-known from the literature that depressed patients are at higher risk of mortality after a myocardial infarction, both through direct and indirect pathways (e.g. Frasure-Smith et al. 1995, Barth et al. 2004, Hare et al. 2013), as illustrated in Figure 1. Depression is associated with poor health behavior in general, including risk factors such as smoking and a poor diet. Joynt, Whellan, et al. (2004) concludes that it is important to treat patients with a depression after an MI to alleviate the risk of a subsequent MI.

[1] Note that antidepressants are not exclusively prescribed to individuals diagnosed with depression or anxiety.

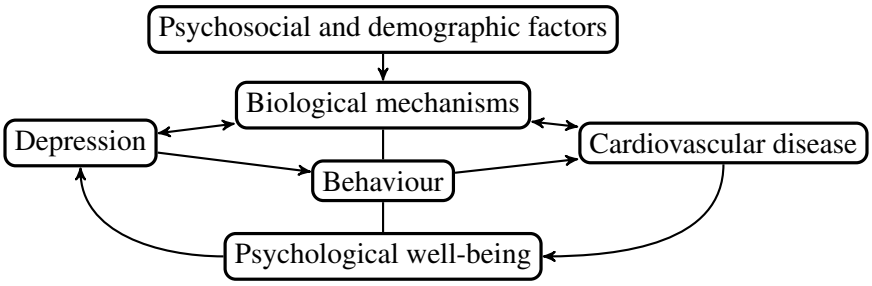


Figure 1. Relationship between cardiovascular disease and depression. Simplified and adapted from Hare et al. (2013).

However, some studies suggest that treatment of depression has not succeeded to decrease the risk of mortality (ENRICHD 2003; Joynt and O'Connor 2005). The effects of treating depression with antidepressants for MI patients are not clear. There is an ongoing discussion whether antidepressants increase or decrease cardiovascular mortality (Narayan and Stein 2009). Melle et al. (2007) does not find an improvement of post-MI long-term depression, or an improvement of the cardiac prognosis. There is a consensus in the literature that the earlier tricyclic antidepressants have cardiac effects, and is contraindicated for MI patients (Cohen et al. 2000; Joynt, Whellan, et al. 2004; Hamer et al. 2010). SSRI antidepressants are considered more safe, and Taylor et al. (2005) find that SSRI decrease mortality. However, Tata et al. (2005) suggest that both TCA and SSRI might increase the risk, whilst Noordam et al. (2016) find that the current use of antidepressants, regarding both TCA and SSRI, are associated with a lower risk of recurrent MI.

In a meta-analysis by Pizzi et al. (2011), the researchers find that the estimated effects differ between RCT:s and observational studies. In RCT:s they find no difference in mortality risk, while the observational studies indicate a decreased risk. The problem with RCT:s are that they often have small samples, short follow-up time, and it is not always clear if the results are externally valid. On the other hand, most observational studies do not use a credible identification strategy, and can only show associations. Thus, there is a need for more studies on SSRI use for MI patients that can utilize the population of MI patients while at the same time allow for a causal interpretation of the results.

3 Data

In this section I will describe the data available. Descriptives are shown later in section 5 and in the appendix.

The data consists of several Swedish population wide registers between the years 2006 and 2013. The Swedeheart registers (RIKS-HIA and SEPHIA) include almost all myocardial infarction patients in Sweden. RIKS-HIA is

nationwide and an almost complete register of all myocardial infarction patients in Sweden. About 90 percent of the hospitals are included. SEPHIA includes follow-up data for patients below 75 years age. The registers from The National Board of Health and Welfare include all prescribed drugs to the individuals, the in- and outpatient diagnoses, and the cause of death.[2]

To create a credible matching on health history, the year before the first MI is used as a measure of pre-MI health history (365 days). Therefore, I exclude all individuals with their first MI in 2006. I exclude individuals with their first MI in 2012 or 2013, to be able to have a follow-up period for individuals experiencing an MI in 2010 or 2011. Additionally, I exclude all patients who had an earlier MI, i.e., I only include patients experiencing their first MI during the years between 2007 and 2011. Some patients experience several MI:s during these years, but I focus only on the first MI. If an individual received antidepressants (of any kind) during the year before the MI he or she is excluded.[3] Individuals receiving antidepressants other than SSRI:s are excluded, as well as observations with missing values on the variables included in the most comprehensive specification.[4] After exclusions the sample consists of in total 38,319 patients.

The pre-MI health history is determined by using the prescribed drug register and the hospital in- and outpatient registers. Second level ATC is used for prescribed drugs, and first chapter level ICD for the diagnoses.[5] I use dummies indicating if the individual has received at least one drug or diagnosis within these classes.[6]

In addition to the pre-MI health variables, the individuals are matched on a large set of variables measuring the severity of the MI and other potential confounders. Variables include year of the MI, sex, employment status, diabetes,

[2]Primary care is not included in the in- and outpatient data, so only diagnoses a patient receives after visiting for example a hospital are included. The in- and outpatient registries include chapters F, I, J and N.

[3]2,036 individuals receive antidepressants within two years of the MI (the follow-up period) but not within six months (the treatment period). They are kept in the sample, in the control group, but the conclusions are no different if they are excluded. See section B in the appendix.

[4]There are 8,389 missing values for Left Ventricular Ejection Fraction. Instead of dropping these observations, I have created a missing value indicator which is included in the propensity score estimation.

[5]ATC stands for the Anatomical Therapeutic Chemical Classification System. There are 14 main categories, for example code C with include drugs for cardiovascular system and code J for antiinfectives for systemic use. The inpatient and outpatient registers are classified according to the International Classification of Diseases (ICD) codes in the corresponding way using the ICD-10 classification. ICD-10 consists of 22 main categories, for example E which includes endocrine, nutritional and metabolic diseases, and M which includes diseases of the musculoskeletal system and connective tissue. Both the ATC and ICD classifications are maintained by the World Health Organization (WHO).

[6]Which ATC and ICD chapters to include was decided after discussions with medical expertise. See Table A5.

Killip class etc. All variables are presented in section A, and the specifications in Table A1 in the appendix.

The outcome is measured as all-cause two-year mortality. The treatment is SSRI antidepressants (ATC code: N06AB), which is the most common class of antidepressants today, for individuals receiving SSRI within six months of the MI (183 days). As I discuss in the next section, the follow-up time differs between the treatment group and the control group. The treated group is followed two years (730 days) from the day of first treatment of SSRI. The untreated group is followed two years from a random day within the first six months of the MI.

4 Empirical strategy

In a randomized controlled trial (RCT) we randomize individuals to a treatment group and a control group. The randomization ensures that the treatment is independent of individual characteristics and self-selection, and the two groups are (in theory) balanced in both observable and unobservable covariates. The causal effect can be estimated by simply running the following regression:

$$Y_i = \alpha + \tau T_i + u_i, \tag{1}$$

where Y is the outcome and T is the treatment for individual i. τ is the estimated coefficient of interest. Without randomization, equation (1) is likely biased due to selection into treatment, i.e., T is correlated with the error term u.

With observational data it is not possible to *ex post* randomize individuals. In the case of antidepressant treatment we know that a selection into treatment exist; depression is included in the error term in equation (1), and the probability of receiving antidepressants is, obviously, higher for individuals who are depressed. Depression is correlated with worse general health, and therefore it is not possible to interpret the estimated coefficient as the causal effect of receiving antidepressants.

Regression Discontinuity Design (RDD), Instrumental Variables (IV) and Difference-in-Difference (DiD) are common methods and designs described in the econometric literature. Another approach to estimate causal effects is matching. Matching has an intuitive appeal: If we only compare individuals who are identical in all covariates, the only difference between them is the treatment status. One problem is that exact matching usually requires very large samples when we have many covariates. However, Rosenbaum and Rubin (1983) show that we do not need to have identical covariates to estimate a causal effect. It suffices with identical (or near-identical) *propensity score*. That is, what is important is that the likelihood of treatment for any given individual is identical.

Two fundamental assumptions are required for propensity score matching: Unconfoundedness and overlap.

Assumption 1. *Unconfoundedness:* $Y(0), Y(1) \perp\!\!\!\perp T|X$

The unconfoundedness assumption tells us that the potential outcomes are independent of the treatment assignment, conditional on a set of covariates X. Obviously, this is a strong assumption on the available data, and does not allow that any unobservable characteristics influence the treatment assignment and potential outcomes simultaneously.

Assumption 2. *Overlap:* $0 < P(T = 1|X) < 1$

The overlap assumption requires that all individuals with the same propensity score has a positive probability of being in either the treatment or control group. The combination of these two assumptions is called *strong ignorabililty*. If the assumptions are fulfilled, we are able to estimate a causal effect.[7]

Unfortunately it is not possible to test if the unconfoundeness assumption holds. We can never know if we have all relevant covariates in our model. Whether matching is a reasonable strategy is a question that must be answered on a case-by-case basis. If we have good knowledge of relevant covariates, matching can be used if we have the data. Myocardial infarction is such a case. The Swedish national quality registers RIKS-HIA and SEPHIA have information on more or less all relevant characteristics regarding the MI for almost the full population of patients. In addition to these variables, there is information on earlier health history, as well as age, sex, employment status etc. I argue that matching on these characteristics fulfills the unconfoundedness assumption.

In contrast, the overlap assumption can be tested. The large dataset of MI patients makes the overlap assumption fulfilled. I discard observations without an overlap using a caliper of 0.2 of the normalized SD of the propensity score, following the advice from Austin (2011) and others.

How can the individuals treatment status differ if strong ignorability is fulfilled? One reason is treatment cultures. After matching, the main source of variation likely stems from cultural practice; between counties and between individual doctors (see Table 2). For example, each county in Sweden has a Läkemedelskommitté (a pharmaceutical committee), which give recommendations of treatment for different diseases and patient groups. These recommendations are supposed to follow the best medical practice, but there are some differences in the recommendations between the committees. These differences can create a variation in the prescription of antidepressants, i.e., in

[7] See Imbens (2015) for a more technical discussion on the assumptions.

one county a depressed patient will receive antidepressants but had not in another county.

I estimate the average treatment effect of the treated (ATT),

$$\begin{aligned}\tau_{ATT} &= E(\tau|T=1) = E[Y(1)|T=1] - E[Y(0)|T=1] \\ &= \frac{1}{N_t}\sum_{i=1}^{N_t}\Big(Y_i(1) - Y_i(0)|T_i=1\Big).\end{aligned} \tag{2}$$

The ATT focus on the outcomes for whom the treatment is intended, i.e., the individuals that are eligible for antidepressant treatment, and estimates the average difference in outcome for those who received treatment compared with the counterfactual if they had not received treatment. Since the number of controls is large relative to the number of treated, the main specification use four neighbors for each treated individual, which will utilize more of the data and decrease the variance.[8]

Two-year all-cause mortality is used as the outcome. The follow-up time begins the day of treatment initiation for patients who receive SSRI within six months of their first MI. For untreated (i.e., individuals who do not receive SSRI within six months of the MI), the follow-up time starts from a random day within the six first months after the MI. This is a way of avoiding survival and immortal time bias which arises because treatment is not fixed at one point in time. Zhou et al. (2005), Suissa (2007) and Lévesque et al. (2010) discusses this problem. The bias arises because an individual could die before the treatment status is fixed, since I allow for a window of six months. The reason to allow for a treatment window is that developing depression, and receive medication, often take some time. Individuals who feel depressed after the MI may not yet identify the mood status as depression, but rather as a direct consequence of the MI, and will not search for help.[9] However, the strategy I use make the follow-up time shorter for some individuals who experience an MI in 2011. There are no reason to believe that this in itself bias the results, however. Figure 2 presents a measurement timeline of the variables.

The individuals are matched on several classes of variables, divided into year, SES, general health, patient history and MI measures. The patient history is measured within the year before the MI. The variables used is presented in section A in the appendix. Figure 3 illustrate the identification problem and the reasoning behind the choice of variables to match on. The general health and socioeconomic status (SES) possibly affect the likelihood of treatment, the mediators and the outcome. Thus, a credible matching must in some way control for the health of the individual (this includes sex, age, and, for example,

[8] Specifications using one-to-one matching are presented in the appendix.

[9] Very few individuals receive antidepressants within the first month of the MI, although the literature suggest that depression is common among MI patients. 2,672 individuals die within six months of the MI, compared with 4,769 individuals within two years, about 56 percent of the deaths.

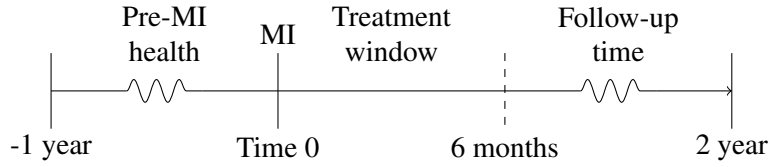


Figure 2. Measurement timeline of covariates, treatment and outcome.

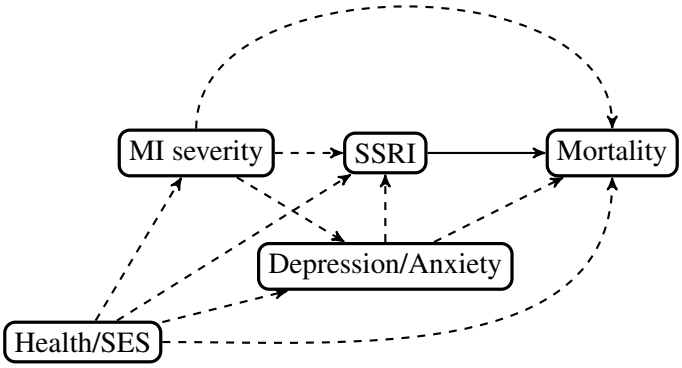


Figure 3. Health and MI affects both the treatment and the outcome. The solid line is the causal effect of interest.

smoking). In the same way, the severity of the MI is likely affecting depression, the likelihood of receiving antidepressants and, obviously, mortality, and must also be controlled for in the matching. Depression affects both the likelihood of receiving treatment and the outcome. Matching on these classes of variables creates statistical twins who differs only in one important aspect, the treatment with SSRI antidepressants.

5 Descriptive statistics

In this section I present general descriptive statistics. Due to the many variables in the data most tables are presented in the appendix.

Table 1 show the share of individuals who received a depression or anxiety diagnosis, and whether they were treated with SSRI antidepressants within six months of the MI. The table also presents the two-year mortality in the sample. It stands clear that relatively few individuals received a diagnosis. About 3 percent received SSRI antidepressants, which is three times the share of those who received a diagnosis. These findings stand somewhat in contrast to the claims in the literature that more than 10 percent of the patients with heart failure are depressed (Joynt, Whellan, et al. 2004). Some of those who are prescribed SSRI antidepressants may receive SSRI for non-obvious reasons,

Table 1. *Diagnosis, antidepressants, and mortality*

	All	Treated	Untreated	Diff
Diagnosis:				
Depression/anxiety	0.01	0.18	0.00	0.18***
	(0.10)	(0.38)	(0.07)	(0.01)
Outcome:				
Mortality	0.13	0.17	0.13	0.04***
	(0.34)	(0.38)	(0.34)	(0.01)
Treatment:				
SSRI	0.03			
	(0.17)			
Observations	38,319	1,101	37,218	38,319

Notes: Diagnosis and treatment are measured within six months of the MI. Outcome is measured as two-year mortality from SSRI treatment for the treated group, and from a random point in day for the untreated group.

or, possibly, are prescribed SSRI due to a mild depression that is not recorded in the registers.[10] As can be seen in the table, 13 percent of the sample died within two years, but of the treated patients, the two-year mortality was about 17 percent.

One of the sources to the random component in antidepressant prescription is differences in prescription cultures between counties. Table 2 presents the percent of SSRI, depression/antiety, mortality, and the mean value of the answer to the anxiety question in the EQ-5D questionnaire (a value between 1 and 3, where 3 is the worst health), divided on county.[11] For example, we can see that Värmland has a low share of depression and anxiety, but a relatively high prescription of SSRI. If we compare Värmland and Dalarna, the difference in SSRI prescriptions and depression/anxiety diagnoses does not seem to be reflected in the EQ-5D questionnaire. Assuming that the patients are more or less the same between the counties, this reflects differences in the prescription culture, which is exploited by the matching method.

Table 3 presents the five most common causes of death in the sample. The four most common causes are problems with the heart, and not causes directly connected to psychological ill-being.

As a back-on-the-envelope calculation, the average time on SSRI for the treated group is about 50 months for those who experienced an MI in 2008,

[10] Note that the individuals may have received a diagnosis in the primary care which is not included in the data.

[11] About 40 percent of the sample have answered the EQ-5D questionnaire. The questionnaire is answered by patients below age 75 about two months after the MI.

Table 2. *SSRI, depression and mortality by county*

	SSRI	Depressed/anxiety	Mortality	EQ-5D: Anxiety
Gotland	4.23	0.47	8.92	1.32
Jämtland	4.04	1.10	14.86	1.39
Västmanland	3.41	1.03	11.58	1.21
Södermanland	3.52	1.52	12.02	1.32
Blekinge	3.27	1.57	14.25	1.32
Gävleborg	3.31	0.52	12.03	1.40
Västra Götaland	3.31	1.03	14.09	1.41
Skåne	3.02	1.16	11.65	1.35
Värmland	3.09	0.50	12.46	1.32
Kronoberg	3.00	1.61	15.34	1.33
Östergötland	2.87	1.37	13.04	1.39
Uppsala	2.81	0.70	8.70	1.40
Örebro	2.94	0.52	12.28	1.34
Jönköping	2.58	0.53	12.62	1.35
Stockholm	2.42	1.15	12.41	1.44
Kalmar	2.70	0.82	13.32	1.40
Norrbotten	2.24	0.58	13.34	1.22
Dalarna	2.23	0.86	10.82	1.33
Västerbotten	2.59	0.81	8.33	1.27
Halland	1.94	0.90	13.15	1.25
Västernorrland	1.64	0.52	12.52	1.27

Notes: Percent of individuals who receive SSRI, depression/anxiety diagnosis and dies within two years of MI, by county. Sorted by SSRI prescription.

Table 3. *Cause of death*

	ICD	Share	Frequency	Description
1	I219	36.43	1,847	Acute myocardial infarction, unspecified
2	I258	6.67	338	Other forms of chronic ischaemic heart disease
3	I259	5.21	264	Chronic ischaemic heart disease, unspecified
4	I251	3.23	164	Atherosclerotic heart disease
5	C349	2.70	137	Malignant neoplasm of bronchus and lung
Other	...	45.76	2,320	522 different codes
Total			5,070	

Notes: ICD-10.

and about 23 months for those with an MI in 2011.[12] Thus, SSRI treatment is in general a longtime treatment.

6 Results

In this section I present the results from the propensity score matching. I estimate four different specifications. In the first column, I match on year of MI, SES and general health variables. In the second column, I include medical variables from the prescribed drugs registry and in- and outpatient history. The third column instead match on variables related to the MI. The last column include all variables, and is the preferred specification. The variables included in each specification can be seen in Table A1 in the appendix, and the propensity scores are estimated by logistic regression models shown in Table A11. The appendix also include results for ordinary OLS regressions using the same specifications, in section B.

Table 4 presents the main results. I estimate the ATT using the four nearest neighbors with a caliper of 0.2 of the normalized SD of the PS.[13] There is a clear pattern in the results, and the estimate shrinks for each specification.

In terms of balance, the third specification is much better than the two first specifications (see the diagnostics section in appendix D). However, the preferred specification is the fourth, which includes all variables available (which should result in the best prediction on the individuals health status), and have a good balance in the covariates. All point estimates are positive, but only the first specification is statistically significant (at the 10 percent level). The three first specifications all have omitted variables, which is likely to bias the estimates upwards, since individuals who receive SSRI have worse health than individuals who do not. However, it might be the case that the most depressed individuals do not receive SSRI, which would attenuate the bias downwards (see the discussion in section 7).

7 Discussion

The results in Table 4 suggest that there is no increase in the risk of two-year mortality for individuals who has experienced a myocardial infarction and receive SSRI. The preferred specification indicate an increase of 0.9 percentage points, but is far from statistically significant. Since the literature suggest that depression and anxiety is common worldwide among patients with cardiovascular disease, this is a positive finding, as many of those patients receive SSRI

[12] Calculated by simply taking the difference between the first and last occurrence in the data.

[13] I first estimate the PS using logistic regression, then trim the sample according to the preferred caliper, and run nearest-neighbor matching on the estimated PS. The difference in sample size is due to the trimming, i.e., lack of overlap in the PS. The trimming is shown in section C. In the appendix I also show the estimates for matching on only one neighbor in Table A7.

Table 4. *Antidepressants and mortality*

	(1)	(2)	(3)	(4)
SSRI	0.024*	0.022	0.020	0.009
	(0.013)	(0.013)	(0.014)	(0.014)
Year/SES/Health	Yes	Yes	Yes	Yes
Medical	No	Yes	No	Yes
MI measures	No	No	Yes	Yes
Observations	38,307	38,291	38,307	38,291
Treated	1,099	1,101	1,099	1,096

Notes: Standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment model: logit. 4 Nearest-neighbor matching on propensity score. Caliper: ± 0.2 normalized sd(PS). Matching with replacement. ATT. Dependent variable: Mortality within two years of treatment start for the treated group, and two year from a random day within six months of MI for the control group. Treatment: SSRI antidepressants within six months of first MI.

antidepressants. This study does, however, not address the question whether antidepressants are effective means against depression, and there is no indication of a *lower* mortality rate due to the treatment. Whether antidepressants should be prescribed for MI patients must therefore depend on their effectiveness on treating depression and anxiety and potential side effects.

Is the chosen empirical strategy applicable in this case? Could the results depend on the choice of method? As with all matching methods the strategy used in this paper relies on observable characteristics. Economists are often worried that there are important unobservable characteristics, which is why methods such as IV, DiD and RDD exploiting exogenous variation are commonly used. While these methods might be the "gold standard" in economics, the method of choice must be judged on a case-by-case basis. Matching methods can be used if we have good knowledge of the possible confounding factors, and have access to a rich dataset. I argue that it is the case here.

There are some important limitations in this study. Relatively few individuals in the sample actually receive a depression or anxiety diagnosis, and since data from the primary care is lacking it is likely that some individuals have a diagnosis which is not seen in the data. Of 38,319 patients, only 371 individuals receive a diagnosis within six months of the MI. During the same period, 1,101 patients are prescribed SSRI antidepressants. Only 18 percent of the individuals receiving treatment have a diagnosis corresponding to the prescription.[14] The most likely reason is the lack of data from the primary

[14]The share of SSRI among patients with depression/anxiety is about 54 percent.

care, but other explanations could be off-label prescription[15], or that the general practitioner does not think that it is necessary to do an ordinary examination before prescribing the drug. Since myocardial infarction is linked to depression, some general practitioners perhaps intervene on early signs of depression. There is evidence that such prescription is becoming more common (Mojtabai and Olfson 2011).

The possibility of omitted variables creates problems for the matching. Ideally, all patients should be identical except for the treatment status. The propensity score matching reduces the many-dimensional problem to a one-dimensional problem, but it cannot solve the problem with unobservable characteristics. In this specific case we can be worried that patients receiving SSRI:s without a diagnosis have worse general health than patients without SSRI:s and no diagnosis, a problem the matching cannot solve. This would create an upward bias, and result in significant positive effects on mortality of SSRI:s. A second problem is that not all depressed individuals receive SSRI:s, as the worst cases are likely to not go to a general practitioner, which could create a bias downwards. Thus, the bias may go in both directions. On the other hand, it is not obvious what this worse health could be; the data is rich on health variables, and since mortality is such a severe outcome it is not unreasonable to assume that the variables used in the matching can take this into account.

8 Conclusions

Using a rich dataset on 38,319 first time myocardial infarction patients in Sweden during 2007-2011, I estimate the causal effect of the use of SSRI antidepressants on mortality using a propensity score matching. I find no evidence that use of SSRI increase the likelihood of mortality within two years.

The individuals are matched on several categories of variables, such as socioeconomic status, earlier health history and the severity of the MI. The most common cause of death is another myocardial infarction or other heart failures.

Matching can only be done on observable characteristics, and there may be unobservable characteristics which could bias the results. Individuals who are depressed or having anxiety are likely to have worse health than individuals who do not. The worst cases, however, may not use SSRI, since it is possible that they do not receive care. The balance tests indicate that the matching is able to create "statistical twins" on the observed characteristics, and it is not obvious in which direction the potential bias of omitted variables may go. While SSRI does not seem to increase the likelihood of mortality, this study cannot answer the question whether SSRI antidepressants are effective means against depression for these patients.

[15]Prescribing a pharmaceutical drug for an unapproved indication or patient group.

References

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5).* American Psychiatric Publishing.

Aso, Shin-ichi, Hiroshi Imamura, Yukio Sekiguchi, Tomomi Iwashita, Ryosuke Hirano, Uichi Ikeda, and Kazufumi Okamoto. 2011. "Incidence and mortality of acute myocardial infarction. A population-based study including patients with out-of-hospital cardiac arrest." *International Heart Journal* 52 (4): 197–202.

Austin, Peter C. 2011. "Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies." *Pharmaceutical Statistics* 10 (2): 150–161.

Barth, Jurgen, Martina Schumacher, and Christoph Herrmann-Lingen. 2004. "Depression as a risk factor for mortality in patients with coronary heart disease: A meta-analysis." *Psychosomatic Medicine* 66 (6): 802–813.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some practical guidance for the implementation of propensity score matching." *Journal of Economic Surveys* 22 (1): 31–72.

Cipriani, Andrea, Corrado Barbui, and John R Geddes. 2005. "Suicide, depression, and antidepressants." *British Medical Journal* 330 (7488): 373–374.

Cohen, Hillel W, Geoffrey Gibson, and Michael H Alderman. 2000. "Excess risk of myocardial infarction in patients treated with antidepressant medications: Association with use of tricyclic agents." *The American Journal of Medicine* 108 (1): 2–8.

Dahlberg, Matz, and Douglas Lundin. 2005. "Antidepressants and the suicide rate: Is there really a connection?" In *Substance Use: Individual Behaviour, Social Interactions, Markets and Politics,* edited by Björn Lindgren and Michael Grossman, 16:121–141. Emerald Group Publishing Limited.

ENRICHD. 2003. "Effects of treating depression and low perceived social support on clinical events after myocardial infarction: The enhancing recovery in coronary heart disease patients (ENRICHD) randomized trial." *JAMA* 289 (23): 3106–3116.

Frasure-Smith, Nancy, Francois Lesperance, and Mario Talajic. 1995. "The impact of negative emotions on prognosis following myocardial infarction: Is it more than depression?" *Health Psychology* 14 (5): 388–398.

Ghassemi, Marzyeh, John Marshall, Nakul Singh, David J Stone, and Leo Anthony Celi. 2014. "Leveraging a critical care database: Selective serotonin reuptake inhibitor use prior to ICU admission is associated with increased hospital mortality." *Chest* 145 (4): 745–752.

Hamer, Mark, G David Batty, Adrie Seldenrijk, and Mika Kivimaki. 2010. "Antidepressant medication use and future risk of cardiovascular disease: The Scottish health survey." *European Heart Journal* 32.

Hare, David L, Samia R Toukhsati, Peter Johansson, and Tiny Jaarsma. 2013. "Depression and cardiovascular disease: A clinical review." *European Heart Journal* 35:1365–1372.

Hawton, Keith, Carolina Casañas i Comabella, Camilla Haw, and Kate Saunders. 2013. "Risk factors for suicide in individuals with depression: A systematic review." *Journal of Affective Disorders* 147 (1–3): 17–28.

Imbens, Guido W. 2015. "Matching methods in practice: Three examples." *Journal of Human Resources* 50 (2): 373–419.

Joynt, Karen E, and Christopher M O'Connor. 2005. "Lessons from SADHART, ENRICHD, and other trials." *Psychosomatic Medicine* 67 (Suppl 1): S63–S66.

Joynt, Karen E, David J Whellan, and Christopher M O'connor. 2004. "Why is depression bad for the failing heart? A review of the mechanistic relationship between depression and heart failure." *Journal of Cardiac Failure* 10 (3): 258–271.

Lévesque, Linda E, James A Hanley, Abbas Kezouh, and Samy Suissa. 2010. "Problem of immortal time bias in cohort studies: Example using statins for preventing progression of diabetes." *British Medical Journal* 340:b5087.

Ludwig, Jens, Dave E Marcotte, and Karen Norberg. 2009. "Anti-depressants and suicide." *Journal of Health Economics* 28 (3): 659–676.

Melle, Joost P van, Peter De Jonge, Adriaan Honig, Aart H Schene, Astrid M G Kuyper, Harry J G M Crijns, Annique Schins, Dorien Tulner, Maarten P van den Berg, and Johan Ormel. 2007. "Effects of antidepressant treatment following myocardial infarction." *The British Journal of Psychiatry* 190 (6): 460–466.

Mojtabai, Ramin, and Mark Olfson. 2011. "Proportion of antidepressants prescribed without a psychiatric diagnosis is growing." *Health Affairs* 30 (8): 1434–1442.

Narayan, Sanjiv M, and Murray B Stein. 2009. "Do depression or antidepressants increase cardiovascular mortality? The absence of proof might be more important than the proof of absence." *Journal of the American College of Cardiology* 53 (11): 959–961".

National Institutes of Health. 2015. "What is a heart attack?" Accessed September 16, 2016. `http://www.nhlbi.nih.gov/health/health-topics/topics/heartattack`.

Noordam, Raymond, Nikkie Aarts, Maarten J G Leening, Henning Tiemeier, Oscar H Franco, Albert Hofman, Bruno H Stricker, and Loes E Visser. 2016. "Use of antidepressants and the risk of myocardial infarction in middle-aged and older adults: A matched case-control study." *European Journal of Clinical Pharmacology* 72 (2): 211–218.

Olfson, Mark, and Steven C Marcus. 2009. "National patterns in antidepressant medication treatment." *Archives of General Psychiatry* 66 (8): 848–856.

Pizzi, Carmine, Anne Wilhelmina Saskia Rutjes, Grazia Maria Costa, Fiorella Fontana, Andrea Mezzetti, and Lamberto Manzoli. 2011. "Meta-analysis of selective serotonin reuptake inhibitors in patients with depression and coronary heart disease." *The American Journal of Cardiology* 107 (7): 972–979.

Reid, Steven, and Corrado Barbui. 2010. "Long term treatment of depression with selective serotonin reuptake inhibitors and newer antidepressants." *British Medical Journal* 340.

Rosenbaum, Paul R, and Donald B Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70 (1): 41–55.

Rubin, Donald B, and Neal Thomas. 1996. "Matching using estimated propensity scores: Relating theory to practice." *Biometrics* 52 (1): 249–264.

Suissa, Samy. 2007. "Immortal time bias in observational studies of drug effects." *Pharmacoepidemiology and Drug Safety* 16 (3): 241–249.

Tata, LJ, J West, C Smith, P Farrington, T Card, L Smeeth, and R Hubbard. 2005. "General population based study of the impact of tricyclic and selective serotonin reuptake inhibitor antidepressants on the risk of acute myocardial infarction." *Heart* 91 (4): 465–471.

Taylor, C Barr, Marston E Youngblood, Diane Catellier, Richard C Veith, Robert M Carney, Matthew M Burg, Peter G Kaufmann, John Shuster, Thomas Mellman, James A Blumenthal, et al. 2005. "Effects of antidepressant medication on morbidity and mortality in depressed patients after myocardial infarction." *Archives of General Psychiatry* 62 (7): 792–798.

Thomas, P, G Kavan Michael, N Elasser Gary, D Pharm, and J Barone Eugene. 2001. "Assessment and treatment of depression following myocardial infarction." *American Family Physician* 64 (4): 641–648.

Thombs, Brett D, Eric B Bass, Daniel E Ford, Kerry J Stewart, Konstantinos K Tsilidis, Udita Patel, James A Fauerbach, David E Bush, and Roy C Ziegelstein. 2006. "Prevalence of depression in survivors of acute myocardial infarction." *Journal of General Internal Medicine* 21 (1): 30–38.

Williams, Redford B. 2011. "Depression after heart attack." *Circulation* 123 (25): e639–e640.

Zhou, Zheng, Elham Rahme, Michal Abrahamowicz, and Louise Pilote. 2005. "Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: A comparison of methods." *American Journal of Epidemiology* 162 (10): 1016–1023.

Ziegelstein, Roy C. 2001. "Depression in patients recovering from a myocardial infarction." *JAMA* 286 (13): 1621–1627.

Appendix

The appendix includes a more detailed description of the variables included the specifications, the propensity score estimations, additional results, and several descriptive tables showing the balance before and after trimming and matching of the sample. I also present figures showing the common support, and love plots showing the balance before and after matching.

A Variable selection

There are different traditions regarding how to select variables to include in matching models. In general, there are two different strains: One focus on selection based on theoretical arguments, the other is more data-driven. Since the goal of the propensity score is to find good balance between the treated and untreated groups, it is not obvious that one is better than another. In fact, as long as the researcher does not see how the variables included affect the outcome, there is no (or very little) danger that he or she selects variables that give the "preferred" outcome.

Another result in the literature is that the bias of including "too many" variables or variables that are unrelated to the treatment and outcome is less than the bias of omitting variables that are important. Thus, in matching it is quite common to include many variables. Only variables measured after treatment or that we know are only related to the treatment and not the outcome should unambiguously be avoided. More variables can, however, make the estimations less precise (Rubin and Thomas 1996; Caliendo and Kopeinig 2008).

My approach has been mostly theoretically driven, and the variables are selected after discussions with medical and statistical expertise. I run four separate specifications. Each specification include and/or remove a class of variables, as shown in Table A1. The first specification includes the year of MI, SES and general health variables. The second specification include earlier health history in the form of prescribed drugs and in- and outpatient data the year before the MI. The third specification removes the health history variables, but instead include variables related to the MI. The fourth, and preferred, specification includes all variables.

Most variables are dummy variables. Age, heart rate and systolic blood pressure at admission are discrete. The ATC and ICD variables take the value 1 if the individual received a diagnosis (ICD) or pharmaceutical drug (ATC) at least once during the year before the MI. See Table A5 for the specific ATC and ICD codes included.

Table A1. *Variables in specifications*

	(1)	(2)	(3)	(4)
Sex	x	x	x	x
Age	x	x	x	x
Employment status	x	x	x	x
Year	x	x	x	x
Smoking status	x	x	x	x
Diabetes	x	x	x	x
Hypertension	x	x	x	x
History of stroke	x	x	x	x
Previous PCI	x	x	x	x
Depression/anxiety	x	x	x	x
ATC		x		x
ICD		x		x
ECG rhythm			x	x
Systolic blood pressure at admission			x	x
Heart rate at admission			x	x
History of CHF			x	x
ECG QRS			x	x
Killip class			x	x
Reperfusion treatment			x	x
Bleeding under care			x	x
CPR or defib			x	x
Mechanical complication			x	x
New atrial fibriliation			x	x
Reinfarct during care			x	x
Left Ventricular Ejection Fraction			x	x
AV block			x	x
Beta blockers at discharge			x	x
Statins at discharge			x	x
Nitrates at discharge			x	x
ACE inhibitors or Angiotensin II at discharge			x	x
Other lipid lowering agents at discharge			x	x
Other antiplatelet at discharge			x	x

Notes: Included ATC codes: B01, C01, C02, C05, C07, C09, C10, N05, N06, N07. Included ICD codes: F, I, J, N. Killip class is constructed using pulmonary rales status and cardiogenic shock.

I also present four tables (Table A2-A5) with descriptives of all variables included in the specifications, before trimming and matching.

B Additional results

This subsection contains a discussion on additional results, such as OLS regressions and matching results using only one nearest-neighbor. I use the same specifications as in the main results.

The point estimates for the OLS regressions in Table A6 are in statistical terms not different from the estimates in the main results. In fact, the standard errors are smaller, and, except for the last specification, the estimates smaller. The OLS is able to use the full sample, compared with the matching that only use the number of treated individuals plus one or four controls. OLS, however, does not ensure that the treated and untreated individuals are comparable (i.e., common support) and may overweight observations with no overlap in the data. In addition, OLS requires a functional form to be specified, which matching does not. However, these potential sources of bias does not seem to matter much for the conclusions in this case.

Table A7 show the results using only one nearest-neighbor. In comparison to the main results, the standard errors are somewhat larger, and both the first and third specification have statistically significant results. However, even though the estimates are some what larger, the conclusion is not different for the preferred fourth specification.

In the main specifications individuals who receive antidepressants (not only SSRI) after the treatment window of six months, but within the two year follow-up, are kept (in total 2,036 observations) and remains in the control group. One could be worried that keeping these individuals in the sample would attenuate the estimates. In Table A8 the observations are dropped, and all specifications are run again. As can be seen in the table the estimates are actually smaller compared with the main specifications, with no statistically significant result.

Table A9 show specifications with two other outcomes. The first row simply use the two year follow-up time from the day of the MI for both the treated and the untreated group, and the second row use the day of SSRI initiation for the treated group, but the day of the MI for the untreated group. The second row have three statistically significant estimates. However, the fourth specification is not significant.

In addition to these tests, I have run each specification without using a caliper and not trimming the data. The estimates and standard errors are only insignificantly different from the main results. I have also run the specifications without replacement with one nearest neighbor. The estimates are in general somewhat smaller than in the main specifications. These results are available upon request.

Table A2. *Descriptives: SES and year covariates*

General covariates	All	Treated	Untreated	Diff
Female	0.32	0.44	0.32	0.12***
	(0.47)	(0.50)	(0.47)	(0.02)
Age				
<50 years	0.06	0.07	0.06	0.01
	(0.24)	(0.26)	(0.24)	(0.01)
50-75 years	0.60	0.54	0.60	-0.06***
	(0.49)	(0.50)	(0.49)	(0.02)
>75 years	0.34	0.39	0.34	0.05***
	(0.47)	(0.49)	(0.47)	(0.01)
Employment status				
Other	0.02	0.02	0.02	0.00
	(0.14)	(0.15)	(0.14)	(0.00)
Employed	0.30	0.28	0.30	-0.02
	(0.46)	(0.45)	(0.46)	(0.01)
Retired	0.68	0.69	0.68	0.02
	(0.47)	(0.46)	(0.47)	(0.01)
Year				
2007	0.20	0.23	0.20	0.02*
	(0.40)	(0.42)	(0.40)	(0.01)
2008	0.19	0.21	0.19	0.02
	(0.40)	(0.41)	(0.39)	(0.01)
2009	0.19	0.19	0.19	-0.00
	(0.39)	(0.39)	(0.39)	(0.01)
2010	0.20	0.18	0.20	-0.02*
	(0.40)	(0.39)	(0.40)	(0.01)
2011	0.21	0.19	0.21	-0.02
	(0.41)	(0.40)	(0.41)	(0.01)
Observations	38,319	1,101	37,218	38,319

Notes: Age is included as a discrete variable when estimating the propensity score.

Table A3. *Descriptives: Health variables measured at MI*

General health	All	Treated	Untreated	Diff
Smoking status				
Never smoker	0.44	0.42	0.44	-0.02
	(0.50)	(0.49)	(0.50)	(0.02)
Former smoker	0.33	0.26	0.33	-0.07***
	(0.47)	(0.44)	(0.47)	(0.01)
Current smoker	0.23	0.32	0.23	0.09***
	(0.42)	(0.47)	(0.42)	(0.01)
Diabetes	0.15	0.17	0.15	0.02*
	(0.36)	(0.38)	(0.36)	(0.01)
Hypertension	0.44	0.47	0.44	0.03*
	(0.50)	(0.50)	(0.50)	(0.02)
History of stroke	0.94	0.91	0.94	-0.03***
	(0.24)	(0.29)	(0.24)	(0.01)
Previous PCI	0.03	0.02	0.03	-0.00
	(0.16)	(0.15)	(0.16)	(0.00)
Depression				
Within year before	0.00	0.01	0.00	0.01***
	(0.04)	(0.09)	(0.04)	(0.00)
Within six months after	0.01	0.18	0.00	0.18***
	(0.10)	(0.38)	(0.07)	(0.01)
Observations	38,319	1,101	37,218	38,319

Notes: Both "Depression within year before" and "Depression within six months after" [the MI] are included as covariates when estimating the propensity score.

Table A4. *Descriptives: MI severity measures*

MI measures	All	Treated	Untreated	Diff
ECG rhythm				
Other	0.02	0.03	0.02	0.01
	(0.15)	(0.17)	(0.15)	(0.01)
Atrial fibrillation	0.09	0.09	0.09	0.01
	(0.28)	(0.29)	(0.28)	(0.01)
Sinus	0.89	0.88	0.89	-0.02
	(0.31)	(0.33)	(0.31)	(0.01)
Systolic blood pressure at admission	148.60	148.35	148.60	-0.25
	(28.87)	(29.26)	(28.86)	(0.89)
Heart rate at admission	79.78	82.32	79.71	2.61***
	(22.09)	(22.16)	(22.09)	(0.68)
History of CHF	0.04	0.04	0.04	0.01
	(0.19)	(0.21)	(0.19)	(0.01)
ECG QRS	0.31	0.35	0.31	0.04**
	(0.46)	(0.48)	(0.46)	(0.01)
Killip class	0.10	0.11	0.10	0.01
	(0.30)	(0.31)	(0.30)	(0.01)
Reperfusion treatment	0.34	0.33	0.35	-0.01
	(0.48)	(0.47)	(0.48)	(0.01)
Bleeding under care	0.01	0.02	0.01	0.01**
	(0.11)	(0.15)	(0.11)	(0.00)
CPR or defib	0.03	0.02	0.03	-0.01
	(0.16)	(0.14)	(0.16)	(0.00)
Mechanical complication	0.00	0.00	0.00	-0.00
	(0.06)	(0.05)	(0.06)	(0.00)
New atrial fibriliation	0.04	0.04	0.04	-0.00
	(0.20)	(0.20)	(0.20)	(0.01)
Reinfarct during care	0.01	0.01	0.01	0.00
	(0.09)	(0.09)	(0.09)	(0.00)
Left Ventricular Ejection Fraction	0.32	0.34	0.32	0.03*
	(0.46)	(0.47)	(0.46)	(0.01)
(LVEF: Missing)	0.22	0.23	0.22	0.01
	(0.41)	(0.42)	(0.41)	(0.01)
AV block	0.02	0.01	0.02	-0.00
	(0.13)	(0.12)	(0.13)	(0.00)
Beta blockers at discharge	0.89	0.89	0.88	0.01
	(0.32)	(0.31)	(0.32)	(0.01)
Statins at discharge	0.87	0.86	0.87	-0.02
	(0.33)	(0.35)	(0.33)	(0.01)
Nitrates at discharge	0.10	0.12	0.10	0.02**
	(0.30)	(0.32)	(0.30)	(0.01)
ACE inhibitors or Angiotensin II at discharge	0.73	0.73	0.73	-0.00
	(0.44)	(0.44)	(0.44)	(0.01)
Other lipid lowering agents at discharge	0.01	0.01	0.01	-0.00
	(0.10)	(0.09)	(0.10)	(0.00)
Other antiplatelet at discharge	0.81	0.76	0.81	-0.05***
	(0.39)	(0.43)	(0.39)	(0.01)
Observations	38,319	1,101	37,218	38,319

Notes: Systolic blood pressure and heart rate at admission are discrete variables. 22 percent of the sample have missing values for Left Ventricular Ejection Fraction, so a missing value indicator are included when estimating the propensity score.

Table A5. *Descriptives: Patient history*

Medical covariates	All	Treated	Untreated	Diff
ATC				
B01	0.31	0.35	0.30	0.05***
	(0.46)	(0.48)	(0.46)	(0.01)
C01	0.15	0.17	0.15	0.02
	(0.36)	(0.37)	(0.36)	(0.01)
C02	0.01	0.01	0.01	0.00
	(0.10)	(0.11)	(0.10)	(0.00)
C05	0.02	0.02	0.02	0.00
	(0.14)	(0.15)	(0.14)	(0.00)
C07	0.30	0.33	0.30	0.04***
	(0.46)	(0.47)	(0.46)	(0.01)
C09	0.31	0.35	0.31	0.05***
	(0.46)	(0.48)	(0.46)	(0.01)
C10	0.21	0.23	0.21	0.02
	(0.41)	(0.42)	(0.41)	(0.01)
N05	0.17	0.33	0.16	0.16***
	(0.37)	(0.47)	(0.37)	(0.01)
N06	0.00	0.00	0.00	0.00
	(0.02)	(0.03)	(0.02)	(0.00)
N07	0.01	0.02	0.01	0.01**
	(0.10)	(0.13)	(0.10)	(0.00)
ICD				
F	0.01	0.03	0.01	0.02***
	(0.09)	(0.16)	(0.09)	(0.00)
I	0.05	0.06	0.05	0.01
	(0.21)	(0.23)	(0.21)	(0.01)
J	0.00	0.00	0.00	0.00
	(0.06)	(0.06)	(0.06)	(0.00)
N	0.00	0.00	0.00	0.00
	(0.05)	(0.06)	(0.05)	(0.00)
Observations	38,319	1,101	37,218	38,319

Notes: Pharmaceutical drug history and in- and outpatient history (excluding primary care) the year before MI. Dummy variables taking the value 1 if the individual received a drug or diagnosis within the category at least once during this time period.

Table A6. *OLS: Antidepressants and mortality*

	(1)	(2)	(3)	(4)
SSRI	0.020*	0.012	0.015	0.011
	(0.011)	(0.011)	(0.011)	(0.011)
Year/SES/Health	Yes	Yes	Yes	Yes
Medical	No	Yes	No	Yes
MI measures	No	No	Yes	Yes
Observations	38,307	38,291	38,307	38,291

Notes: Robust standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. OLS. Dependent variable: Mortality within two years of treatment start for the treated group, and two year from a random day within six months of MI for the control group. Treatment: SSRI antidepressants within six months of first MI.

Table A7. *Antidepressants and mortality (1 NN)*

	(1)	(2)	(3)	(4)
SSRI	0.030**	0.027	0.030*	0.012
	(0.014)	(0.016)	(0.016)	(0.017)
Year/SES/Health	Yes	Yes	Yes	Yes
Medical	No	Yes	No	Yes
MI measures	No	No	Yes	Yes
Observations	38,307	38,291	38,307	38,291
Treated	1,099	1,101	1,099	1,096

Notes: Standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment model: logit. 1 Nearest-neighbor matching on propensity score. Caliper: ± 0.2 sd(PS). Matching with replacement. ATT. Dependent variable: Mortality within two years of treatment start for the treated group, and two year from a random day within six months of MI for the control group. Treatment: SSRI antidepressants within six months of first MI.

Table A8. *Antidepressants and mortality (subsample)*

	(1)	(2)	(3)	(4)
SSRI	0.021	0.010	0.013	0.006
	(0.013)	(0.014)	(0.014)	(0.014)
Year/SES/Health	Yes	Yes	Yes	Yes
Medical	No	Yes	No	Yes
MI measures	No	No	Yes	Yes
Observations	36,272	36,262	36,268	36,259
Treated	1,098	1,101	1,096	1,097

Notes: Standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment model: logit. 4 Nearest-neighbor matching on propensity score. Caliper: ± 0.2 sd(PS). Matching with replacement. ATT. Dependent variable: Mortality within two years of treatment start for the treated group, and two year from a random day within six months of MI for the control group. Treatment: SSRI antidepressants within six months of first MI.

Table A9. *Antidepressants and mortality (other outcomes)*

Outcome: Within two years	(1)	(2)	(3)	(4)
... of MI	0.020	0.017	0.015	0.004
	(0.013)	(0.013)	(0.013)	(0.013)
... of SSRI treatment	0.032**	0.029**	0.027**	0.017
	(0.013)	(0.013)	(0.013)	(0.014)
Year/SES/Health	Yes	Yes	Yes	Yes
Medical	No	Yes	No	Yes
MI measures	No	No	Yes	Yes
Observations	38,307	38,291	38,307	38,291
Treated	1,099	1,101	1,099	1,096

Notes: Standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Treatment model: logit. 4 Nearest-neighbor matching on propensity score. Caliper: ± 0.2 sd(PS). Matching with replacement. ATT. Dependent variable: The first row use a two year follow-up from MI for both the treatment and control group. The second row use a two year follow-up from treatment start for the treated group, and two year from MI for the control group. Treatment: SSRI antidepressants within six months of first MI.

Table A10. *Receiving SSRI conditional on depression/anxiety diagnosis*

Diagnosis within six months	All	Yes	No	Diff
SSRI				
Within two months	0.01	0.24	0.01	0.23***
	(0.10)	(0.43)	(0.09)	(0.02)
Within six months	0.03	0.54	0.02	0.51***
	(0.17)	(0.50)	(0.15)	(0.03)
Observations	38,319	371	37,948	38,319

Notes: Depression or anxiety ICD codes F32, F33, F41. ATC code N06AB.

Overall, these robustness checks confirm the results in the main section, and we can conclude that there is no evidence of an increased risk of two-year mortality for MI patients receiving SSRI.

Individuals with depression or anxiety

Table A10 presents descriptive statistics of the number of individuals with a depression or anxiety diagnosis in the in- and outpatient care, conditional on whether they receive SSRI within two or six months of the MI.

While a significantly larger share of the patients with a diagnosis received SSRI compared with patients without a diagnosis, in absolute numbers there are more patients who receive SSRI without a diagnosis. This can be explained by at least two things. Only diagnoses from the in- and outpatient registry is included, excluding primary care. Thus, it is likely that there are some individuals in the sample with a diagnosis than can be seen here. Second, especially for psychiatric drugs, it is not uncommon that individuals receive medication without a corresponding diagnosis (Mojtabai and Olfson 2011). It should be noted that only about than half of the patients with a (confirmed) diagnosis receive SSRI.

C Propensity score and trimming

Table A11 show the estimations of the propensity score, using logistic regression with treatment status as outcome. The propensity score is predicted and used as a matching variable using nearest-neighbor matching.

Table A12 show the number of observations which lack overlap, i.e., no neighbor within 0.2 of the normalized SD of the PS, in the data for each respective specification. Observations without overlap is trimmed (dropped in the sample).

Table A11. *Estimating the propensity score: Logistic regression*

	(1)	(2)	(3)	(4)
Year				
2008	-0.0758 (0.0983)	-0.0813 (0.0986)	-0.0631 (0.0984)	-0.0696 (0.0987)
2009	-0.0942 (0.100)	-0.0938 (0.101)	-0.0764 (0.101)	-0.0791 (0.101)
2010	-0.207** (0.101)	-0.217** (0.102)	-0.189* (0.102)	-0.204** (0.102)
2011	-0.199** (0.0997)	-0.202** (0.100)	-0.171* (0.101)	-0.178* (0.101)
Female	0.436*** (0.0685)	0.354*** (0.0700)	0.442*** (0.0693)	0.361*** (0.0707)
Age	0.00579 (0.00408)	0.00168 (0.00424)	0.00256 (0.00439)	0.0000441 (0.00447)
Employment status				
Employed	-0.00745 (0.230)	0.0442 (0.231)	0.00895 (0.230)	0.0566 (0.231)
Retired	-0.0360 (0.239)	-0.0386 (0.240)	-0.0195 (0.240)	-0.0302 (0.240)
Smoking status				
Former smoker	-0.0562 (0.0808)	-0.0818 (0.0812)	-0.0563 (0.0811)	-0.0816 (0.0814)
Current smoker	0.412*** (0.0833)	0.377*** (0.0844)	0.416*** (0.0839)	0.375*** (0.0850)
Diabetes	0.136 (0.0879)	0.0569 (0.0920)	0.0903 (0.0892)	0.0366 (0.0928)
Hypertension	0.0711 (0.0679)	-0.0647 (0.0806)	0.0525 (0.0695)	-0.0630 (0.0813)
History of stroke	-0.422*** (0.116)	-0.321*** (0.122)	-0.409*** (0.116)	-0.328*** (0.123)
Previous PCI	0.0280 (0.207)	-0.0908 (0.219)	0.00277 (0.210)	-0.0632 (0.220)
Depression				
Within year before	1.972*** (0.349)	1.412*** (0.417)	1.988*** (0.350)	1.449*** (0.417)
Within six months after	3.837*** (0.112)	3.748*** (0.113)	3.832*** (0.112)	3.746*** (0.114)
ATC				
B01		0.146 (0.0898)		0.136 (0.0908)
C01		-0.0791 (0.102)		-0.121 (0.107)
C02		0.130 (0.290)		0.114 (0.291)
C05		-0.183 (0.225)		-0.181 (0.226)

Table A12. *Trimming of the data*

Overlap	Treatment		
±.01036	0	1	Total
0	10	2	12
1	37208	1099	38307
Total	37218	1101	38319

Notes: Specification 1.

Overlap	Treatment		
±.0106	0	1	Total
0	28	0	28
1	37190	1101	38291
Total	37218	1101	38319

Notes: Specification 2.

Overlap	Treatment		
±.01044	0	1	Total
0	10	2	12
1	37208	1099	38307
Total	37218	1101	38319

Notes: Specification 3.

Overlap	Treatment		
±.01066	0	1	Total
0	23	5	28
1	37195	1096	38291
Total	37218	1101	38319

Notes: Specification 4.

D Diagnostics

Covariate balance after matching

Tables A13-A16 presents the covariate balance of the variables after trimming, before and after matching. It is ideal that the standard difference of the matched variables is 0, and the matched ratio is 1. The matching is not very successful in the first two specifications. It is, however, much better in the third and fourth specifications. See also the love plots in this section.

Common support

Figure A1 show the common support before trimming the data. It is clear from the figures that the propensity score, or the likelihood of treatment, is relatively low for both treated and untreated individuals. There are, however, some individuals with quite high propensity scores. The figures show the density of individuals with the corresponding propensity score. Since there are so many more untreated individuals the figures hide the fact that there are almost the same amount of individuals with a propensity score above 0.3 in the both groups (173 individuals in the untreated group, and 199 individuals in the treated group, irrespective of specification).

Love plots

Figure A2-A5 show love plots for the respective specification. The love plots can be compared with the standardized difference in the raw and matched samples in Tables A13-A16. It is clear from these figures that the first two specifications do not succeed to create comparable groups, but the third and fourth specifications are successful.

Table A13. *Covariate balance: Specification 1*

	Std. diff.		Ratio	
	Raw	Matched	Raw	Matched
Year				
2008	0.040	0.017	1.063	1.025
2009	-0.003	0.035	0.997	1.060
2010	-0.056	0.012	0.917	1.021
2011	-0.042	-0.023	0.940	0.966
Female	0.240	-0.035	1.130	0.992
Age	0.023	-0.034	1.103	1.046
Employment status				
Employed	-0.044	0.017	0.962	1.017
Retired	0.033	-0.034	0.975	1.030
Smoking status				
Former smoker	-0.150	0.017	0.875	1.019
Current smoker	0.202	-0.002	1.231	0.999
Diabetes	0.051	0.015	1.101	1.026
Hypertension	0.054	0.011	1.010	1.001
History of stroke	-0.120	-0.085	1.484	1.308
Previous PCI	-0.008	0.100	0.953	2.138
Depression				
Within year before	0.101	0.000	6.444	1.000
Within six months after	0.636	0.000	32.697	1.000

Notes: ATT, 4 NN.

Table A14. *Covariate balance: Specification 2*

	Std. diff.		Ratio	
	Raw	Matched	Raw	Matched
Year				
2008	0.042	-0.015	1.065	0.979
2009	-0.003	0.035	0.995	1.061
2010	-0.055	0.039	0.920	1.070
2011	-0.043	-0.001	0.939	0.998
Female	0.240	-0.032	1.130	0.993
Age	0.021	0.031	1.104	0.991
Employment status				
Employed	-0.045	-0.004	0.961	0.996
Retired	0.034	0.005	0.974	0.996
Smoking status				
Former smoker	-0.147	0.069	0.877	1.084
Current smoker	0.201	-0.058	1.230	0.959
Diabetes	0.053	-0.016	1.104	0.973
Hypertension	0.054	0.058	1.010	1.010
History of stroke	-0.122	-0.033	1.495	1.103
Previous PCI	-0.008	0.022	0.951	1.149
Depression				
Within year before	0.118	0.026	10.816	1.330
Within six months after	0.638	0.001	32.957	1.001
ATC				
B01	0.104	0.062	1.080	1.043
C01	0.045	-0.003	1.090	0.995
C02	0.022	0.018	1.223	1.182
C05	0.020	0.023	1.142	1.162
C07	0.083	0.032	1.069	1.024
C09	0.096	0.053	1.074	1.037
C10	0.049	0.037	1.070	1.051
N05	0.385	0.027	1.606	1.021
N06	0.014	0.000	1.690	1.000
N07	0.070	0.020	1.852	1.163
ICD				
F	0.149	-0.009	3.715	0.950
I	0.047	0.023	1.208	1.093
J	0.005	0.005	1.082	1.081
N	0.017	-0.007	1.365	0.889

Notes: ATT, 4 NN.

Table A15. *Covariate balance: Specification 3*

	Std. diff.		Ratio	
	Raw	Matched	Raw	Matched
Year				
2008	0.043	0.002	1.066	1.003
2009	-0.003	0.017	0.997	1.028
2010	-0.056	0.011	0.917	1.018
2011	-0.042	-0.021	0.940	0.969
Female	0.239	-0.030	1.130	0.993
Age	0.023	0.008	1.105	1.026
Employment status				
Employed	-0.046	0.015	0.960	1.015
Retired	0.034	-0.016	0.974	1.014
Smoking status				
Former smoker	-0.145	0.039	0.878	1.045
Current smoker	0.200	-0.021	1.229	0.984
Diabetes	0.054	-0.022	1.105	0.964
Hypertension	0.055	0.015	1.010	1.002
History of stroke	-0.122	-0.017	1.495	1.050
Previous PCI	-0.008	0.018	0.953	1.122
Depression				
Within year before	0.110	0.012	7.309	1.142
Within six months after	0.634	0.001	32.189	1.002
ECG rhythm				
Atrial fibrillation	0.030	-0.018	1.090	0.953
Sinus	-0.049	0.003	1.124	0.994
Systolic blood pressure at admission	-0.010	0.011	1.027	0.960
Heart rate at admission	0.118	-0.008	1.000	0.909
History of CHF	0.045	-0.028	1.237	0.888
ECG QRS	0.075	0.010	1.057	1.006
Killip class	0.020	-0.007	1.054	0.984
Reperfusion treatment	-0.026	0.019	0.983	1.014
Bleeding under care	0.074	0.057	1.747	1.503
CPR or defib	-0.043	-0.011	0.763	0.928
Mechanical complication	-0.006	-0.037	0.892	0.547
New atrial fibriliation	-0.014	-0.019	0.939	0.917
Reinfarct during care	0.005	-0.016	1.052	0.852
Left Ventricular Ejection Fraction	0.057	-0.005	1.044	0.997
(LVEF: Missing)	0.032	0.020	1.043	1.026
AV block	-0.024	-0.021	0.828	0.847
Beta blockers at discharge	0.034	0.023	0.919	0.944
Statins at discharge	-0.050	-0.001	1.114	1.001
Nitrates at discharge	0.074	0.006	1.207	1.013
ACE inhibitors or Angiotensin II at discharge	0.001	-0.001	1.000	1.001
Other lipid lowering agents at discharge	-0.010	-0.018	0.902	0.835
Other antiplatelet at discharge	-0.127	0.019	1.195	0.978

Notes: ATT, 4 NN.

Table A16. *Covariate balance: Specification 4*

	Std. diff.		Ratio	
	Raw	Matched	Raw	Matched
Year				
2008	0.042	-0.014	1.066	0.980
2009	-0.004	0.035	0.995	1.060
2010	-0.052	0.012	0.923	1.020
2011	-0.043	-0.024	0.939	0.964
Female	0.236	-0.004	1.129	0.999
Age	0.023	0.023	1.108	1.008
Employment status				
Employed	-0.042	0.004	0.963	1.004
Retired	0.031	-0.000	0.976	1.000
Smoking status				
Former smoker	-0.144	0.056	0.880	1.067
Current smoker	0.196	-0.043	1.226	0.968
Diabetes	0.050	-0.041	1.098	0.934
Hypertension	0.055	0.014	1.010	1.002
History of stroke	-0.123	-0.007	1.500	1.021
Previous PCI	-0.008	0.006	0.956	1.037
Depression				
Within year before	0.108	0.027	8.926	1.381
Within six months after	0.631	0.004	31.805	1.007
ATC				
B01	0.106	0.030	1.081	1.020
C01	0.045	-0.001	1.089	0.999
C02	0.014	0.004	1.145	1.040
C05	0.020	0.019	1.146	1.133
C07	0.084	0.012	1.071	1.009
C09	0.096	0.013	1.074	1.009
C10	0.050	-0.013	1.071	0.984
N05	0.378	0.013	1.598	1.010
N06	0.014	-0.014	1.698	0.667
N07	0.063	-0.035	1.759	0.778
ICD				
F	0.144	-0.025	3.588	0.861
I	0.044	0.029	1.196	1.122
J	0.005	-0.007	1.087	0.889
N	0.002	0.000	1.040	1.000
ECG rhythm				
Atrial fibrillation	0.031	0.005	1.093	1.015
Sinus	-0.050	-0.002	1.127	1.005
Systolic blood pressure at admission	-0.010	0.001	1.026	0.935
Heart rate at admission	0.114	-0.013	1.000	0.918
History of CHF	0.046	0.019	1.241	1.091
ECG QRS	0.071	0.008	1.055	1.005
Killip class	0.021	0.000	1.057	1.000
Reperfusion treatment	-0.024	0.000	0.984	1.000
Bleeding under care	0.069	0.011	1.683	1.077
CPR or defib	-0.042	0.027	0.767	1.218
Mechanical complication	-0.006	-0.023	0.894	0.668
New atrial fibriliation	-0.013	0.001	0.942	1.005
Reinfarct during care	0.006	-0.014	1.061	0.871
Left Ventricular Ejection Fraction	0.057	0.003	1.044	1.002
(LVEF: Missing)	0.029	0.009	1.040	1.012
AV block	-0.024	-0.024	0.830	0.824
Beta blockers at discharge	0.033	0.015	0.922	0.961
Statins at discharge	-0.049	-0.012	1.111	1.025
Nitrates at discharge	0.069	0.011	1.193	1.026
ACE inhibitors or Angiotensin II at discharge	0.003	-0.001	0.998	1.001
Other lipid lowering agents at discharge	-0.010	-0.018	0.904	0.835
Other antiplatelet at discharge	-0.126	0.021	1.194	0.976

Notes: ATT, 4 NN.

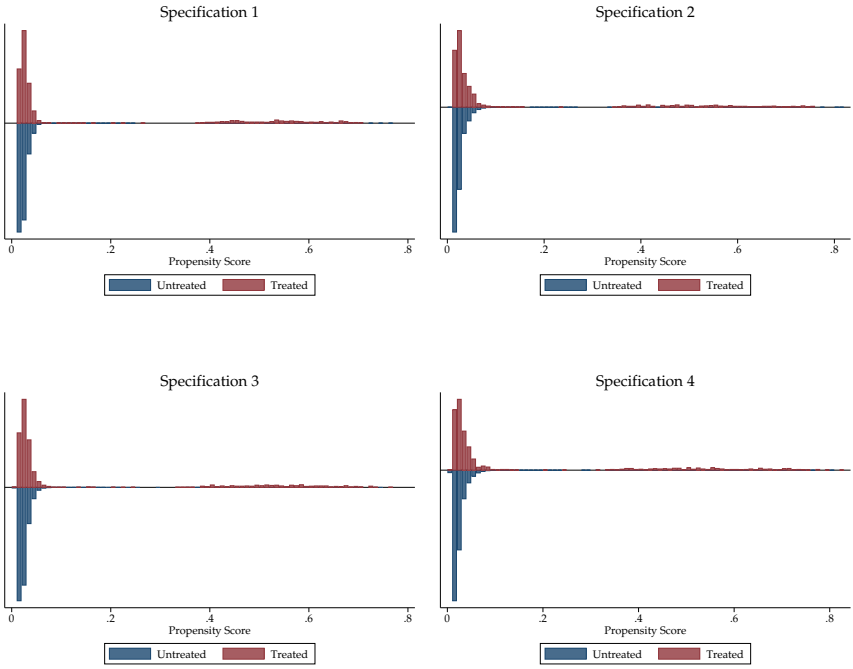


Figure A1. Common support.

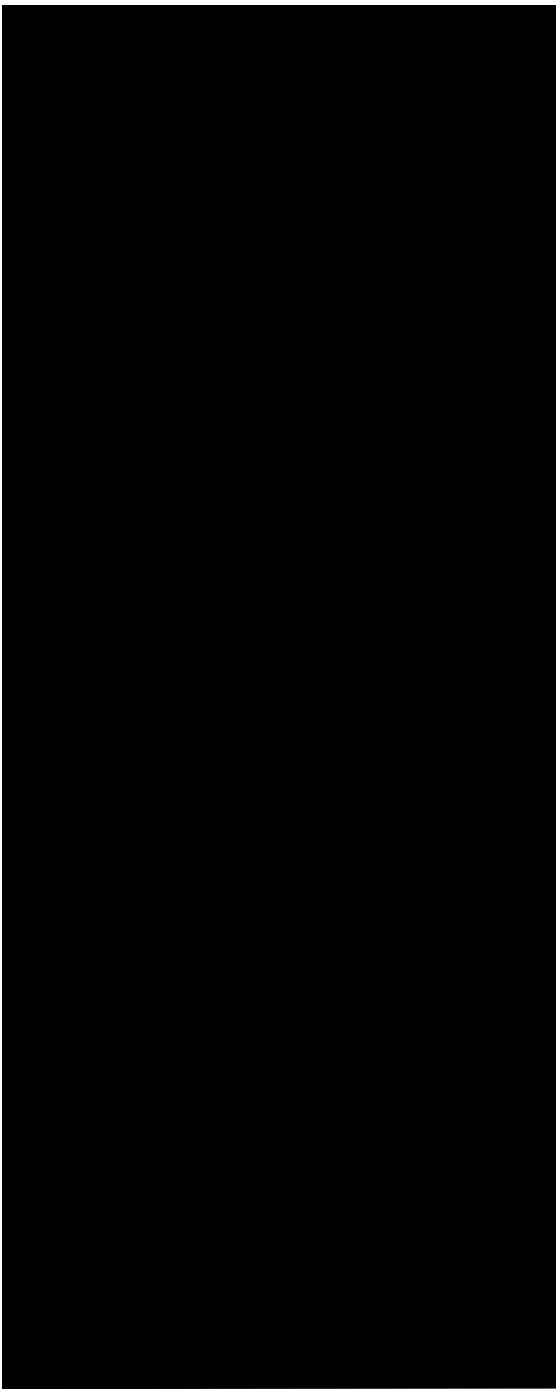


Figure A2. Specification 1. ATT, 4 NN.

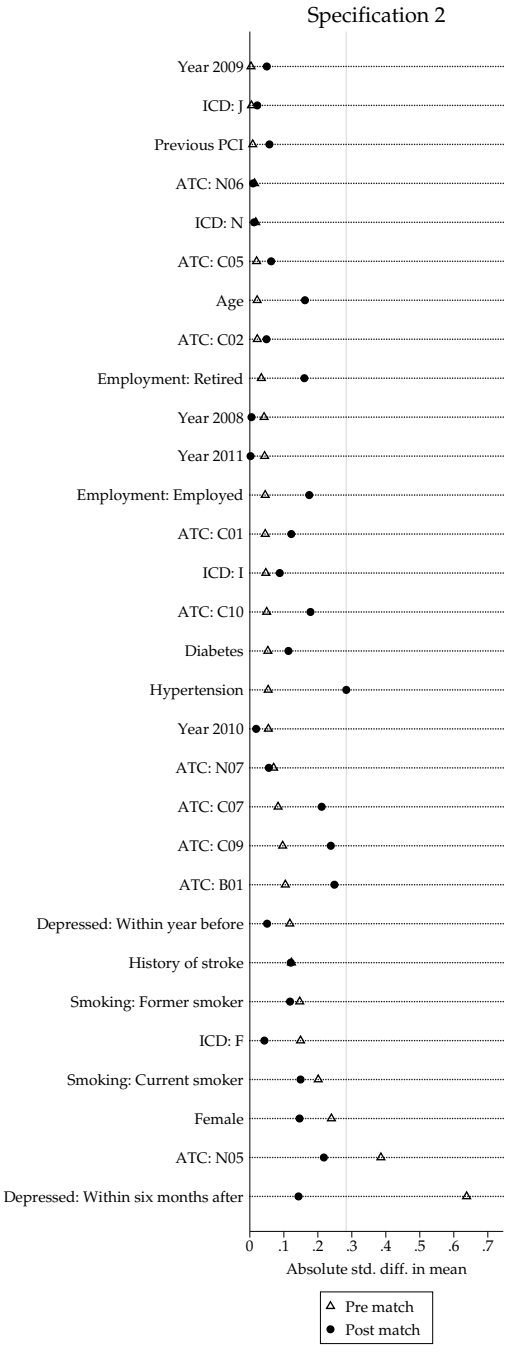


Figure A3. Specification 2. ATT, 4 NN.

Figure A4. Specification 3. ATT, 4 NN.

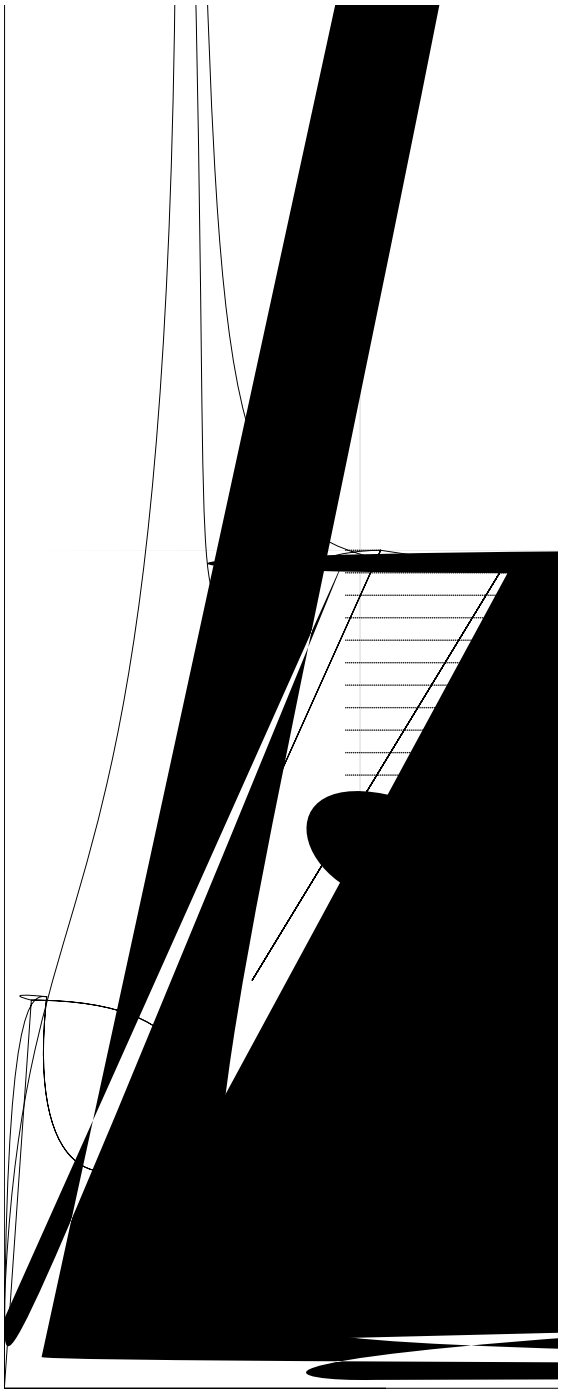


Figure A5. Specification 4. ATT, 4 NN.