IFAU
Institute for Evaluation of Labour
Market and Education Policy

Microeconometric analyses of individual behavior in public welfare systems

Daniel Avdic

DISSERTATION SERIES 2014:1

Presented at the Department of Economics, Uppsala University

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P O Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

This doctoral dissertation was defended for the degree of Doctor in Philosophy at the Department of Economics, Uppsala University, February 28, 2014. Essays 1, 3 and 4 have previously been published by IFAU and are revised versions of Working papers 2013:13, 2014:7 and 2011:12, respectively.

ISSN 1651-4149

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Ekonomikum, Kyrkogårdsgatan 10, Uppsala, Friday, 28 February 2014 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Colin Cameron (University of California, Davis).

Abstract

Avdic, D. 2014. Microeconometric Analyses of Individual Behavior in Public Welfare Systems. Applications in Health and Education Economics. *Economic studies* 143. xiv +161 pp. Uppsala: Department of Economics. ISBN 978-91-85519-50-7.

This thesis consists of four self-contained essays.

Essay 1: Women have more absenteeism while simultaneously live longer than men. This pattern suggests that men and women's preferences for sickness absence might differ due to e.g. distinct health behaviors. These behaviors could, in turn, arise from the traditional gender division of labor within households, in which it might be more important to invest in the woman's health. We empirically analyze these hypotheses using administrative health data and find robust evidence for gender differences in preferences for health-related absenteeism.

Essay 2: The paper analyzes whether residential proximity from an emergency room affects health outcomes from suffering an acute myocardial infarction (AMI). Previous research has suffered from empirical problems relating to health-based spatial sorting of involved agents and data limitations on out-of-hospital mortality. Using policy-induced variation in hospital distance, arising from emergency room closures, and data on all AMI deaths in Sweden over a twenty-year period, results show a clear and gradually declining probability of surviving an AMI as hospital distance increases.

Essay 3: Although learning-by-doing is believed to be an important source of productivity, there is limited evidence that increased production volume enhances productivity. We document evidence of learning-by-doing in a high-skill activity where stakes are high; advanced cancer surgery. For this purpose, we introduce a novel instrument that exploits changes in the number of public hospitals across time and space, affecting the number of cancer surgeries performed in Swedish hospitals. Using detailed register data, our results suggest substantial positive effects of operation volume on post-surgery survival rates.

Essay 4: The paper analyzes whether student choice of college financing affects study durations by exploiting an intervention in the Swedish student aid system. The reform provided incentives for college students to reallocate time from studies to market work. We evaluate this time reallocation hypothesis by estimating relative changes in earnings and completed academic credits attributed to the intervention for students from different socioeconomic backgrounds. Applying detailed Swedish administrative data, we find that the intervention both increased relative earnings and decreased the relative study pace for students from a lower socioeconomic background.

Keywords: causal effect, public policy, health behavior, gender norms, learning-by-doing, health care access, time-to-graduation

Daniel Avdic, Department of Economics, Box 513, Uppsala University, SE-75120 Uppsala, Sweden.

ISSN 0283-7668
ISBN 978-91-85519-50-7
urn:nbn:se:uu:diva-215654 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-215654)

Till Nurea

Acknowledgements

My Swedish Lutheran heritage makes writing acknowledgments a daunting task. It is not that I do not feel thankful as much as it is the actual public expression of gratitude that makes the, allegedly, simple task of putting together a "thank" and a "you" surprisingly complicated. However, my recent endeavors into the Catholic part of continental Europe might have facilitated this objective to some extent. Below is an incomplete list of persons I would like to thank who, in various ways, helped me during the four years I spent completing this thesis. There are numerous more to whom I am indebted, but I hope that anyone unintentionally not mentioned here can find it in their hearts to forgive me for this.

First of all, I would like to extend my most sincere gratitude to the outstanding thesis supervision conducted by Professor Per Johansson, who have been a steadfast cliff of academic expertise in a tempest of uncertainty for me ever since I began work on my Master's thesis. One of the many things he has taught me over the years is that inference in empirical analysis fundamentally depends on the *a priori* assumptions put on the missing data. Despite this insight, I would still venture to say that I could not in any counter-factual universe have had found a better supervisor. Thank you Per for always being available to give advice in times of need, always taking your time to discuss research and always helping me out the best you can even when I have the feeling that it might have involved substantial additional work for you. I would also want to take this opportunity to say that not only are you a great supervisor but — in my opinion — also one of the best researchers out there.

I am also deeply indebted to my co-advisor Johan Vikström for being an excellent complement to Per; while I have sometimes been slightly distracted by irrelevancies, Johan has been a master of efficiency clarifying when, what and where things should be done in a very structural way. Without this assistance I am positively sure the thesis would still be labeled "work in progress". Moreover, apart from being an excellent researcher, Johan is also a person who always makes one feel in a good mood — something one cannot get to much of while writing a PhD thesis.

I would also like to express my thanks to the discussant at my final seminar, Martin Karlsson for traveling all the way from Essen in Germany to comment on my papers, and to Colin Cameron for accepting to be the discussant for my defense.

I have been blessed with being associated to several different excellent research institutions over the years; first, I would like to thank everyone at IFAU

for taking me in and sharing with me your productive environment of research and policy mixed in just the right proportions. Just to mention a few, I would like to give credits to Maria Karlsson, for her inexhaustible patience in helping me out with all the travel expenses I constantly bothered her with; IFAU:s grand old man, Anders Forslund, for employing me and simply being the "Anders" we all are very fond of; Olof Åslund, for being a great director of the institute and for all the input he generously supplied me with during our yearly reviews; and, of course, my fellow room-mates Martin, Arizo and Ulrika, for being patient enough to endure my endless monologues of academic chatter.

I also convey a large chunk of gratitude to the people at the Institute for Futures Studies during the time I was a part of this inspiring multi-disciplinary environment. Thanks to the late Professor Thomas Lindh, Daniel Hallberg, Christer Rosen, Charlotte Thulstrup and Gustav Öberg of the "NTA" team and also to Joakim Palme, Jakob Larsson, Martin Korpi, Torbjörn Lundkvist and Ewa Koff. The year I spent working at IF gave me valuable insights and experiences in academic practical life before I started my PhD.

Even though I did not spend too much time in the Department of Economics at Uppsala University, I would still allow myself to thank my fellow classmates, especially Gabriella Chirico and Tove Eliasson (thank you Tove for sharing with me that wonderful semester in Chicago). Thanks also to the department administration, especially Katarina Grönvall for being so nice and helpful at all occasions.

While I did not really travel much for professional reasons until my third year, I am now spending more time outside of than inside Sweden. My personal international experiences, in particular all interesting persons I have had the pleasure to meet along the way, has been the dominating cause for this trend reversal. Starting with the IZA summer school of '12, I would like to thank everyone who made this such a great event — I have never had that much fun while working — especially my friends Arna Vardardottir and Maren Michaelsen. Further, I have benefited much from the semester I spent as a visiting scholar at the Department of Economics at Northwestern University, where I specifically would like to give a big shout-out to Professor Chuck Manski who, taking time with both consultation and lecturing, fundamentally have changed the way I approach econometrics and identification in empirical analysis. I am also grateful to Nicolas Ziebarth for inviting me to Cornell University and letting me take part of his inspiring Health Economics research environment. Finally, I want to thank everyone at CINCH and the chair of health economics at the University of Duisburg-Essen for their warm welcoming and the assistance they provided me with during the, still ongoing, transition to a new country, i.e. overcoming the notorious German bureaucracy. All in all, without my international experiences, I would be intellectually much poorer than I am today.

I also like to thank all my friends outside of academia — none mentioned, none forgotten — whom to a large extent I have neglected while writing this

piece of work. While most of them now have families and live ordinary family lives, I still maintain the direction of my academic pursuit insatiably. In this respect, a special mention goes to Marie where, sadly, what first brought us together eventually also drove us apart.

Last, but certainly not least, I would like to thank the members of my immediate family: My mother Ann for always making me feel welcome at the place where I once was raised; my father Dan for tolerantly accepting who I am despite our sometimes diametrically opposite political views; my brother Adam for keeping me off my high horses when it is necessary; my sister Karin for persistently keeping in touch when I do not; and, finally, my grandmother Nurea — to whom i dedicate this thesis — for all the great and life-lasting memories she provided me with during my childhood. I love you all very much.

Essen, December 2013

Daniel Avdic

Contents

Introduction

Although most of today's modern societies have developed comprehensive welfare systems over the course of history, the format and extent of these systems varies considerably in different parts of the world. From the Anglo-Saxon tradition of a *residualist* welfare norm, where the social security system is primarily designed for the poor and intended to be a minimal financial safety net against social deprivation; over the *solidaristic* French system, where policies for including people who for various reasons (e.g. unemployment) have been excluded from society is the main target for government support; and the *corporativist* welfare system in Germany, where independent social security funds manage most of the social support systems with only a subsidiary role for the federal government; to the *institutional* welfare systems of the Scandinavian countries, where social security is extensive, provided for the entire population and is based on notions of mutual responsibility and income redistribution (Spicker, 2000).

Irrespective of the institutional context or whether the arguments for welfare are humanitarian, democratic or practical, social policy — i.e. the actions with which governments implement welfare and social protection systems — is the essential constituent in the design and maintenance of all modern welfare systems (cf. Spicker (2008); Kennedy (2013)). However, to enable the creation of social policies that eventually achieve their original purposes, it is essential that a knowledge base with respect to the expected effects of such policies are first created — not only in retrospect, but also in order to allow for forecasts of possible *future* policy measures. To inform decision-makers with a relevant and objective basis of knowledge that enables them to formulate and implement policies designed to achieve political objectives is perhaps the most important role of researchers in the social sciences. These joint efforts of academia and elected representatives of the democratic system is an vital component in the development of our modern societies (Nathan, 2000).

Whilst the task of contributing to a factual basis for policy is important, there is also another, methodological, reason to conduct policy analysis. Specifically, policy interventions set the stage for empirical investigations of important research questions as they can often be utilized as *natural experiments*; that is, studies which exploit variation that is not manipulated for the purposes of research by combining a research design and analytical features of the data to allow causal inferences to be drawn. Natural experiments widen the range of policies that can usefully be evaluated beyond those that are amenable to planned experimentation, i.e. those that are difficult to manipulate in a controlled experimental setting for either practical, political or ethical reasons (cf.

Robinson *et al.* (2009); Rosenzweig and Wolpin (2000); Diamond and Robinson (2010)). Since the identification of causal effects is the main goal — but also the greatest difficulty — of policy analysis, the importance of these interventions for making inference in observational studies can not be overemphasized.

The main theme of this dissertation comprises empirical studies of individual behavior and public policy in the context of a modern welfare state. It consists of four self-contained essays, each focusing on different aspects of the interaction between the publicly provided health and education systems and individual behavior in Sweden over the last two decades. The national health and education systems are key policy elements within the modern welfare state as important propellants for growth and personal well-being. Thus, it is my hope that the essays included in this monograph will be able to contribute with new and valuable knowledge about key aspects in these areas of social policy.

The first paper in this thesis addresses the persistent gender gap in absenteeism observed in many countries with a focus on whether different health behavior among men and women may explain this difference. The second and third paper analyzes the organization of the national health care systems with respect to the degree of centralization of inpatient care. The first of these articles analyzes the health effects of a reduced access to emergency health care in terms of geographical distance to an emergency clinic. The second article examines whether a higher surgical volume can increase the quality of cancer care through learning-by-doing effects. Finally, the last paper in this thesis focuses on the effectiveness of the national system of higher education with respect to the relationship between students academic performance and the amount of labor they supply during their studies.

Gender, absenteeism and the morbidity-mortality paradox

The first paper of the thesis (Essay 1) addresses the causes for the observed gender gap in health-related absenteeism. In most developed countries women have considerably higher levels of sickness absence than men (Mastekaasa and Olsen, 1998). Yet, while most morbidity measures show a similar overrepresentation among women, there is one major exception to this rule — the remaining life expectancy — which has led some scholars to label this relationship the *morbidity-mortality paradox* (cf. Lee (2010)). One suggested explanation for this apparently inconsistent pattern has been the existence of gender differences in health behavior where women use common measures of morbidity proactively in order to keep healthier, thus prolonging their lives relative to men (cf. Nathanson (1975); Verbrugge (1982); Stronegger *et al.* (1997); Uitenbroek *et al.* (1996)). Moreover, Paringer (1983) suggested

that these differences could arise from households higher relative valuation of womens' health. Specifically, due to the traditional household gender division of labor, where women typically perform dual roles while men specialize, an illness in the woman does not only imply (insurable) forgone earnings but also (uninsurable) household production losses. Therefore, households may rationally optimize by *investing* more in womens' health using e.g. increased absenteeism as a prevention device. As sickness absence is almost exclusively discussed from an adverse standpoint, an analysis of this beneficial investment perspective of absenteeism may entail valuable implications for health policy.

To empirically analyze whether womens' higher absenteeism can be explained by gender-specific health behavior, we sample men and women with an observed hospital admission and estimate the relative gender response in sickness absence from these admissions. To this end, we use linked Swedish longitudinal administrative data on inpatient care, sickness absence and mortality. The empirical design allows us to control for unobserved gender differences in health, economic incentives and other factors confounding the relationship between sickness absence and sex. Moreover, since the sexes may generally suffer from different types of diseases, we also condition on the type of disease causing the hospitalization and assess the relative post-admission health of men and women by means of mortality data.

We find that sickness absence increases more for women than for men after the hospital admission, irrespective of the type of disease considered. This result is also remarkably robust to a number of sensitivity analyses. Furthermore, we find that the mortality rate is relatively lower for women after hospitalization, which leads us to rule out the explanation that different degrees of severity causing the admissions could explain the heterogeneous reactions. Finally, a considerable share of the difference in sickness absence response is driven by women with children, indicating that household health investments may be important in explaining womens' higher absence rates.

Geographical accessibility and quality of health care

The second and third papers of the thesis (Essays 2 and 3) are related and concern the organization of inpatient care with respect to the degree of centralization of resources and the quality of provided care. The first of these papers empirically analyzes the adverse health consequences of a reduced geographical access to emergency health services for individuals who suffered from an acute myocardial infarction (AMI). AMI is one of the most common causes for both morbidity and mortality in the Western world and a disease where time plays a crucial role in the likelihood of successful treatment. (Nationellt register för hjärtstopp, 2011). The time aspect is further reinforced by the fact that medical assistance may often be unavailable when the potentially life-threatening condition occurs, since most AMIs happen outside a hospital

clinic. This disease characteristic is also reflected in the national mortality statistics where e.g. a majority of all annual reported AMI deaths in the U.S. occur outside hospitals (American Heart Association, 2012). Due to the public health significance of AMI in combination with recent trends of health care centralization in many countries, it is not only interesting to analyze the health effects of geographical access to emergency health care *per se*, but also to study the policy implications of these agglomeration processes.

The specific question asked in this paper is to what extent the quality of the resulting health care is influenced by the distance an individual who sustained an AMI resides from an emergency hospital. While this relationship has been previously analyzed (cf. Bachmann *et al.* (1986); Piette and Moos (1996); Norris (1998); Pell *et al.* (2001)) this essay further contributes by addressing a number of hitherto unsolved empirical problems by exploiting variation in the number of emergency rooms in Sweden over the past two decades. Specifically, patients whose designated "home" hospital was closed were forced to switch to another "referral" hospital subsequent to the closure, generally located at a different distance from the patient's home. Utilizing the variation in hospital distance generated by these closures credibly circumvents problems arising from patients potential endogenous sorting on distance from the hospital with respect to health status.

The estimation results from sampling all AMI occurrences in Sweden over a twenty year period, using data on both hospitalizations and mortality, show that an increased distance to an emergency hospital predicts a lower likelihood of surviving an infarction by as much as two percentage points for every ten kilometer distance. Furthermore, the results indicate that this effect is mainly driven by an increased out-of-hospital mortality and that a symmetric effect (i.e. an *increased* survival probability) is discernible for patients who experienced a shorter hospital distance as a result of the hospital closures. Taken together, these findings provide considerable evidence that a centralization of inpatient care, while potentially beneficial in the context of planned surgery (see e.g. Luft *et al.* (1987); Maerki *et al.* (1986); Hamilton and Ho (1998) and below), may also carry negative health consequences in terms of reduced availability to medical treatment in emergency situations.

Learning-by-doing in advanced cancer surgery

The third paper in the dissertation (Essay 3) discuss health care organization from the favorable side of resource centralization by exploring the relationship between the number of surgeries performed in a hospital clinic and postoperative quality measures of care. The observation that larger hospitals tend to perform better has been coined the *volume-outcome* relationship in the medical literature (cf. Halm *et al.* (2002); Kizer (2003); Birkmeyer *et al.* (2003)) but is also closely related to the general economic literature on *learning-*

by-doing effects (cf. Arrow (1962); Lucas (1988); Thornton and Thompson (2001); Thompson (2001); Levitt *et al.* (2012)). The intuition for this relationship is simple — the more persons or organizations perform a specific task, the better they may become at doing it. This paper examines learning-by-doing in cancer surgery; a specialized profession associated with high stakes, as reflected in the high mortality risk of the disease, making potential learning effects highly relevant from a health policy perspective.

Although it is generally acknowledged that increased production enhances productivity through learning-by-doing effects in virtually all sectors of the economy, it has been surprisingly difficult to empirically pin down a credible causation of this phenomenon. The main reason for this problem of inference has been that conventional measures of experience are correlated with unobserved factors that, in turn, are associated with productivity. To deal with endogeneity, we introduce a new instrument that generates quasi-experimental variation in the number of cancer surgeries performed in Swedish public hospitals between 1998-2008. The instrument exploits regional variation in the closures and openings of cancer surgery clinics over time, affecting the surgical volume in nearby clinics in ways that are arguably unrelated to individual cancer surgery outcomes — due to the public nature of the Swedish health care system.

The paper uses detailed health data on the three most common types of cancer in the Western world; breast, prostate and colorectal cancer — a total of more than 100,000 cancer surgeries — performed at Swedish hospitals. The data includes rich and detailed information about each cancer case and is linked to individual-level register data on mortality and socioeconomic characteristics for the entire Swedish population. The estimation results suggests rapid gains of increased operation volume on individual post-surgical health outcomes. In particular, relating the size of the estimated volume effect across hospitals in Sweden, an increase in surgical volume of a hospital from the 25^{th} to the 50^{th} percentile of the hospital volume distribution would reduce the hospital's predicted four year cancer mortality rate by nearly twenty percent. Moreover, we find that learning among individual surgeons is important and increases with the complexity of the procedure; that other outcomes, such as the likelihood of a subsequent cancer surgery, is affected by the clinic surgical volume; and that a decreasing effect of volume, with a distinct learning threshold, appears when we allow for non-linearities of the effect. Hence, the substantial learning effects we identify in the data entail the important policy implication that surgical treatment of cancer might be significantly enhanced by an increased centralization of Swedish cancer care.

Financial aid, labor supply and student academic achievement

The fourth and final paper of the thesis (Essay 4) analyzes the role of the college financial aid system for the study-to-work transition. Since long study durations involve both individual costs, from lower life-time earnings, and social costs, in the form of reduced aggregate labor supply, most OECD countries provide financial aid for college students to increase college throughput rates (OECD, 2008). However, the enormous variation in average study durations observed all over the world makes it essential for education policy to identify the factors that contribute to these differences in order to design efficient education systems. This paper deals with one such factor, the degree of market work college students supply during their studies. The average OECD student in 2003 was employed about 27% of full-time while studying, see e.g. OECD (2005). Although the effects from working while studying can be positive for subsequent labor market outcomes (cf. Light (2001); Hotz *et al.* (2002); Häkkinen (2006); Geel and Backes-Gellner (2012)), there are also potential adverse effects associated with an excessive student labor supply, such as an increased risk of dropping out and an extension of the time-to-degree (cf. Ehrenberg and Sherman (1987); Stinebrickner and Stinebrickner (2003); Bound *et al.* (2012)).

This paper uses a comprehensive reform of the Swedish student financial aid system to analyze the effect of student labor supply on their academic achievements. The reform changed the relative cost of financing a college education by means of student loans (an increased interest rate) and by engaging in market work (an increase in the exempt amount of earnings), respectively. These changes gave students incentives to replace their student loans for labor income and, thus, potentially to reallocate some their time endowments from studying to market work, which might have had an adverse impact on their study pace. We exploit the timing of the intervention in the public financial aid system to test this *time-reallocation* hypothesis in order to contribute with important insights on the extent to which the design of the student aid system may affect the study duration of college students.

We sample college students enrolled prior to the intervention and follow them over time, as the reform was implemented, using detailed Swedish administrative data on incomes and academic achievements. Furthermore, as previous research has found students from a lower socioeconomic status to be more dependent on the financial aid system (cf. Ehrenberg and Sherman (1987); Card (1999)), we estimate relative effects for students from different socioeconomic backgrounds to improve on identification. The results show that students from a lower socioeconomic background increased their relative earnings by 25 percent and reduced their relative study pace by about ten percent as a consequence of the changed rules, hence lending substantial support for the time-reallocation hypothesis. This conclusion is further supported by

results from a number of supplementary robustness checks. All in all, the results from this paper show that the design of the student aid system is likely to be an important policy instrument affecting the average duration of studies, in particular for students with limited funding opportunities.

References

AMERICAN HEART ASSOCIATION (2012). *Heart Disease and Stroke Statistics – 2013 Update*. Annual report, American Heart Association, http://my.americanheart.org/professional/General/AHA-Heart-Disease-and-Stroke-Statistics-2013-Update_UCM_445937_Article.jsp.

ARROW, K. (1962). The Economic Implications of Learning by Doing. *Review of Economic Studies*, **29** (3), 155–173.

BACHMANN, J., MCDONALD, G. and O'BRIEN, P. (1986). A Study of Out-of-Hospital Cardiac Arrests in Northern Minnesota. *The Journal of the American Medical Association*, **256** (4), 477–483.

BIRKMEYER, J., STUKEL, T., SIEWERS, A., GOODNEY, P., WENNBERG, D. and LUCAS, F. (2003). Surgeon Volume and Operative Mortality in the United States. *New England Journal of Medicine*, **349** (22), 2117–2127.

BOUND, J., LOVENHEIM, M. and TURNER, S. (2012). Increasing Time to Baccalaureate Degree in the United States. *Education Finance and Policy*, **7** (4), 375–424, MIT Press.

CARD, D. (1999). Causal effects of Education on Earnings. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics 3A*, Amsterdam: Elsevier Science., pp. 1801–1863.

DIAMOND, J. and ROBINSON, J. A. (2010). *Natural Experiments of History*. Cambridge: Harvard University Press.

EHRENBERG, R. and SHERMAN, D. (1987). Employment While in College, Academic Achievment, and Postcollege Outcomes: A Summary of Results. *Journal of Human Resources*, **22** (1), 1–23.

GEEL, R. and BACKES-GELLNER, U. (2012). Earning While Learning: When and How Student Employment is Beneficial. *LABOUR*, **26** (3), 313–40.

HÄKKINEN, I. (2006). Working while Enrolled in University: Does it Pay? *Labor Economics*, **13** (2), 167–189, Elsevier.

HALM, E. A., LEE, C. and CHASSIN, M. R. (2002). Is Volume Related to Outcome in Healthcare? A Systematic Review and Methodological Critique of the Literature. *Annals of Internal Medicine*, **137** (6), 511–520.

HAMILTON, B. and HO, V. (1998). Does Practice Make Perfect? Examining the Relationship Between Hospital Surgical Volume and Outcomes for Hip Fracture Patients in Quebec. *Medical Care*, **36** (6), 892–903.

HOTZ, V. J., XU, L. C., TIENDA, M. and AHITUV, A. (2002). Are There Returns to the Wages of Young Men From Working While in School? *Review of Economics and Statistics*, **84** (2), 343–75.

KENNEDY, P. (2013). *Key Themes in Social Policy*. London: Taylor & Francis.

KIZER, K. (2003). The Volume-Outcome Conundrum. *New England Journal of Medicine*, **349** (22), 2159–2161.

LEE, C. (2010). Gender, Health and Health Behaviors. In J. Chrisler and D. R. McCreary (eds.), *Handbook of Gender Research in Psychology*, *20*, Springer Science, pp. 471–493.

LEVITT, S. D., LIST, J. A. and SYVERSON, C. (2012). *Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant*. Working paper 18017, NBER.

LIGHT, A. (2001). In-school Work Experience and the Return to Schooling. *Journal of Labor Economics*, **19** (1), 65–93.

LUCAS, R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, **22** (1), 3–42.

LUFT, H. S., HUNT, S. S. and MAERKI, S. C. (1987). The Volume-Outcome Relationship: Practice-Makes-Perfect or Selective-Referral Patterns? *Health Services Research*, **22** (2), 157–182.

MAERKI, S. C., LUFT, H. S. and HUNT, S. S. (1986). Selecting Categories of Patients for Regionalization. Implications of the Relationship Between Volume and Outcome. *Medical Care*, **24** (2), 148–158.

MASTEKAASA, A. and OLSEN, K. (1998). Gender, Absenteeism and Job Characteristics: A Fixed Effects Approach. *Work and Occupations*, **25** (2), 195–2228.

NATHAN, R. P. (2000). *Social Science in Government: The Role of Policy Researchers*. London: Rockefeller Institute Press.

NATHANSON, C. (1975). Illness and the Feminine Role: A Theoretical Review. *Social Science and Medicine*, **9** (2), 57–62.

NATIONELLT REGISTER FÖR HJÄRTSTOPP (2011). *Årsrapport 2011*. Annual report, Svenska Rådet för Hjärt-lungräddning, http://www.hlr.nu/sites/hlr.nu/files/attachment/Rapport%202011.pdf.

NORRIS, R. M. (1998). Fatality Outside Hospital from Acute Coronary Events in Three British Health Districts, 1994-5. *British Medical Journal*, **316** (7137), 1065–1070, United Kingdom Heart Attack Study Collaborative Group.

OECD (2005). *Education at a Glance 2005: OECD Indicators*. Tech. rep., OECD Publishing, http://dx.doi.org/10.1787/eag-2005-eg.

— (2008). *Education at a Glance 2008: OECD Indicators*. Tech. rep., OECD Publishing, http://dx.doi.org/10.1787/eag-2008-eg.

PARINGER, L. (1983). Women and Absenteeism: Health or Economics? *American Economic Review*, **73** (2), 123–127, Papers and Proceedings of the Ninety-Fifth Annual Meeting of the American Economic Association.

PELL, J. P., SIREL, J. M., MARSDEN, A. K., FORD, I. and COBBE, S. M. (2001). Effect of Reducing Ambulance Time on Deaths from Out-of-Hospital Cardiac Arrest: Cohort Study. *British Medical Journal*, **322** (7299), 1385–1388.

PIETTE, J. and MOOS, R. (1996). The Influence of Distance on Ambulatory Care Use, Death, and Readmission Following a Myocardial Infarction. *Health Services Research*, **31** (5), 573–591.

ROBINSON, G., MCNULTY, J. E. and KRASNO, J. (2009). Observing the Counterfactual: The Search for Political Experiments in Nature. *Political Analysis*, **17**, 341–357.

ROSENZWEIG, M. R. and WOLPIN, K. (2000). Natural 'Natural Experiments' in Economics. *Journal of Economic Literature*, **38** (4), 827–874.

SPICKER, P. (2000). *The Welfare State: A General Theory*. London: Sage.
— (2008). *Social policy: Themes and Approaches*. Bristol: Policy Press, 2nd edn.
STINEBRICKNER, R. and STINEBRICKNER, T. R. (2003). Working During School and Academic Performance. *Journal of Labor Economics*, **21** (2), 473–91.
STRONEGGER, W.-J., FREIDL, W. and RASKY, V. (1997). Health Behaviour and Risk Behaviour: Socioeconomic Differences in an Austrian Rural County. *Social Science and Medicine*, **44** (3), 423–426.
THOMPSON, P. (2001). How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study. *Journal of Political Economy*, **109** (1), 103–137.
THORNTON, R. and THOMPSON, P. (2001). Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding. *American Economic Review*, **91** (5), 1350–1368.
UITENBROEK, D. G., KEREKOVSKA, A. and FESTCHIEVA, N. (1996). Health Lifestyle Behaviour and Socio-Demographic Characteristics. A Study of Varna, Glasgow and Edinburgh. *Social Science and Medicine*, **43** (3), 367–377.
VERBRUGGE, L. M. (1982). Sex Differentials in Health. *Public Health Report*, **97** (5), 417–437.

Essay 1: Gender differences in preferences for health-related absenteeism*†

1 Introduction

Health-related absenteeism has been persistently higher for women than for men in most developed countries over the last decades. For example, Figure 1.1 illustrates the percentage gender difference in reported sickness absence for a number of European countries over time. Similar gender patterns of absence emerge regardless of whether one looks at the extensive or intensive margin of absenteeism, or whether the data originates from surveys or administrative records (Mastekaasa and Olsen, 1998). Furthermore, these observations are also in line with observed gender differences in many common measures of morbidity such as medical care utilization and self-reported health (Sindelar, 1982).

If one were to use sickness absence as an objective measure of health, one would quickly reach the conclusion that women have poorer health than men. At the same time, however, women outlives men. In fact, the remaining life expectancy is higher for women than for men in all ages and in nearly all parts of the world. The global average gender difference in life expectancy was about four years in 2010 and has been persistently so for a long time (Lee, 2010).

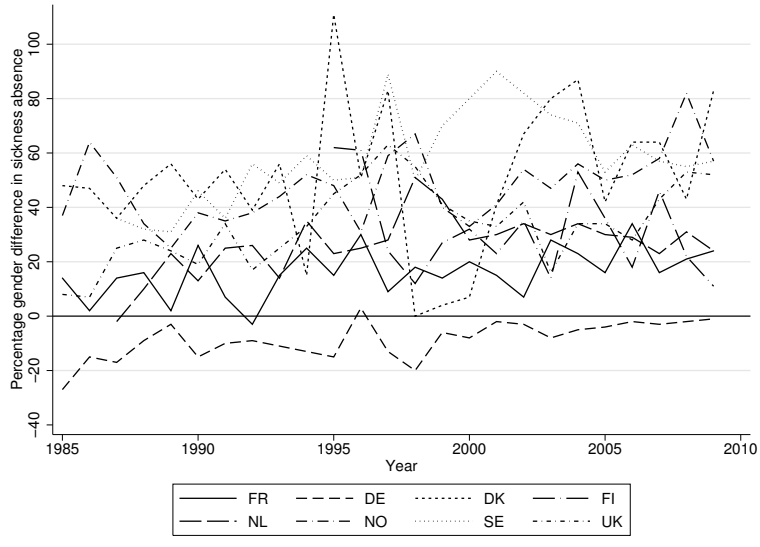
A common explanation for these seemingly contradictory observations is the existence of gender differences in health-related behavior; in particular, that women generally act more proactively than men in matters related to health (Nathanson, 1975; Verbrugge, 1982; Sindelar, 1982; Schappert and Nelson, 1999; Stronegger *et al.*, 1997; Uitenbroek *et al.*, 1996).[1] This explanation is supported by experimental evidence showing that women in general are more risk averse than men (see e.g. surveys in Eckel and Grossman

*Co-authored with Per Johansson.

†The authors thank Nicolas Ziebarth, Johan Vikström, Will White and seminar participants at the ESPE and EALE conferences 2012, the Department of Economics and the Health Economic Forum at Uppsala University, the New Perspectives on Sickness Absence and Retirement workshop in Oslo and the workshop on Absenteeism and presenteeism in Trier, University of Duisburg-Essen and the 15th IZA Summer School in Labor Economics. Financial support from the Swedish Council for Working Life and Social Research (DNR 2004-2005 and 2009-0826) is also gratefully acknowledged.

[1]This particular explanation for the so-called *morbidity-mortality* paradox was discussed already in the 17th century. The English demographer John Graunt elaborated on the observation that both the birth and death rates of men were higher than for women:

Figure 1.1. Gender difference in sickness absence prevalence by country, 1980-2010



NOTE.— Source: Eurostat. The vertical axis is defined as the percentage difference in the share of women divided by the share of men that reported being absent from work for health reasons during a particular survey week.

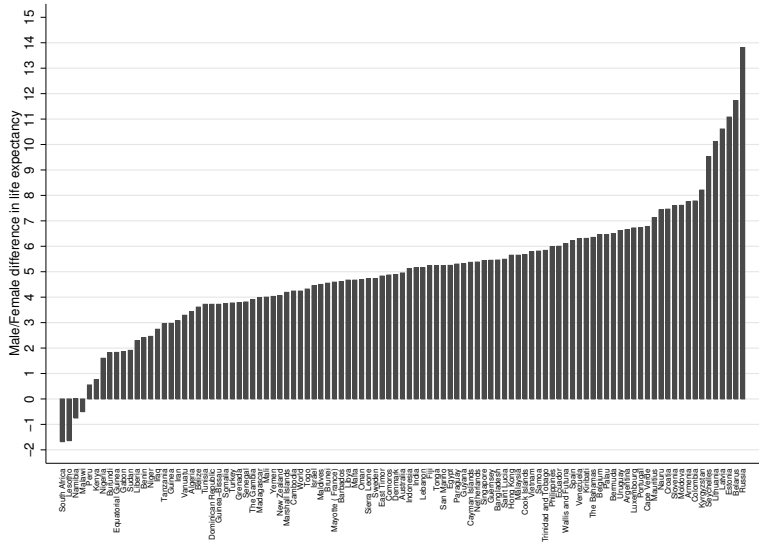
(2008); Croson and Gneezy (2009); Bertrand (2010)). Thus, if women pay more attention to potential future diseases, e.g. by utilizing medical services or sickness absence more frequently, poor health may be detected at an earlier stage, remediated and, as a consequence, increase the relative life expectancy for women in relation to men.

Why would men and women differ in their preferences toward risk and prevention? In principle, these differences may arise either from biological differences between the sexes or from historical social processes that form perceptions and norms of what constitutes and distinguishes masculine and feminine behavior. Although the task of separating the effects of heredity and environment on social outcomes is perhaps heroic, the latter explanation is to some extent supported by the large observed variation in relative life expectancy around the world displayed in Figure 1.2. In particular, the difference exceeds

"It appearing, that it were fourteen men to thirteen women [born], and that they die in the same proportion also, yet I have heard Physicians say, that they have two women Patients to one man, which Assertion seems very likely...Now, from this it should follow, that more women should die than men, if the number of Burials answered in proportion to that of Sicknesses; but this must be salved, either by the alledging, that the Physicians cure those Sicknesses, so as few more die, if none were sick; or else that men, being more intemperate than women, die as much by reason of their Vices, as women do by the Infirmitie of their Sex, and consequently, more Males being born then Females, more also die." (Graunt, 1662).

14 years in Russia, is approximately five years in northern Europe, three in Asia and in South Africa men outlive women by 1.5 years. Such large gender differences in life expectancy can hardly be attributed solely to the biological differences between men and women in different geographical regions, but also likely to contain elements of gender-specific health behaviors based on specific cultural norms in different parts of the world.

Figure 1.2. Gender difference in life expectancy by country, 2011



NOTE.— Source: CIA factbook 2011. The vertical axis is defined as the expected difference in years of life of a woman and a man for each listed country.

The reasons for the gender differences in health-related absence has scarcely been investigated so far, in spite of the geographic scope and the vast implications for e.g. gender equality policies. The few studies that do exist in the field have predominantly focused on the roles that financial incentives, health and labor market conditions play in creating these differences.

Mastekaasa and Olsen (2000) examine whether the gender differences arise from segregation in the labor market in which, the authors claim, women mainly work in more unhealthy sectors. They find, however, on the contrary that gender differences in sickness absence increase when men and women in the same occupation and labor market sector are compared. Broström *et al.* (2004) and Angelov *et al.* (2011) use Swedish data to perform Oaxaca-Blinder decompositions of the gender absenteeism gap into differences in observed attributes and (unobserved) preferences. The main conclusion from these latter two studies is that economic incentives explain part of while the working environment and health does not explain any of the differences. Interestingly, a substantial part of the variation is explained by differences in the response to self-reported health.

Angelov *et al.* (2011) study the effects of parenthood on relative absenteeism finding that mothers increase their absence by on average 0.5 days per month more than fathers for as long as 18 years after the birth of the first child. Åkerlind *et al.* (1996) similarly investigate the relation between absenteeism, age and family status and find that young women and men with children utilize the sickness insurance scheme the most and the least, respectively. Both studies conclude that gender inequality regarding the responsibility for child rearing seems to be a major factor behind the gender gap in absenteeism.

Laaksonen *et al.* (2008) use data from Finland and find that both short and medium absence spells (less than 60 days) are more common among women; a finding they contribute to a greater physical work load and work fatigue for women, rather than psychosocial working conditions and family-related factors. Finally, Evans and Steptoe (2002) investigate the importance of socially constructed gender identities using survey data of men and women in occupations where they are in a minority (accounting for women and nurses for men). In particular, they investigate whether being the minority gender at the work place is associated with poorer health and mental well-being and find empirical support for their hypothesis.

This paper contributes by first analyzing gender differences in preferences for sickness absence and, subsequently, by examining a particular reason why such a gender difference might arise. Specifically, we estimate the relative response on absenteeism for men and women who were the subject of an adverse health shock, measured by an observed individual hospital record. Our conjecture is that, if preferences for sickness absence differ between the sexes, we expect men and women to respond differently to the health shock, *ceteris paribus*. We base our analysis on detailed Swedish longitudinal register data on inpatient care, to which we have merged administrative data on sickness absence, mortality and socioeconomic characteristics. The longitudinal aspect of the data allows us to condition on pre-hospitalization gender differences in health, economic incentives and other factors which may confound the relationship between sickness absence and sex we want to evaluate.

We find that women increase their sickness absence more than men after a hospital admission. This estimated difference in absence response is remarkably stable across both labor market sectors and different categories of diseases as well as robust to a number of sensitivity analyses. Moreover, we find that post-hospitalization mortality rates are significantly lower for women than for men. This latter finding allows us to rule out the possibility that the relative response in sickness absence is driven by differences in health shock severance of men and women. Taken together, we find compelling evidence that men and women differ in their preferences for health-related absenteeism.

As a next step in the analysis, we examine whether the probed gender differences in sickness absence arise from the gender division of household responsibilities. To this end, we study whether these differences could essentially be attributed to women with, relative to without, children. The argument un-

derlying this test is the empirical observation that women typically has the main responsibility for household production, even when she also participates in the labor market, while her (male) spouse typically specializes in the labor market. The implication of this household division of labor is that women with children often performs dual roles within the household. Two principal theories exist as to why these dual roles of women may contribute to the observed gender differences in sickness absence: The first theory emphasizes the economic consequences of the division of labor in a household, while the second focuses on the consequences for health from the greater psychological pressure of these multiple roles.

The first theory, denoted the *household health investment theory*, conjectures that, due to her dual roles, a woman's health is more important for the household than the health of her spouse. The reason is that her illness does not only include the foregone earnings from being unable to perform market work, but it also creates an additional cost arising from her incapacity to perform home production (Paringer, 1983). Therefore, it may in this context be rational for the household to respond more to a negative health shock of the woman (e.g. by starting a new or extending an already ongoing spell of sickness absence) than for an equivalent health shock of the spouse. In other words, it is optimal for the household to *invest* more in the woman's — rather than in her spouse's — health.[2]

In contrast to the theory of health investments, advocates of the second theory have emphasized that the gender absenteeism gap could arise from the psychological pressure of dual roles — referred to here as the "double burden" of women (see e.g. Bratberg *et al.* (2002)).[3] Advocates of this *role strain theory* argue that multiple roles are detrimental for the well-being of the individual and may thus increase sickness absence. This could be the case, for example, if switching between the roles of performing tasks at work and at home entails a fixed cost, in particular for individuals where these roles are very different; for instance, working in an office and taking care of children.[4]

[2]Note that we do not consider preferences as being stable or innate. This means that individuals without families may have different preferences for, e.g., risk than individuals with families. Any observed difference between individuals with and without families could thus be from sorting (i.e. innate or biological) or simply formed by society or family.

[3]Time use studies in Sweden (SCB, 2009) have consistently shown that the total time worked is approximately the same for men and women. This similarity in total time worked corresponds well with statistics from time use studies in the USA, Germany and the Netherlands (Burda *et al.*, 2008). Hence, the double burden hypothesis should not be interpreted as an effect of a greater workload for women in comparison to men, but rather as an effect of the psychological strain of switching between roles.

[4]There is also literature advocating the potential benefits of having multiple roles (the *role enhancement* theory). According to this theory, individuals may feel that their lives are more meaningful when they perform several roles, which subsequently increase their well-being. Thus, the role enhancement theory gives exactly the opposite predictions of the role strain theory. See e.g. Mastekaasa (2000).

Our estimation results lend some support to the household health investment theory. In particular, the health investment effect is able to explain approximately one third of the estimated relative gender response in absenteeism from the hospitalizations.

The paper is organized as follows; the next section describes the relevant aspects of the Swedish sickness insurance system. Section three provides a stylized model framework from which we deduce empirically testable hypotheses regarding the absence behavior of men and women. Section four describes the data and the sampling method used for the estimation. Section five presents the results and section six offers some concluding remarks.

2 The Swedish sickness insurance system

All workers in Sweden, both employed and unemployed, are covered by a public health, sickness and disability insurance. The compensation levels in the sickness and disability insurance are — in an international comparison — high (around 80 and 65 percent, respectively) and the degree of monitoring is lax.[5] Based on the information in a medical certificate, the Swedish Social Insurance Agency (SSIA) determines whether the illness in question causes reduced work capacity. The proportion of cases where the SSIA decides against the physician's recommendation is, however, very small. For example, in 2006 the request for sick pay was rejected in 1.5 percent of all new cases (SSIA, 2007). The length of a sick spell is to a large extent determined by the reasons given by the insured's own motivation (Arrelöv *et al.*, 2006). Moreover, physicians often make decisions against their better judgment by, for example, prescribing too long sickness absence spells (Englund, 2001). Within this context, it is thus reasonable to assume that sickness absence is not only determined by objective health, but that there is also room for the insured individual's own judgment of his or her health.

3 Methodological framework

Our point of departure is the following stylized (Swedish) world: Market prices for household goods are high due to high minimum wages and income taxes. The income tax system is based on individual rather than household income[6] and a general sickness and disability insurance exists which replaces the income of sick individuals who are unable to take part in the labor market.

[5] See e.g. Engström and Johansson (2012) for a recent detailed description of the institutions.

[6] In 1971, Sweden changed from household taxation to individual income taxation. Selin (2009) estimated that female labor supply increased by 10 percentage points as a result of the reform.

We assume that individuals with children live in households consisting of a man, a woman and at least one child while individuals without children are assumed to be living alone.

Due to the individual-based income tax and relatively costly household services, households would benefit in economic terms, under realistic assumptions of productivity, if one of the household members specializes in labor market production and the other divides his or her time between labor market and household production. We assume the former (specializing) household member to be the male and the latter (diversifying) to be the female.[7] We also assume that the time available for household production is increasing in health (see e.g. Grossman (1972) for a theoretical argument justifying this assumption). Finally, while productivity at the workplace is difficult to monitor, home production is not subject to this type of information asymmetry since the former usually involves working for someone else while the latter more resembles a form of self-employment. The implication is that shirking ones duties may be possible at the workplace but not in the household. These assumptions together suggest that the sickness absence incentives in a household are greater for women than for men for any given level of health.

Moreover, the gender difference in absenteeism may arise from the highly segregated Swedish labor market (see e.g. SOU (2004)). For instance, the difference could be explained by that women predominantly are employed in sectors and establishments with poorer working conditions than in more male-dominated sectors and establishments.[8] These and other confounding factors complicate testing for a behavior difference in absenteeism between men and women. Our strategy to address these empirical problem is to estimate the relative response in sickness absence of men and women from an adverse health shock, measured by a hospital record, utilizing the longitudinal features of the data we have at our disposal. This way we are able to condition on the gender difference in absenteeism prior to the health shock due to e.g. economic incentives and labor market segregation. However, before we describe our empirical strategy in more detail, we discuss a stylized theoretical model for the decision to be absent from work.

3.1 Theoretical framework

Let household preferences be represented by a direct utility function

$$u(C, \mathbf{s}, \mathbf{L}), \tag{3.1}$$

[7] The reason for this, admittedly, simplification of reality is that women, traditionally, perform most of the home production (see e.g. Tichenor (1999); Booth and Van Ours (2005); Evertsson and Nermo (2007); Boye (2008)).

[8] In general, however, we would expect the opposite. That is, men face poorer working conditions than women (see e.g. Broström *et al.* (2004)).

where C is the household's consumption of goods, $\mathbf{s} = (s_f, s_m)'$ represents the vector of desired daily hours of absence for the (f)emale and the (m)ale, and $\mathbf{L} = (L_f, L_m)'$ is the vector of contracted leisure hours for the spouses. Household services are produced by the female, h_f^p, and the male, h_m^p, but are jointly consumed. We assume that there are no productivity differences between the genders and normalize the price to one (that is, the value of consumed goods is normalized by the price for household services). Hence, the household production is $\mathbf{1}'\mathbf{h}^p$, where $\mathbf{h}^p = (h_f^p, h_m^p)$ and $\mathbf{1} = (1,1)$.

The utility function is maximized under the following household budget constraint

$$\mathbf{w}'\mathbf{h} + y - (1-\delta)\mathbf{w}'\mathbf{s} - \mathbf{1}'\mathbf{h}^p = C, \tag{3.2}$$

where $\mathbf{w} = (w_f, w_m)$ is the vector of net wage rates, $\mathbf{h} = (h_m, h_f)$ are the contracted daily hours of work, $\delta \in (0,1)$ is the level of compensation in the sickness insurance and y is the family's non-labor income. Individuals also face the time constraint

$$T = L_j + h_j + h_j^p,\ j = f, m,$$

where $h_j = s_j + h_j^w$ with h_j^w as the desired daily hours of work.

Substitution of C from (3.2) into (3.1) and maximizing with respect to the absence of an arbitrary spouse, say the woman, yields the household member's demand for absence conditional on both spouses' labor supply. Under the assumption of weak separability with respect to s_m, the solution can be written as

$$s_f = f(\mathbf{h}, \mathbf{h}^p, C_{sf}, \mu_f), \tag{3.3}$$

where $\mu_f = \mathbf{wh} + y - (1-\delta)w_m s_m$ is the income net of sickness compensation (the virtual income) and $C_{sf} = \partial C/\partial s_f = -(1-\delta)w_f$ is the net cost of being absent from work. Due to the weak separability assumption, men's absence time s_m only has an income effect through μ_f in (3.3). By simply changing s_m for s_f we obtain a corresponding expression for men, i.e. for s_m, yielding $\mu_m = \mathbf{wh} + y - (1-\delta)w_f s_f$ and $C_{sm} = -(1-\delta)w_m$. For singles we have that $\mu_j = w_j h_j + y$ and $h_j^p \equiv 0,\ j = f, m$.

An increase in the cost of absenteeism (virtual income) would decrease (increase) the demand for absence; hence $\partial s_j/\partial C_{sj} (= \phi_j) < 0$ and $\partial s_j/\partial \mu_j (= \zeta_j) > 0$. Furthermore, it is reasonable to assume that increased hours in home production would increase demand for absence, hence $\partial s_j/\partial h_j^p (= \omega_j) > 0$. This last effect may arise for at least three reasons: First, it may be an effect from reduced health because home production may impair health *per se*, i.e. the double burden hypothesis (see e.g. Bratberg *et al.* (2002)). Second, performing a greater degree of household production implies a greater potential payoff from responding to an illness at an early stage in order to reduce future lost home production, i.e. household investments in health (see e.g. Paringer (1983)). Finally, sickness absence does not necessarily prohibit some home production. Hence, in a context where monitoring is lax or difficult

due to asymmetric information, sickness absence and home production can be regarded as substitutes.

Introducing heterogeneity in health, η_i (where a higher value of η_i implies better health), we get

$$s_{ij} = f(\eta_i, \mathbf{h}_i, \mathbf{h}_i^p, C_{isj}, \mu_{ij}).$$

It is reasonable to assume

$$\partial s_{ij}/\partial \eta_i (= \gamma_{ij}) < 0,$$

and

$$\partial^2 s_{ij}/\partial \eta_i \partial h_i^p (= \delta_{ij}) < 0.$$

In other words, better health reduces sickness absence in general but at an accelerating rate with respect to household production. Thus, if we could observe (exogenous) changes in health together with data on sickness absence and household production we could recover the average demand for absenteeism of men and women, $\gamma_f = E(\gamma_{if}) < 0$ and $\gamma_m = E(\gamma_{im}) < 0$ with $\kappa = \gamma_f - \gamma_m < 0$, and for men and women with different levels of household production, $\delta_j = E(\delta_{ij}) < 0, j = m, f$, respectively.

3.2 Empirical modeling

To provide an intuition for our empirical strategy, first assume that the latent variables — i.e. health status and household production — are observed and that equation (3.3) is linear in its arguments. For a sample of n individuals observed in T time periods, the regression model is subsequently given as

$$s_{itj} = \gamma_j \eta_{itj} + \lambda_{j1} h_{ij} + \lambda_{j2} h_{ij'} + \omega_{j1} h_{ij}^p + \omega_{j2} h_{ij'}^p + \delta_j h_{ij}^p \times \eta_{itj} +$$
$$\phi_j C_{sj} + \zeta_j \mu_{ij} + \varepsilon_{itj},\ i = 1, ..., n,\ t = 1, ..., T,\ j' \neq j \tag{3.4}$$

where s_{itj} and η_{itj} are days of sickness absence and health status, respectively, for individual i of gender j at time period t. By model assumption, ε_{itj} is a regression error; that is, i.i.d. and independent of the included covariates. Moreover, let c_i be a household indicator variable that takes the value one if an individual belongs to a household and zero otherwise. We assume *i)* $\mathrm{E}(h_{im}^p | c_i = 1) = 0$, *ii)* $\mathrm{E}(h_{im}^p | c_i = 0) = \mathrm{E}(h_{if}^p | c_i = 0)$ and *iii)* $\mathrm{E}(h_{if}^p | c_i = 1) - \mathrm{E}(h_{if}^p | c_i = 0) > 0$. That is, we assume *i)* that men with children specialize in market work and perform no home production, *ii)* that men and women without children perform equal amounts of home production on average and *iii)* that women with children on average perform strictly more home production than women and men without children. Finally, define

$$\delta_f^{\#} = \delta_f [\mathrm{E}(h_{if}^p | c_i = 1) - \mathrm{E}(h_{if}^p | c_i = 0)].$$

Hence, $\delta_f^{\#}$ measures the change in sickness absence due to a change in health stemming from the greater household production $\mathrm{E}(h_{if}^p|c_i=1)-\mathrm{E}(h_{if}^p|c_i=0)$.

Assume that labor supply and household production are fixed between two time periods, $t=(0,1)$, but individual health is lower in the second period, i.e. $\eta_{i1j}-\eta_{i0j}=\Delta_{ij}<0$. The change in sickness absence over the two time periods is thus given by

$$s_{i1j}-s_{i0j}=\gamma_j\Delta_{ij}+\delta_f^{\#}c_iF_i\times\Delta_{if}+\varepsilon_{i1j}-\varepsilon_{i0j},\ i=1,...,n_j,$$

where $n_j=n_{j1}+n_{j0}$ is the the total number of women ($j=f$) and men ($j=m$) aggregated over household status, $c_i=(0,1)$.

While individual health status is unobserved, we do observe an indicator of a change in health; a hospital admission. In this context, it is useful to think of the hospital admission in the following way: An individual decides to visit a hospital if his or her health falls below a certain threshold. If this threshold is different for men and women, we can write the data generating process for a hospital admission as

$$H_i=\mathbf{1}(\eta_i<\tau+\tau_FF_i),$$

where $\mathbf{1}(\cdot)$ is the indicator function (returning one if the expression inside the parenthesis is true and zero otherwise). Thus, an admission is observed if health is below a certain threshold τ for men and $\tau+\tau_F$ for women. Suppose that $\tau_F>0$ and that average health in the time period before the visit is the same for men and women. In this setting, admitted women will on average be in better health than admitted men.[9]

Let $\overline{\eta}_{jt}$ be mean health of men and women, $j=(m,f)$, at time t and Δ_j be the mean change in health of men and women between time periods zero and one. Suppose *i)* that the health of men and women is the same before the hospital admission and *ii)* that men and women have the same health thresholds for visiting a hospital (i.e. $\tau_F=0$). Under these assumptions, the changes in health of hospitalized men and women will on average be equal; that is, $\Delta_m=\Delta_f=\Delta$. Under these circumstances gender preferences for sickness absence could be identified from the hospitalizations by estimating the following model with OLS

$$s_{it}=b_0+b_1F_i+\gamma_m^*H_{it}+\kappa^*(F_i\times H_{it})+\alpha X_i+\varepsilon_{it}, \tag{3.5}$$

where $t=-l,\ldots,0,\ldots,l$, $\gamma_m^*=\gamma_m\Delta>0$ and $\kappa^*=\kappa\Delta$. Here H_{it} is a step function assuming a value of one for all time periods subsequent to the hospitalization and zero otherwise (i.e. $H_{it}=1$ for $t>0$ and 0 otherwise). X_i is a vector of control variables and b_0 and b_1 are mean parameters derived from the labor supply model (3.4). For details on the derivation see Appendix B.

[9] We assume that doctors cannot perfectly screen the health of individuals and make correct decisions whether the visit calls for an admission or not. The implication is that if women make more visits than men they also have more admissions.

This model setup ensures that the "difference-in-differences" specification in (3.5) eliminates any differences in levels of sickness absence across the groups. Hence, time-invariant differences in health, economic incentives, working conditions and other potentially confounding factors are taken into account by the empirical design. However, if men and women on average have different health thresholds when they decide to visit a hospital, i.e. $\tau_F \neq 0$, the health shocks causing the hospitalizations will on average be different between men and women, implying that $\Delta_m \neq \Delta_f$. Under the alternative hypothesis women care more about their health than men, implying that the (unobserved) health shocks causing the hospital admissions should on average be less severe for women than for men (i.e. $\tau_F > 0$). We show in Appendix B that under these assumptions the OLS estimator of κ^* based on equation (3.5) will be biased downwards. Thus, a positive OLS estimate of κ^* will in this case provide a lower bound of the gender difference in sickness absence response from the hospitalizations.

Even if the hospitalizations are considered as objective measures of health shocks, they are also in many cases an ex-post measure of any health change observed by the individual (except for e.g. accidents). The implication is that behavioral changes affecting sickness absence would for many conditions occur prior to the hospital admission. In order to address this potential problem we have performed sensitivity analyses where we estimate model (3.5) but lag the hospitalization one and two years. That is, we assume that the individual may observe his or her health change one and two years before we observe it in the data.

Theoretically, hospital admissions for women should be caused by less severe health shocks and, as a consequence, the estimated relative gender response on sickness absence would be downward biased. Empirically, however, the relationship between health and gender at the time of the hospitalization is an open question. In order to address the concern of whether women are affected by more severe health shocks than men, we evaluate the post-hospitalization (objective) health of men and women by estimating Cox hazard regression models, i.e.

$$\Pr(\text{exit}_{it} = 1 | \text{exit}_{it-1} = 0) = \lambda_0(t)\exp(\delta_1 F_i + \alpha X_i), \quad t > 0. \tag{3.6}$$

Here $\lambda_0(t)$ is the baseline hazard (i.e. the hazard rate for men) and δ_1 and α are parameters to be estimated. Specifically, we study the exits to death and to a new hospital admission. A positive estimate of δ_1 for these transitions provides evidence of relatively poorer post-admission health of women compared to men.

Next, given that a relative gender difference in the estimation of (3.5) is found, we seek to analyze whether this difference could be attributed to household health investments. To this end, we compare the relative sickness absence response of women with and without children by re-estimating (3.5) above and substituting the female indicator variable for an indicator for whether the in-

dividual have children (as a proxy for household production). We show in Appendix B that the OLS estimator will be biased downwards if the health of women with children is better than the health of women without children in relation to the hospitalization. Note, however, that we are also able to adjust for this potential bias by including the subsample of men in the estimation. To implement this procedure, we set up a triple differences model by additionally including an indicator variable for having children along with the three first level interactions of gender, the health shock and having children. Formally, we estimate (OLS)

$$\begin{aligned} s_{it} &= b_0 + b_1 F_i + b_2 c_i + b_3 F_i c_i + \gamma^*_{m1} H_{it} + \gamma^*_{m0} H_{it} c_i \\ &+ \gamma^*_f F_i H_{it} + \delta^*_f F_i H_{it} c_i + \varepsilon_{it}, \end{aligned} \tag{3.7}$$

where $\delta^*_f = \delta^{\#}_{f1} \Delta_f$ and γ^*_{mc} is the response for men with ($c = 1$) and without ($c = 0$) children. The OLS estimator of δ^*_f is consistent under the assumption that the difference in the health threshold for being admitted to a hospital for men with or without children is greater than the difference in threshold for women with or without children (see Appendix B for details).

Finally, we also evaluate relative post-hospitalization health by estimating Cox-regression models with exits to mortality and re-admission by household status,

$$\Pr(\text{exit}_{it} = 1 | \text{exit}_{it-1} = 0) = \lambda_0(t) \exp(\delta_1 F_i + \delta_2 c_i + \delta_3 F_i c_i + \alpha X_i), \quad t > 0. \tag{3.8}$$

Here $\lambda_0(t)$ is the baseline hazard (i.e. the hazard rate for men without children). A positive estimate of δ_3 suggests that women with children have poorer post-hospitalization health than women without children. In this model we adjust for potential health differences between women with and without children by using men with or without children as the counter-factual impact from having children. As we cannot control for differences in pre-hospitalization health of men and women without children, consistent estimation of the parameter of interest is subject to a stronger identifying assumption here than in the estimation of the effects on sickness absence.

3.3 Empirical deduction of the posed hypotheses

Finding that women, both with and without children, do not increase their absenteeism relative to men after the hospital admission will lead us to refute both the gender and household investment hypotheses, at least for our specific sample. On the other hand, if we find that women increase their sickness absence more than men — while not having poorer objective health measures after the hospitalization — will support the conjecture that women take more preventive action than men. Furthermore, finding that these differences in absence response are primarily attributed to women with children would support

the health investment hypothesis of Paringer (1983), unless these womens' objective health is also relatively poorer than the health of women without children. In contrast, if the higher absence rates of women with children are accompanied by higher mortality and re-admission rates compared to women without children, we would reject the health investment hypothesis in favor of the double burden hypothesis of e.g. Bratberg *et al.* (2002)).

4 Data

Our empirical analysis exploits micro data originating from administrative registers. The data, collected and maintained by Statistics Sweden, covers the entire Swedish population between 16 and 65 for years 1987-2000, and individuals aged 16 to 74 for years 2001-2010 and contains annual information on a wide range of socioeconomic variables. Furthermore, information on hospitalizations was provided by the National Board of Health and Welfare and covers all inpatient medical contacts in public hospitals between 1987-1996. This is not a major restriction since virtually all inpatient medical care in Sweden at the time was performed by the public sector and, from 1997 onwards, the register also includes privately operated health care. For a person to be registered with a health impairment, he or she must have been hospitalized which, as a general rule, means that he or she must have spent one night in a hospital. However from 2002, the registers also covers outpatient medical contacts within specialized care.

An important feature of the data is that it contains the exact cause for each hospitalization. The diagnoses, made by physicians, are classified according to the World Health Organization's International Statistical Classification of Diseases and Related Health Problems (ICD-10). ICD-10 is a seven digit coding of diseases and signs, symptoms, abnormal findings, complaints, and external causes of injury or diseases. We include all diagnosis categories in the analysis with the exception of diagnoses related to pregnancy. To take account of possible heterogeneity between the sexes in the type and severity of disease causing the hospitalizations, we include controls for disease category in the regressions. In addition, we also estimate separate models for the four most common groups of diseases of the data: ischemic heart disease, musculoskeletal disorders, cancer and mental health problems.

We use the annual number of days an individual received sickness benefits as the primary dependent variable in the regressions. Information on sickness benefits was obtained from the Swedish Social Insurance Agency (SSIA), covering all individual spells of publicly paid sick leave in Sweden. Since sickness insurance is compulsory, this does not restrict our analysis with respect to the population of interest. However, since 1992, sickness spells less than two weeks are no longer registered due to the introduction of an employer contribution period. Thus, in effect, we will only capture the effects of sickness

spells longer than two weeks in our estimations. Hence, this restriction excludes absences due to minor health problems such as common colds. We do not believe this data limitation causes any problems for the analysis of the behavioral effects that we seek to identify. In fact, the restriction might even be advantageous as it reduces the variation in sickness absence across e.g. labor market sectors.

Nevertheless, there are two other potential shortcomings with the definition of sickness absence we use. First, all workers are, in addition to the sickness insurance, also covered by a public disability insurance. These two insurances are intimately related; both are administrated by the SSIA and, as the level of compensation in the sickness insurance is higher than in the disability insurance, disabled individuals would in most cases prefer to receive sickness rather than disability benefits. Therefore, most individuals admitted to disability insurance schemes are likely to have a prehistory of sickness absence. The implication of this for our analysis is that being on a disability insurance scheme is also a likely outcome of the health change. Furthermore, being on a disability insurance scheme may also be a relevant health investment decision for the same reason as with sickness absence. To investigate this possibility we performed sensitivity analyses where we include individuals on disability insurance schemes in the sickness absence outcome definition. Second, group differences in mortality rates could potentially bias the results as the hypothetical level of sickness absence of deceased individuals is unobserved. In order to investigate this potential source of bias, we also include deceased individuals by setting their absence to zero for all years subsequent to their death.

We take a random sample of 40 percent of all employed[10] individuals between 20 and 50 years of age in 1992 who experienced an in-hospital care record at some point between 1993-2004. The sample is restricted to employed individuals for all the years prior to the hospitalization but, for obvious reasons, relax it thereafter.

Since most individuals have children at some point in their lives, defining a household as a unit of observation with at least one child would make household status a time-varying variable. Given this definition, comparing individuals with and without children at the time of the hospitalization would imply that the estimated results would also reflect life cycle variation in e.g. health and absenteeism. Moreover, the hospitalization itself could also have an effect on the probability of having a child. In order to circumvent these problems, we further restrict our sample to only include men and women who were between 40 and 45 at the time of the hospital admission and define individuals with at least one child at the age of 40 as having a household. The advantage of this definition is that most individuals have completed their fertility by the age of 40 so that few will change family status after their hospitalization.

[10] An individual is defined as employed if he or she had earnings in the tax register in November 1992.

With the above restrictions our final sample consists of roughly 63,000 employed men and women with at least one hospital admission between 40 and 45.[11] To get an idea of how much the sample restrictions limit the inferences that can be drawn to the total population of working individuals, we construct a representative sample of individuals with the same age distribution during the same time period, but without a record of a hospital admission.[12] Table A.3 presents descriptive statistics of our analysis sample and the age-matched comparison sample. While the slightly higher wages and lower sickness absence for the comparison sample provides an indication that our analysis sample are in a somewhat poorer health condition, these differences are surprisingly small. Based on these descriptive statistics we conclude that the analysis sample is roughly similar to the comparison sample with regard to key characteristics such as observed health status and income.

5 Results

This section begins with a presentation of the gender analysis, which is subsequently followed by the household analysis as outlined in section three. The section concludes with a discussion of the results from a number of sensitivity analyses.

5.1 Gender differences in sickness absence and health

We first present the results from the relative gender response from the hospitalization on sickness absence. This analysis is then followed by a discussion of the results from the evaluation of the relative post-hospitalization health outcomes.

Sickness absence

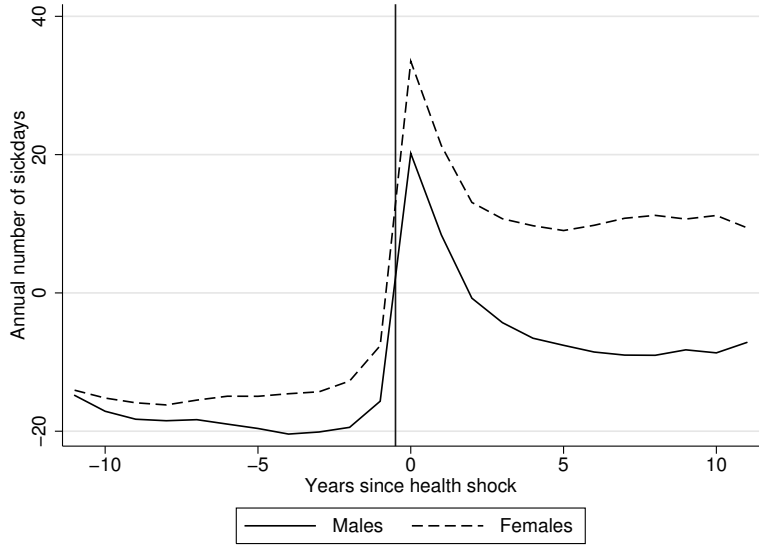
Before turning to the formal regression results, we first present some graphical evidence of the relationship between gender and sickness absence in relation to the hospitalization. Figure 5.1 plots the average number of sick days for men (solid line) and women (dashed line) by years from the hospital admission. The figure clearly shows that women have on average more sick days

[11] The original population of all labor market active individuals between 20-50 in 1992 amounted to 2,587,580 individuals. After taking a random sample of 40 percent (leaving 1,035,032 individuals) and conditioning on individuals having been hospitalized between 1993-2004 (leaving 470,587 individuals), aged between 40 and 45 at the time of hospitalization (leaving 75,880 individuals) and finally removing individuals with labor market interruptions in the years before the hospital admission we are left with 63,599 individuals in our analysis sample. See Table A.2 in Appendix A for descriptive statistics of the sample and the included variables.

[12] Specifically, each sampled individual from our analysis sample (i.e. with a hospital admission record) is matched with a person of the same age without a hospital admission record in the age range 40-45.

both before — but in particular after — the hospitalization compared to men. This observation is thus an initial indication that men and women exert different absence behavior after the hospitalization spell. The figure also reveal that the gender absence gap increases somewhat in the years prior to the hospitalization. This pattern would be expected if women acted more proactively than men and reacted to a change in their health before this change generated the observed hospital admission. We discuss this potential problem below.

Figure 5.1. Annual days of sickness absence for men and women in relation to a hospital admission



NOTE.— The figure is constructed by plotting the residuals from an ordinary least squares regression of year and age fixed effects on days of sickness absence. Sickness absence is defined as the annual number of days on sickness absence. The vertical line indicates the time of the hospital admission.

Turning to results from estimation, Table 5.1 presents the estimated κ^* from equation (3.5) using the same observations making up Figure 5.1. Hence, the estimated parameter is the annual relative response in sickness absence of men and women over in total twelve years after the hospitalization. The estimated results largely confirm the gender pattern from Figure 5.1. Our preferred specification, given in column (3) of the table, reports that women have on average an additional twelve days of sickness absence after the hospitalization relative to men.[13] Note that the inclusion of control variables only marginally influences the estimated parameters. To the extent that the estimates change, however, the inclusion of fixed year and age effects and other control variables slightly *increases* the estimated relative response in sickness absence. Given that the inclusion of control variables captures gender differences in health,

[13] We have also performed analyses where we have estimated the effects two, four, six, eight and ten years after the hospital admission.

this pattern indicates that women have on average better pre-admission health than men.[14]

Table 5.1. *Estimated relative response of men and women on sickness absence from a hospital admission*

	(1)	(2)	(3)
Full sample	10.980***	11.401***	12.582***
	(0.319)	(0.318)	(0.317)
Diagnosis type			
Heart	7.727***	8.440***	9.328***
	(1.257)	(1.259)	(1.261)
Cancer	2.188	2.941**	4.019***
	(1.480)	(1.479)	(1.470)
Mental	17.799***	18.558***	21.251***
	(1.851)	(1.849)	(1.867)
Musculoskeletal	15.352***	17.153***	17.862***
	(1.438)	(1.431)	(1.435)
Industry Sector			
Manufacturing	9.798***	10.316***	11.100***
	(0.897)	(0.896)	(0.890)
Public	14.110***	14.513***	15.208***
	(1.097)	(1.097)	(1.095)
Education	13.667***	12.899***	12.672***
	(1.137)	(1.133)	(1.131)
Health	11.607***	12.045***	12.271***
	(0.966)	(0.965)	(0.959)
Controls		✓	✓
Fixed effects			✓

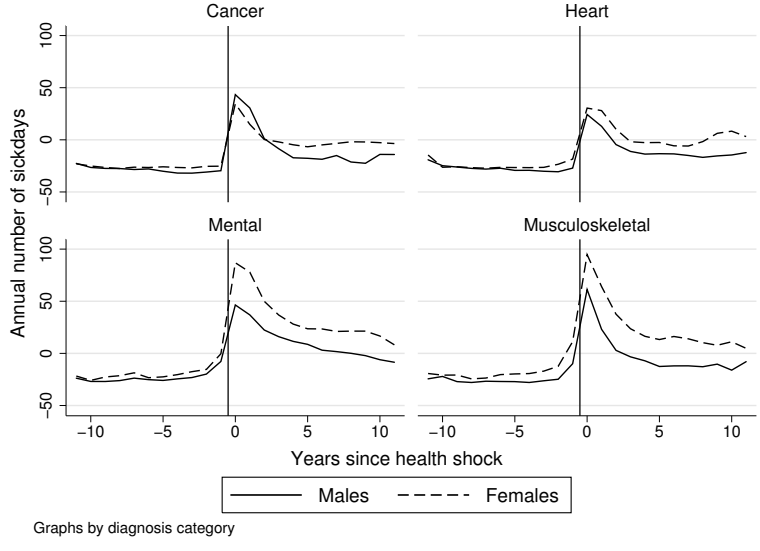
NOTE.— The table reports the estimated κ^* parameters (standard errors) of model (3.5) in the empirical section. Column (1) includes no controls while column (2) includes controls for own and spouse's annual earnings and after-tax income and indicators for being a high earner, being the primary earner of the household and having post-secondary education and column (3) additionally includes year, age, industry and diagnosis category fixed effects. Standard errors are estimated using a robust covariance matrix.*, ** and *** denote significance at the 10, 5 and 1 percent levels. See Table A.1 for detailed variable definitions.

So far, our results strongly suggest that women act in a more proactive manner than men by utilizing health insurance more when they experience a negative change in health. However, if the assumption that the hospitalizations of men and women in our sample are comparable is invalid, this result may simply be the product of gender differences in the severity of the illness affecting the individuals. While the inclusion of fixed effects for diagnosis type and labor market sector left the results qualitatively unchanged, heterogeneous responses may still be prevalent across these categories. In particular, Table A.4 and Table A.5 demonstrate that men and women in our sample in general suffer from different kinds of illnesses as well as work in different industries. The range is substantial; from a female share of 0.84 for cancer diseases to only 0.37 for heart diseases and from a share of 0.87 employed in the health sector to only 0.09 employed in the construction sector.

[14] See Appendix B for an informal proof of this proposition.

Figure 5.2 plots the sickness absence of men and women before and after the hospitalization for the four most common diseases in our data; cancer, heart, mental and musculoskeletal diseases. Interestingly, the pre-admission trends are essentially parallel for all categories, suggesting that the gender difference in the aggregate pre-admission trend is to some extent driven by gender differences in diagnosis category. The sickness absence of men is higher in the first year after admission for cancer patients but falls below that of women in subsequent years. For all other categories the post-admission pattern closely follows the aggregate results.

Figure 5.2. Annual days of sickness absence for men and women in relation to a hospital admission, by diagnosis type



NOTE.— Each panel pertains to a specific disease category. The figures are constructed by plotting the residuals from an ordinary least squares regression of year and age fixed effects on days of sickness absence. Sickness absence is defined as the annual number of days on sickness absence. The vertical line indicates the time of the hospital admission.

Results from separate estimations of model (3.5) conditional on diagnosis category and on four different labor market sectors (manufacturing, public administration, education and health) are displayed in rows 2-9 of Table 5.1. What is noteworthy from the table is the consistent pattern of a relative increase in female absenteeism irrespective of analyzed category and the inclusion of controls. In fact, including control variables slightly increases the estimated relative difference in all sub-analyses, indicating that women also *within* each diagnosis and labor market sector category on average were in a better health condition than men at the time of the hospitalization.

Interestingly, from rows 2-5 of Table 5.1 we find a large spread in the estimated relative sickness absence response across different diagnosis categories. In particular, a woman with a cancer diagnosis has roughly four more days of

sickness absence after the admission than a man with a cancer diagnosis. In contrast, a woman with a mental disease has on average over twenty days more of sickness absence than a man with the same diagnosed condition. For the latter, more vaguely defined, diagnosis it is likely that the resulting sickness absence spell would be more affected by the patient's own interpretation — and less by the physician's assessment — of the illness (see e.g. Englund (2001)). Hence, this result provides further evidence that men and women reacted differently in terms of absenteeism in relation to the hospitalization.

Finally, the results from separate estimations of the four labor market sectors are displayed in rows 6-9. The results are remarkably consistent across sectors and provide supporting evidence that the difference in sickness absence response of men and women is not driven by the gender-segregated labor market in Sweden.

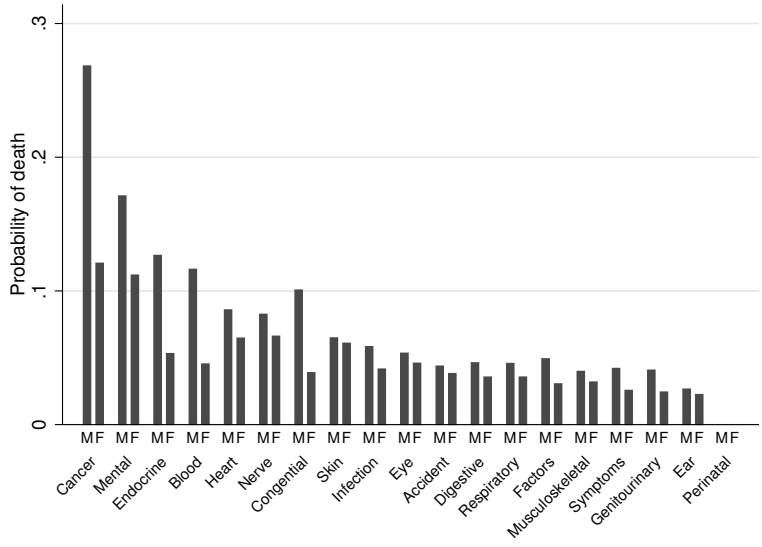
Post-hospitalization health outcomes

The results of the previous section showed a relative increase in women's sickness absence after a hospital stay compared with men. However, this result may result from that men and women are affected differently by similar health shocks, even within diagnostic categories, and thus the question remains whether this difference in sickness absence response is driven by poorer health among women rather than gender differences in health behavior.

As a starting point for analyzing post-admission health, Figure 5.3 present sample statistics of the fraction of men and women in our sample who died within three years after the hospitalization by diagnosis category. Figure 5.4 further disaggregates the mortality data by family status. The resulting pattern is unambiguous; men have a higher mortality rate for all diagnosis categories. In particular, the mortality rate for men is more than twice as high as that of women for cancer diseases (0.27 compared to 0.12) and even higher when comparing men and women without children (0.41 compared to 0.13). Since cancer is a disease for which the mortality risk is strongly correlated with the time of detection (see e.g. Levin *et al.* (2008), Brett (1969) and UK trial of early detection of breast cancer group (1988)), this massive gender difference in cancer mortality provides a hint of how a more preventative health behavior among men could enhance their longevity. This difference is also likely a consequence of the since 1986 ongoing public breast cancer screening program for women in Sweden.[15] Finally, the large gender difference in post-admission cancer mortality rate may also explain some of the relatively small estimated differences in sickness absence between men and women diagnosed with cancer in the previous section.

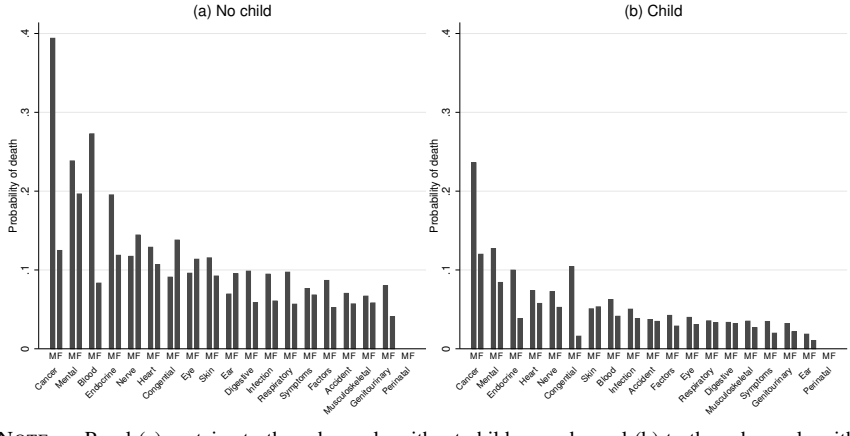
[15]There is an important discussion questioning the value of screening for breast cancer today (see e.g. McPherson (2010)). This criticism should, however, not be interpreted as criticism of the value of early detection. Today most women are aware of the risk of breast cancer and of the possibilities of self-screening which was not the case almost 30 years ago when screening was introduced.

Figure 5.3. Mortality risk after a hospital admission, by gender and diagnosis type



NOTE.— Mortality risk is defined as the fraction of individuals who deceased within three years after the hospital admission. The bins in the figure pertains to (F)emales and (M)ales for each diagnosis category.

Figure 5.4. Mortality risk after a hospital admission, by gender, household status and diagnosis type



NOTE.— Panel (a) pertains to the subsample without children and panel (b) to the subsample with children. Mortality risk is defined as the fraction of individuals who deceased within three years after the hospital admission. The bins in the figure pertains to (F)emales and (M)ales for each diagnosis category.

Next, Table 5.2 reports the results from estimating the Cox regression model (3.6) with exits to death and a second hospital admission.[16] Column (1) of the

[16]We have also estimated linear probability models for the same outcomes occurring within two, three etc. years after the hospital admission. The results are qualitatively similar to the Cox regression model.

table reports results without including controls while column (2) includes the full set of controls. As in Table 5.1, the first row of Table 5.2 shows the estimation results for the full sample while rows 2-4 (5-8) display the results when estimating the model separately for the four different diagnosis categories (industries). The reported coefficient is the estimated mortality risk of women relative to men, i.e. $\hat{\delta}_1$ from model (3.6).

Table 5.2. *Estimated relative post-hospitalization mortality and re-admission risks of men and women*

	Mortality		Re-admission	
	(1)	(2)	(3)	(4)
Full sample	-0.211***	-0.326***	0.049***	0.003
	(0.046)	(0.048)	(0.012)	(0.013)
Diagnosis type				
Heart	-0.590***	-0.691***	-0.042	-0.097**
	(0.193)	(0.197)	(0.043)	(0.045)
Cancer	-0.951***	-1.099***	-0.293***	-0.380***
	(0.090)	(0.094)	(0.051)	(0.054)
Mental	-0.521***	-0.498***	0.022	0.011
	(0.138)	(0.142)	(0.045)	(0.046)
Musculoskeletal	-0.264	-0.372	0.273***	0.220***
	(0.221)	(0.228)	(0.043)	(0.045)
Industry Sector				
Manufacturing	-0.155	-0.295**	0.071**	0.019
	(0.124)	(0.126)	(0.031)	(0.032)
Public	-0.337*	-0.443**	0.061	-0.005
	(0.188)	(0.196)	(0.046)	(0.049)
Education	-0.031	-0.109	0.064	0.034
	(0.185)	(0.189)	(0.048)	(0.049)
Health	-0.205	-0.272*	-0.002	-0.097**
	(0.157)	(0.164)	(0.037)	(0.039)
Controls		✓		✓
Fixed effects		✓		✓

NOTE.— The table reports the estimated δ_1 parameters (standard errors) of model (3.6) in the empirical section for different subsamples. Estimation is performed under the assumption of a Cox proportional hazards model using an exact maximum likelihood estimator. Columns (2) and (4) includes controls for own and spouse's annual earnings and after-tax income and indicators for being a high earner, being the primary earner of the household and having post-secondary education and year, age, industry and diagnosis category fixed effects. *, ** and *** denote significance at the 10, 5 and 1 percent levels. See Table A.1 for detailed variable definitions.

From the first row of column (1) we find that women have a lower estimated mortality risk, but a higher risk of being re-admitted to a hospital after the hospitalization. However, when adding controls to the models, the effect on mortality increases in magnitude while the re-admission risk turns insignificant and indistinguishable from zero. Even though we include a rich set of control variables, it is nevertheless likely that some important aspects of health are omitted in the estimation. However, the implication from the coefficient pattern when adjusting for observable characteristics is that, if anything, the estimates in columns (2) and (4) should still be biased towards zero. Thus,

taken together we interpret the results from this section as establishing that women have on average better health than men in our sample, not only before, but also subsequent to the hospitalization.

The estimation results conditional on diagnosis category (rows 2-5) are in general consistent with the aggregate results. The negative effect on mortality is, as expected, greater and the inclusion of control variables increases (decreases) the magnitude of the negative (positive) estimates. Interestingly, women have an estimated lower, albeit insignificant, mortality risk for musculoskeletal diseases but at the same time a significantly greater re-admission risk, even when adjusting for observable characteristics. Since musculoskeletal diseases are, in general, more of a symptom diagnosis than e.g. cancer, we may interpret the increased relative re-admission risk for this diagnosis more as a consequence of preventive behavior among women rather than an indication of their poorer relative post-admission health. Finally, the results from the within-industry regressions (rows 6-9) are estimated with lower precision but the reported point estimates are generally in line with the aggregate results.

5.2 Household differences in sickness absence and health

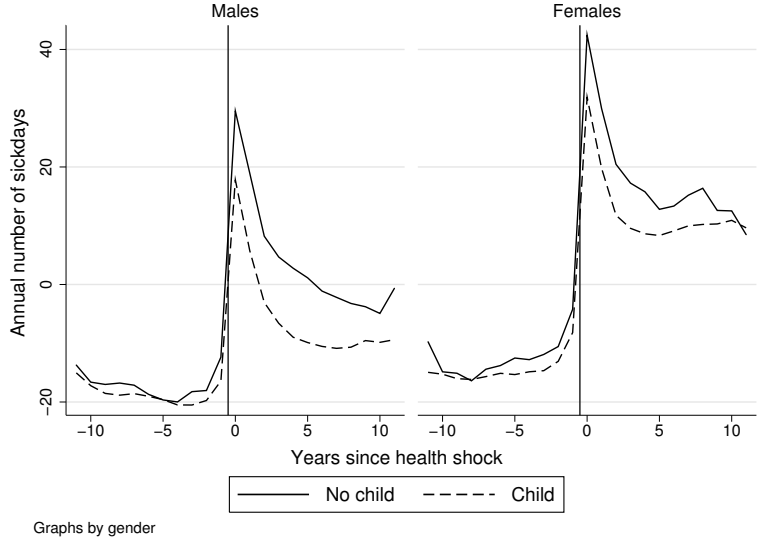
As in the former subsection we first present the estimation results from the analysis on sickness absence and thereafter the results from the evaluation of post-hospitalization relative health outcomes. Finally, we discuss the results from omitting the cancer diagnosis category from the analysis.

Sickness absence

Before discussing the estimation results, we provide some initial graphical evidence. Figure 5.5 illustrates the average annual number of sick days for men (left panel) and women (right panel) with (dashed line) and without children (solid line) in relation to the hospitalization. The figure shows that individuals with children are on average less absent for health reasons both before and after the hospitalization. This pattern was expected given the descriptive sample statistics in Table A.3. The figure further reveals that, while the increase in sickness absence after the admission is greater for individuals without children, this increase is less pronounced for women. Hence, this descriptive analysis provides some preliminary evidence in support of both the health investment hypothesis of Paringer (1983) and the double burden hypothesis of Bratberg *et al.* (2002); i.e., women with children have relatively more days of absence than women without children following a change in individual health.

The estimation results from two different regression models using the same observations making up Figure 5.5 are displayed in panel A of Table 5.3. The coefficients reported in columns (1) through (6) of the table are results from estimating regression models separately for each panel of Figure 5.5, while the results in column (7) refer to results from estimating equation (3.7) in which

Figure 5.5. Annual days of sickness absence in relation to a hospital admission, by gender and household status



NOTE.— The left panel pertains to the results for men and the right panel to the results for women. The figures are constructed by plotting the residuals from an ordinary least squares regression of year and age fixed effects on days of sickness absence. Sickness absence is defined as the annual number of days on sickness absence. The vertical line indicates the time of the hospital admission.

the right panel of Figure 5.5 is subtracted from the left. We show in Appendix B that the estimated effects in columns (1) through (6) are biased downwards if the health of hospitalized women (men) with children is better than the health of hospitalized women (men) without children.

In line with the coefficient pattern observed in Figure 5.5, estimates from Table 5.3 report that the sickness absence response from the hospitalization is smaller for individuals with children, irrespective of sex. Furthermore, while men with children increase their absence between five to seven days less relative to men without children, the corresponding difference for females is only between one and three days. The estimated coefficients in the regressions are lower when control variables are included in the model, suggesting that health of individuals with children is better than for individuals without children (see Appendix B). Finally, the results from estimating equation (3.7), reported in column (7) of the table, similarly show that women with children increase their absenteeism by on average 4.6 days more than women without children when also controlling for the general effect of having children. Relating this result to the results from the gender analysis in the last section, we may conclude that women with children are responsible for roughly one-third of the excess gender response in absenteeism.

Results from estimation conditional on each of the four selected diagnosis categories are displayed in rows 2-5 in Table 5.3. These estimations are ac-

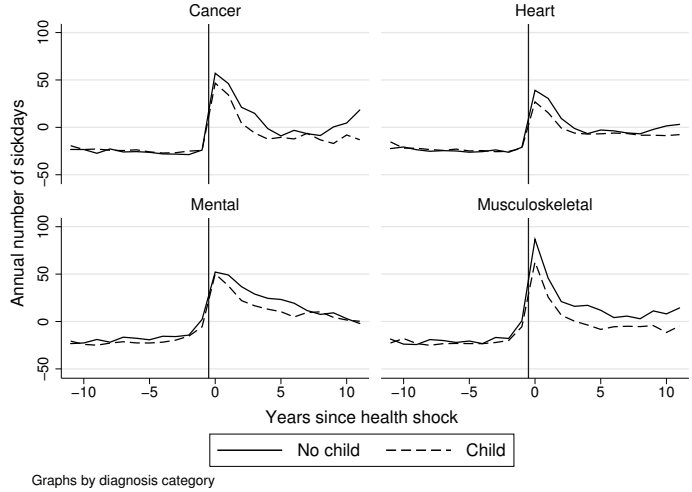
companied by Figures 5.6–5.7, which plot the corresponding average annual number of sick days over time from the hospitalization by diagnosis category, gender and family status. The absence pattern for men is much in line with the aggregate results while the pattern is somewhat more heterogeneous for women. For cancer or a mental illness, women with children increase their absence more than women without children, which is also true with respect to the long-run effect for women with a musculoskeletal diagnosis. Once again, the results from estimation correspond closely to the pattern illustrated in the corresponding figures and are broadly consistent with the estimates from the aggregate analysis. It is interesting to note that the differences in sickness absence again increase with the vagueness of the diagnosis. This suggests that any potential health investment effect is greatest in diagnoses where the individual freedom in deciding whether or not to be absent from work is greatest, i.e. for mental and musculoskeletal diseases.

Table 5.3. *Estimated relative sickness absence response from a hospital admission, by gender and household status*

	Males			Females			Both genders
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. All diagnoses included							
Full sample	-7.497***	-7.641***	-5.163***	-2.754***	-2.859***	-1.020	4.622***
	(0.541)	(0.540)	(0.542)	(0.729)	(0.728)	(0.727)	(0.905)
Diagnosis type							
Heart	-5.989***	-6.087***	-4.638***	-5.917*	-6.049*	-6.045*	-0.694
	(1.712)	(1.712)	(1.732)	(3.202)	(3.177)	(3.195)	(3.636)
Cancer	-12.274***	-12.752***	-10.335***	1.669	1.675	2.810*	13.511***
	(3.753)	(3.732)	(3.736)	(1.520)	(1.516)	(1.519)	(4.035)
Mental	-1.289	-3.322	-0.657	13.683***	10.701***	9.827***	14.362***
	(2.451)	(2.453)	(2.493)	(3.542)	(3.552)	(3.626)	(4.315)
Musculoskeletal	-11.917***	-12.775***	-10.383***	11.703***	9.777***	10.049***	23.678***
	(2.424)	(2.412)	(2.401)	(3.695)	(3.652)	(3.664)	(4.382)
Panel B. Excluding Cancer diagnoses							
Full sample	-7.373***	-7.496***	-4.888***	-3.345***	-3.476***	-1.507*	4.101***
	(0.548)	(0.547)	(0.546)	(0.830)	(0.828)	(0.826)	(0.987)
Industry Sector							
Manufacturing	-8.284***	-8.359***	-8.129***	-5.242**	-4.828**	-5.041**	2.621
	(1.004)	(0.999)	(0.991)	(2.324)	(2.321)	(2.309)	(2.502)
Public	-8.332***	-8.876***	-7.322***	3.142	3.459	4.421	12.138***
	(1.980)	(1.994)	(1.974)	(2.881)	(2.880)	(2.875)	(3.473)
Education	-6.475**	-6.195**	-4.961*	-2.194	-2.675	-4.103	1.484
	(2.588)	(2.580)	(2.599)	(2.859)	(2.856)	(2.862)	(3.868)
Health	-6.315***	-6.603***	-7.054***	-10.112***	-10.149***	-9.333***	-2.559
	(2.392)	(2.385)	(2.374)	(1.503)	(1.499)	(1.490)	(2.808)
Controls		✓	✓		✓	✓	✓
Fixed effects			✓			✓	✓

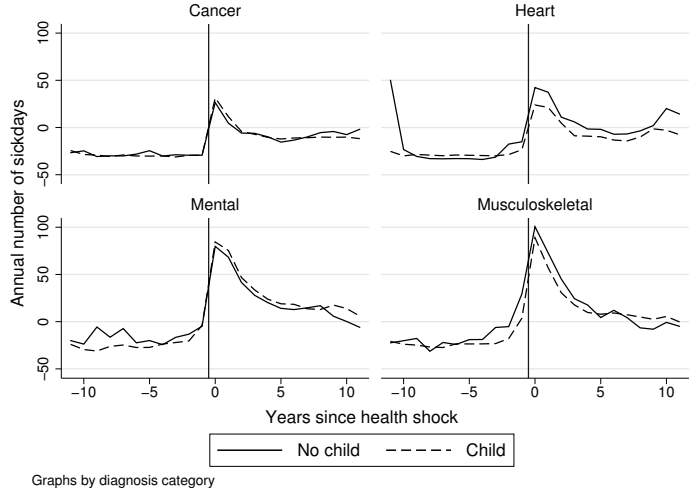
NOTE.— The table reports the estimated κ^* and δ^* parameters (standard errors) of models (3.5) and (3.7) in the empirical section, for different samples. Columns (1)-(4) pertains to estimating model (3.5) on the subsample of men (columns (1)-(2)) and on the subsample of women (columns (3)-(4)) while columns (5)-(6) pertains to estimating model (3.7) on the full analysis sample. Controls include own and spouse's annual earnings and after-tax income and indicators for being a high earner, being the primary earner of the household and having post-secondary education and fixed effects include year, age, industry and diagnosis specific effects. Standard errors are estimated using a robust covariance matrix.*, ** and *** denote significance at the 10, 5 and 1 percent levels. See Table A.1 for detailed variable definitions.

Figure 5.6. Annual days of sickness absence for *men* in relation to a hospital admission, by household status and diagnosis category



NOTE.— Each panel pertains to a specific disease category. The figures are constructed by plotting the residuals from an ordinary least squares regression of year and age fixed effects on days of sickness absence. Sickness absence is defined as the annual number of days on sickness absence. The vertical line indicates the time of the hospital admission.

Figure 5.7. Annual days of sickness absence for *women* in relation to a hospital admission, by household status and diagnosis category



NOTE.— Each panel pertains to a specific disease category. The figures are constructed by plotting the residuals from an ordinary least squares regression of year and age fixed effects on days of sickness absence. Sickness absence is defined as the annual number of days on sickness absence. The vertical line indicates the time of the hospital admission.

Post-hospitalization health outcomes

Results from the estimation of Cox hazard models in order to analyze relative post-admission objective health measures for individuals with and without children are reported in Table 5.4. The estimates from our preferred model specification (3.8) are displayed in columns (5) and (6) and, for completeness, results from separate estimations for men and women are also presented in columns (1)-(4) of the table.

The estimated coefficients reported in the first row and columns (1) and (3) of panel A show a lower estimated mortality risk for both men and women with children. Specifically, the annual relative difference in mortality risk is approximately 36 percent lower for women and approximately 56 percent lower for men relative to individuals without children, respectively.[17] Consequently, from the results reported in column (5), where the sample of men is included in order to capture health differences by family status of the women, we also find an *increased* mortality risk for women with children of about 33 percent. This relation is also reflected in the results for the risk of being re-admitted to a hospital. Taken together, these results hence suggest a relatively poorer level of health for women with children at the time of the hospitalization, which would then lead us to reject the health investment hypothesis in favor of the double burden hypothesis.

[17] This risk reduction is obtained as $100*(\exp(\text{coefficient}) - 1)$.

Table 5.4. *Estimated relative post-hospitalization mortality and re-admission risks, by gender and household status*

	Males		Females		Both genders	
	Mortality (1)	Re-admission (2)	Mortality (3)	Re-admission (4)	Mortality (5)	Re-admission (6)
Panel A. All diagnoses included						
Full sample	-0.894***	-0.188***	-0.484***	-0.123***	0.416***	0.066**
	(0.065)	(0.021)	(0.079)	(0.022)	(0.102)	(0.030)
Diagnosis type						
Heart	-0.673***	-0.115*	-0.161	-0.121	0.555	-0.006
	(0.207)	(0.062)	(0.424)	(0.091)	(0.468)	(0.110)
Cancer	-0.687***	-0.099	0.111	0.071	0.791***	0.183
	(0.162)	(0.112)	(0.130)	(0.056)	(0.208)	(0.125)
Mental	-0.587***	-0.296***	-0.432*	-0.215***	0.102	0.089
	(0.166)	(0.062)	(0.238)	(0.075)	(0.286)	(0.096)
Musculoskeletal	-1.026***	-0.181**	-0.752*	-0.239***	0.367	-0.060
	(0.292)	(0.079)	(0.384)	(0.078)	(0.475)	(0.110)
Panel B. Excluding Cancer diagnoses						
Full sample	-0.953***	-0.191***	-0.779***	-0.163***	0.165	0.028
	(0.072)	(0.021)	(0.101)	(0.024)	(0.123)	(0.032)
Industry Sector						
Manufacturing	-0.763***	-0.173***	-0.626**	-0.160**	0.019	0.001
	(0.148)	(0.041)	(0.305)	(0.071)	(0.334)	(0.082)
Public	-0.862***	-0.131	-1.126***	-0.263***	-0.287	-0.134
	(0.318)	(0.088)	(0.373)	(0.085)	(0.486)	(0.122)
Education	-0.844**	-0.172*	-0.915**	-0.244***	-0.008	-0.066
	(0.349)	(0.102)	(0.362)	(0.082)	(0.498)	(0.130)
Health	-0.814**	-0.167**	-0.440**	-0.146***	0.424	0.062
	(0.333)	(0.081)	(0.223)	(0.042)	(0.397)	(0.090)
Controls		✓		✓		✓
Fixed effects		✓		✓		✓

NOTE.— The table reports the estimated δ_1 and δ_3 parameters (standard errors) of models (3.6) and (3.8) in the empirical section for different subsamples. Estimation is performed under the assumption of a Cox proportional hazards model using an exact maximum likelihood estimator. Columns (1)-(6) pertains to estimating model (3.6) on the subsample of men (columns (1)-(3)) and on the subsample of women (columns (4)-(6)) while column (7) pertains to estimating model (3.8) on the full analysis sample. Controls include own and spouse's annual earnings and after-tax income and indicators for being a high earner, being the primary earner of the household and having post-secondary education and fixed effects include year, age, industry and diagnosis specific effects. Standard errors are estimated using a robust covariance matrix.*, ** and *** denote significance at the 10, 5 and 1 percent levels. See Table A.1 for detailed variable definitions.

Results from excluding cancer diagnoses

From observing the results conditional on disease category, it is clear that the inference in the last subsection is primarily driven by the subsample of cancer patients. However, men and women suffer different forms of cancer and, for this reason, men admitted to hospital for cancer may not be a valid control for unobserved health differences of women with respect to family status. This empirical issue is further complicated by the vast difference in post-hospitalization mortality rates we observed for men with and without children in Figure 5.4. One reason for these large mortality differences across gender and family status might be the ongoing (cervical and breast) public cancer screening programs in Sweden which only targets women. While the beneficial health effect of cancer screening is clear, its focus on women and the potential spill-over effect it might have on men with families (through their spouses) would imply that only men without families are left out from this policy. Using men as controls for the counter-factual effect of having children is hence likely to violate our identifying assumption, in particular with respect to the case of cancer patients. For this reason, we extend our analysis by presenting results after excluding individuals with a cancer diagnosis from the sample.

Panel B in Table 5.4 reports results for post-hospitalization health after cancer patients have been excluded from the analysis. From columns (1) and (3) we, once again, observe an estimated lower mortality risk for individuals with children. The yearly relative mortality risk is approximately 61 and 54 percent lower for men and women, respectively. When men are included as controls for unobserved differences in health across family status, the estimated difference in post-hospitalization mortality for women with household responsibility decreases sharply in magnitude and becomes insignificant — while still being precisely estimated. Furthermore, a similar pattern is found for the re-admission risk health outcome. These results are also largely robust across the four industry sectors in rows 2-5 of panel B. Hence, family status among women is no longer associated with poorer post-hospitalization health when we exclude individuals who were hospitalized with a cancer diagnosis.

As a next step, we investigate whether the results on sickness absence changes when we exclude cancer patients from the analysis. These results are reported in panel B of Table 5.3 and are in all relevant aspects similar to the results for the full sample reported in panel A. The results by industry sector from rows 2-5 in panel B also supports this interpretation. Three out of four sectors report positive point estimates, but, due to lower precision from smaller sample sizes, only the result for the public sector is statistically significant.

5.3 Robustness checks

We have performed a number of robustness checks in order to validate our main results, most of which are discussed at length in the previous sections. In this section we give only a brief summary of the analyses made and the reported results.

First, we re-estimated the all models restricting the outcome variables to include only observations less than or equal to two, four, six, eight and ten years subsequent to the hospitalization (compared to twelve years in the main specification). These results (not reported here) are very similar to the results from our main specifications. If anything, the effects on sickness absence increase slightly with the time window; for the gender analysis the difference in sickness absence increases from 8.9 days two years after the admission to 12.6 days ten years after the admission. The corresponding numbers are 3.2 and 3.9 days of sickness for the household analysis.

Second, we re-estimated the Cox hazard regression models where we additionally controlled for the length of sickness absence and diagnosis category using stratified Cox regressions. The argument underlying this robustness check is the following: under the null hypothesis of no difference in absence behavior of men and women, the length of sickness absence and the diagnosis category are valid measures of pre-hospitalization health status of the individuals. Consequently, we should not find any differences in mortality and readmission once we control for these factors. Interestingly, estimation results from the stratified models provide the same inference as in our main specification (not reported here). In addition, the proportional hazards assumption in the Cox regression models could not be rejected for our data.[18]

We have also evaluated the sensitivity of the results by varying the definition of the outcome variables. First, the hospitalization is likely to be an ex-post measure of the actual health change observed by the individual. As a result, any health prevention behavior reflected in sickness absence could in many cases have begun before the hospitalization occurred. We investigate this phenomenon by lagging the hospital admission one and two years back in time. Second, we have estimated the effects on sickness absence prevalence. We define prevalence as a binary variable, assuming a value of one if the individual received any sickness benefits during the year and zero otherwise. Third, we have included individuals receiving disability benefits by imputing their annual sickness absence to 365 days. Finally, we have retained all deceased individuals in the estimation by setting their sickness absence to zero for all years subsequent to their death.

Table 5.5 reports results from the estimation of the same sickness absence regression models as previously, but where the hospitalization has been lagged with one and two years. We have also reported the main results in columns

[18] The key assumption of proportionality is tested by analyzing the Schoenfeld residuals following the generalization by Grambsch and Therneau (1994).

Table 5.5. *Robustness checks I: Specifying different times until the onset of the health shock*

	Gender differences model			Household differences model		
	Years from health shock					
	0 years	1 year	2 years	0 years	1 year	2 years
	(1)	(2)	(3)	(4)	(5)	(6)
Full sample	12.582***	11.321***	11.502***	4.101***	3.349***	4.192***
	(0.317)	(0.386)	(0.440)	(0.905)	(1.124)	(1.291)
Diagnosis code						
Heart	9.328***	7.079***	7.846***	-0.694	-4.536	-6.609
	(1.261)	(1.602)	(1.851)	(3.636)	(4.814)	(5.620)
Cancer	4.019***	4.060**	1.904	13.511***	10.160*	12.492**
	(1.470)	(1.916)	(2.220)	(4.035)	(5.227)	(6.102)
Mental	21.251***	14.856***	18.418***	14.362***	14.801**	9.081
	(1.867)	(2.470)	(2.923)	(4.315)	(5.878)	(7.235)
Musculoskeletal	17.862***	16.549***	16.975***	23.678***	18.413***	21.155***
	(1.435)	(1.814)	(2.091)	(4.382)	(5.498)	(6.441)
Industry Sector						
Manufacturing	11.100***	7.867***	6.797***	2.621	1.665	1.895
	(0.890	(1.051)	(1.189)	(2.502)	(2.857)	(3.178)
Public	15.208***	11.954***	14.491***	12.138***	9.484**	4.858
	(1.095)	(1.352)	(1.540)	(3.473)	(3.877)	(4.481)
Education	12.672***	12.962***	15.504***	1.484	5.857	6.521
	(1.131)	(1.387)	(1.544)	(3.868)	(4.439)	(3.970)
Health	12.271***	9.726***	9.733***	-2.559	-0.259	6.538*
	(0.959)	(1.173)	(1.338)	(2.808)	(3.402)	(3.970)

NOTE.— The table reports the estimated κ^* and δ^* parameters (standard errors) of models (3.5) and (3.7) in the empirical section for different samples when artificially moving the timing of the hospital admission 0-2 years back in time. Columns (1)-(3) pertains to estimation of model (3.5) and columns (4)-(6) pertains to estimation of model (3.7), respectively. All specifications include controls for own and spouse's annual earnings and after-tax income and indicators for being a high earner, being the primary earner of the household and having post-secondary education and year, age, industry and diagnosis category fixed effects. Standard errors are estimated using a robust covariance matrix.*, ** and *** denote significance at the 10, 5 and 1 percent levels. See Table A.1 for detailed variable definitions.

(1) and (4) in the table for comparison. The results for the gender analysis are remarkably stable while the results from the household analysis are slightly less stable, due to more imprecisely estimated parameters. The general pattern remains the same, however.

The estimated effects for the gender and household analyses using the continuous outcome are presented in columns (1) and (3) of Table 5.6, respectively. The corresponding gender and household estimates for the prevalence outcome measure are reported in columns (2) and (4) of the same table. For ease of comparison, we also report the results from the baseline model in the first row of the table. It is clear from the first row that inference is not contingent on the type of sickness absence outcome measure we use. Turning to the second row, which reports results from including disability beneficiaries, the estimated gender difference in absence response increases for both outcomes. With respect to the household analysis, however, inference from using the

Table 5.6. *Robustness checks II: Varying the outcome measure and including deceased and disabled individuals*

	Gender differences model		Household differences model	
	Days of absence (1)	Prevalence (2)	Days of absence (3)	Prevalence (4)
Baseline	12.582***	0.045***	4.101***	0.025***
	(0.336)	(0.002)	(0.987)	(0.006)
Include disability	24.743***	0.067***	1.744	0.020***
	(0.443)	(0.002)	(1.380)	(0.006)
ITT	12.754***	0.046***	3.968***	0.024***
	(0.333)	(0.002)	(0.969)	(0.006)
ITT and disability	25.005***	0.068***	1.441	0.018***
	(0.440)	(0.002)	(1.366)	(0.006)

NOTE.— The table reports the estimated κ^* and δ^* parameters (standard errors) of models (3.5) and (3.7) in the empirical section for different samples and outcomes. Columns (1)-(2) pertains to estimation of model (3.5) and columns (3)-(4) pertains to estimation of model (3.7), respectively. Days of absence is defined as the annual number of days on sickness absence for an individual and prevalence is defined an indicator variable for whether the individual was sick during the year of not. The 'Baseline' row pertains to the main results from Table 5.1 and Table 5.3, respectively. The 'Include disability' row includes individuals who were on disability in addition to individuals on sickness absence in the outcome measure (sickness absence is set to 365 days for disabled individuals), the 'ITT' row also includes deceased individuals in the years after their death in the sample and the 'ITT and disability' row includes both disabled and deceased individuals in the estimation. All specifications include controls for own and spouse's annual earnings and after-tax income and indicators for being a high earner, being the primary earner of the household and having post-secondary education and year, age, industry and diagnosis category fixed effects. Standard errors are estimated using a robust covariance matrix.*, ** and *** denote significance at the 10, 5 and 1 percent levels. See Table A.1 for detailed variable definitions.

continuous outcome is no longer statistically significant. This result suggests that women without children have a higher probability to enter the disability insurance program after the hospitalization, relative to women with children. In the last two rows of the table, we present the results for the intent-to-treat analysis, i.e. where we set the absence of deceased individuals to zero. From these two rows it is clear that inferences are not affected by the inclusion of deceased individuals.[19]

In conclusion, the results for the gender analysis are robust to all robustness checks we have performed in this section. Results for the household analysis are somewhat more mixed, mainly due to more imprecisely estimated parameters. In particular, including disability beneficiaries in the measure of sickness absence reduces the magnitude of the estimated effect making it only statistically significant for the prevalence outcome.

6 Summary and concluding remarks

Women are on average more absent from work for health reasons than men in most developed countries. Surprisingly little research has been devoted to

[19] We have also estimated the same models keeping individuals with cancer from which we obtain qualitatively similar results.

explaining the reasons behind this phenomenon, however. The few studies that do exist have mainly focused on factors such as differences in economic incentives, health and labor market characteristics.

We take a different approach, focusing on the potential role of gender-specific health behaviors in explaining the persistent gender gap in absenteeism. In particular, we first ask whether a relatively more proactive and health-promoting behavior of women may serve as an explanation for the gender difference. In a subsequent second step, we investigate whether these preference differences can be attributed to utility-maximizing households in which the traditional household division of labor makes women's health more valuable compared to the health of their (male) spouse. In particular, as women have traditionally held dual roles as producers of both market and household goods, the potential benefit from investing in her health, e.g. through increased absenteeism (or recuperation), may be rational from the point of view of an optimizing household.

We utilize Swedish administrative data on sickness benefits, mortality and inpatient care to empirically test the hypothesis that the gender gap in absenteeism is driven by different health behaviors of men and women. We conjecture that sickness absence is partly determined by an individual's own subjective preferences for absenteeism, while data on mortality and hospitalizations are (more) objective measures of individual health. As differences in absenteeism between men and women may also arise from unobserved factors, such as labor market segregation and economic incentives, estimates based on simple covariate adjustments will generally be confounded. For this reason, we sample men and women who experienced an adverse health shock, measured by a hospital admission, and compare the relative change in sickness absence of men and women before and after the hospitalization. We show that, if the severity of the health shock causing a hospital admission is on average the same for men and women, we can interpret a relative gender difference in sickness absence response from the hospitalization as a behavioral effect.

We find that women respond more than men in terms of sickness absence after the hospitalization. This result is interpreted as that women in general display a more preventative behavior when they are affected by deteriorating health. In addition, when comparing the sickness absence of women with and without children, we find that approximately one-third of the excess gender response in sickness absence can be attributed to women with children, lending some support for the household investment hypothesis. By analyzing post-hospitalization objective health measures, we find that the increased sickness absence response is unlikely to be driven by more severe health shocks for women (with children). In particular, we find that women have lower risks of both mortality and a subsequent hospital re-admission than men after the hospitalization.

We perform an extensive number of sensitivity analyses to corroborate our main results. The results from the gender analysis are remarkably robust

across all specifications while the results for the household analysis are more mixed, mainly due to more imprecisely estimated parameters. In particular, including disability beneficiaries in the analysis reduces the magnitude of the relative absence response, rendering it only statistically significant when sickness absence is implemented as a prevalence measure.

In conclusion, the traditional morbidity-mortality paradox is commonly explained by the fact that women exert a more preventative and risk-aversive behavior than men. The results obtained in this study lend support to the conjecture that differences in health-related behavior of men and women are important in explaining the gender gap in absenteeism observed around the world. Furthermore, we also find that a non-trivial share of these preference differences can be attributed to household investments in women's health. We believe that the results obtained in this paper should be viewed as providing a more complex and multifaceted picture of the patterns of sickness absence and health across different demographic and socioeconomic groups. It would be a heroic task to attempt to isolate the effect of any particular health investment behavior to the gender difference in life expectancy. However, it is possible to argue in favor of this notion by noticing the negative correlation between the recent narrowing gender gap and the relative increase in averse health behavior among women, such as e.g. obesity and smoking, observed in, for example, the U.K. (see e.g. LSAP (2012)). While the use of sickness absence as a means of investing in health may not be on parity with potential effects of losing weight or quitting smoking, it is probably still worth taking seriously in our modern society where stress-related diseases are emerging as a major health concern.

References

ÅKERLIND, I., ALEXANDERSSON, K., HENSING, G., LEIJON, M. and BJURULF, P. (1996). Sex Differences in Sickness Absence in Relation to Parental Status. *Scandinavian Journal of Public Health*, **24** (1), 27–35.

ANGELOV, N., JOHANSSON, P., LINDAHL, E. and LINDSTRÖM, E.-A. (2011). *Kvinnors och Mäns Sjukfrånvaro*. Rapport 2011:2, The Institute for Labor Market Policy Evaluation (IFAU), Uppsala, Sweden.

ARRELÖV, B., EDLUND, C. and GOINE, H. (2006). Grindvakterna och Sjukförsäkringen — Samspel och Motspel. In E. Palmer (ed.), *SKA Projektet: Sjukförsäkring, Kulturer och Attityder*, *10*, Försäkringskassan, Stockholm, pp. 479–530.

BERTRAND, M. (2010). New Perspectives on Gender. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics 4b*, *17*, Elsevier Ltd., pp. 1545–1592.

BOOTH, A. and VAN OURS, J. (2005). *Hours of Work and Gender Identity: Does Part-Time Work Make the Family Happier?* Discussion Paper 5438, CEPR, London.

BOYE, K. (2008). *Happy hour? Studies on Well-Being and Time Spent on Paid and Unpaid Work*. Ph.D. thesis, Swedish Institute for Social Research, Dissertation Series no. 74.

BRATBERG, E., DAHL, S.-Å. and RISA, A. E. (2002). The Double Burden — Do Combinations of Career and Family Obligations Increase Sickness Absence Among Women? *European Sociologial Review*, **18** (2), 233–249.

BRETT, G. (1969). Earlier Diagnosis and Survival in Lung Cancer Study. *British Medical Journal*, **4** (5678), 260–262.

BROSTRÖM, G., JOHANSSON, P. and PALME, M. (2004). Economic Incentives and Gender Differences in Work Absence Behavior. *Swedish Economic Policy Review*, **11**, 33–63.

BURDA, M., HAMERMESH, D. and WEIL, P. (2008). The Distribution of Total Work in the EU and US. In T. Boeri, M. Burda and F. Kramarz (eds.), *Working Hours and Job Sharing in the EU and USA: Are Europeans Lazy? Or Americans Crazy?*, Oxford: Oxford University Press.

CROSON, R. and GNEEZY, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, **47** (2), 1–27.

ECKEL, C. and GROSSMAN, P. J. (2008). Men, Women and Risk Aversion: Experimental Evidence. In C. Plott and V. Smith (eds.), *Handbook of Experimental Economics Results vol 1*, *113*, Elsevier Ltd., pp. 1061–1073.

ENGLUND, L. (2001). *Förändringar i Distriktsläkares Sjukskrivningspraxis Mellan Åren 1996 och 2001 i ett Svenskt Landsting*. mimeo, Centrum för Klinisk Forskning Dalarna, Falun.

ENGSTRÖM, P. and JOHANSSON, P. (2012). The Medical Doctors as Gatekeepers in the Sickness Insurance? *Applied Economics*, **44** (28), 3615–3625.

EVANS, O. and STEPTOE, A. (2002). The Contribution of Gender-Role Orientation, Work Factors and Home Stressors to Psychological Well-Being and Sickness Absence in Male- and Female-Dominated Occupational Groups. *Social Science and Medicine*, **54** (4), 481–92.

EVERTSSON, M. and NERMO, M. (2007). Changing Resources and the Division of Housework: A Longitudinal Study of Swedish Couples. *European Sociological Review*, **23** (4), 455–470.

GRAMBSCH, P. M. and THERNEAU, T. M. (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika*, **81** (3), 515–526.

GRAUNT, J. (1662). *Natural and Political Observations Mentioned in a Following Index and Made Upon the Bills of Mortality*. London: Martin, Allestry and Dicas.

GROSSMAN, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, **80** (2), 223–255, the University of Chicago Press.

LAAKSONEN, M., MARTIKAINEN, P., RAHKONEN, O. and LAHELMA, E. (2008). Explanations for Gender Differences in Sickness Absence: Evidence from Middle-Aged Municipal Employees from Finland. *Occupational and Environmental Medicine*, **65** (5), 325–30.

LEE, C. (2010). Gender, Health and Health Behaviors. In J. Chrisler and D. R. McCreary (eds.), *Handbook of Gender Research in Psychology*, *20*, Springer Science, pp. 471–493.

LEVIN, B., LIEBERMAN, D. A., MCFARLAND, B., SMITH, R. A., BROOKS, D., ANDREWS, K. S., DASH, C., GIARDIELLO, F. M., GLICK, S., LEVIN, T. R., PICKHARDT, P., REX, D. K., THORSON, A. and WINAWER, S. J. (2008). Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline From the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer and the American College of Radiology. *CA: A Cancer Journal for Clinicians*, **58** (3), 130–160.

LSAP (2012). *Life Expectancy: Past and Future Variations by Gender in England and Wales*. Tech. rep., Longevity Science Advisory Panel, http://www.longevitypanel.co.uk/docs/life-expectancy-by-gender.pdf.

MASTEKAASA, A. (2000). Parenthood, Gender and Sickness Absence. *Social Science and Medicine*, **50** (12), 1827–1842.

— and OLSEN, K. (1998). Gender, Absenteeism and Job Characteristics: A Fixed Effects Approach. *Work and Occupations*, **25** (2), 195–2228.

— and — (2000). Do Women or Men Have the Less Healthy Jobs? An Analysis of Gender Differences in Sickness Absence. *European Sociological Review*, **16** (3), 267–86.

NATHANSON, C. (1975). Illness and the Feminine Role: A Theoretical Review. *Social Science and Medicine*, **9** (2), 57–62.

PARINGER, L. (1983). Women and Absenteeism: Health or Economics? *American Economic Review*, **73** (2), 123–127, Papers and Proceedings of the Ninety-Fifth Annual Meeting of the American Economic Association.

MCPHERSON, K. (2010). Should we Screen for Breast Cancer? *British Medical Journal*, **341** (7766), 233–235.

SCB (2009). *Tid för Vardagsliv – Kvinnors och Mäns Tidsanvändning 1990/91 och 2000/01*. Report 99, Statistiska Centralbyrån, Levnadsförhållanden.

SCHAPPERT, S. M. and NELSON, C. (1999). National Ambulatory Medical Care Survey: 1995-96 Summary. *Vital and Health Statistics*, **142** (13), 1–122, http://www.cdc.gov/nchs/products/series.htm.

SELIN, H. (2009). *The Rise in Female Employment and the Role of Tax Incentives. An Empirical Analysis of the Swedish Individual Tax Reform 1971*. Working Paper 2009:3, Uppsala Center for Fiscal Studies.

SINDELAR, J. L. (1982). Differential Use of Medical Care by Sex. *Journal of Political Economy*, **90** (5), 1003–1019.

SOU (2004). *Den Könsuppdelade Arbetsmarknaden*. Tech. Rep. 2004:43, Statens offentliga utredningar (SOU).

SSIA (2007). *Försäkringskassan Analyserar: Nej till sjukpenning. Vad Hände Sen?* Tech. Rep. 2007:1, Swedish Social Insurance Agency (SSIA).

STRONEGGER, W.-J., FREIDL, W. and RASKY, V. (1997). Health Behaviour and Risk Behaviour: Socioeconomic Differences in an Austrian Rural County. *Social Science and Medicine*, **44** (3), 423–426.

TICHENOR, V. (1999). Status and Income as Gendered Resources: The Case of Marital Power. *Journal of Marriage and the Family*, **61** (3), 638–650.

UITENBROEK, D. G., KEREKOVSKA, A. and FESTCHIEVA, N. (1996). Health Lifestyle Behaviour and Socio-Demographic Characteristics. A Study of Varna, Glasgow and Edinburgh. *Social Science and Medicine*, **43** (3), 367–377.

UK TRIAL OF EARLY DETECTION OF BREAST CANCER GROUP (1988). First Results on Mortality Reduction in the UK Trial of Early Detection of Breast Cancer. *The Lancet*, **332** (8608), 411–416.

VERBRUGGE, L. M. (1982). Sex Differentials in Health. *Public Health Report*, **97** (5), 417–437.

Appendix A Tables and figures

Table A.1. *List of included variables*

Variable	Description
Female	1 if female.
Child	1 if the individual has at least one child.
Age	The age in years of an individual.
Earnings	Annual earnings in 2001 currency.
High Earner	Annual earnings $>$ 7.5 Basic Amounts.
Income	After-tax income in 2001 currency.
Virtual Income	Spouse's income in 2001 currency.
Primary Earner	1 if earnings more than spouse's earnings.
High Education	1 if individual has post-secondary education.
Days of Sickness Absence	Annual days of sickness absence.
Admission	1 for all years after a hospital admission.
Re-admission	1 if a second hospital admission occurred.
Disability	1 if observed to receive disability benefits.
Death	1 for all years after death.

NOTE.— Incomes are measured in 2001 Swedish Basic Amounts (BA). One BA was equal to approximately € 3,300 in 2001.

Table A.2. *Means and standard deviations of the included variables, by year*

	Year												
Variable	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Total
Female	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Child	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82
Age	38.96	39.96	40.96	41.96	42.96	43.96	44.96	45.96	46.96	47.96	48.96	49.96	44.46
Earnings	5.04	5.28	5.45	5.76	5.99	6.15	6.34	6.55	6.75	6.92	7.02	7.13	6.19
	(2.46)	(3.73)	(2.96)	(3.07)	(3.37)	(3.75)	(4.21)	(4.60)	(4.81)	(5.28)	(5.12)	(5.33)	(4.21)
High Earner	0.10	0.13	0.16	0.2	0.23	0.26	0.29	0.33	0.38	0.41	0.44	0.46	0.28
Income	8.26	8.69	9.06	9.61	10.04	10.40	10.9	11.53	12.21	12.73	13.14	12.33	10.72
	(4.17)	(5.38)	(4.89)	(5.29)	(5.79)	(6.25)	(13.34)	(25.47)	(39.78)	(48.93)	(58.03)	(8.67)	(26.37)
Virtual Income	3.22	3.41	3.61	3.85	4.05	4.24	4.56	4.98	5.46	5.81	6.12	5.19	4.53
	(3.48)	(3.91)	(3.90)	(4.21)	(4.56)	(4.80)	(12.54)	(24.99)	(39.46)	(48.63)	(57.80)	(6.21)	(25.97)
Primary Earner	0.31	0.31	0.32	0.32	0.32	0.32	0.33	0.33	0.33	0.33	0.33	0.32	0.32
High Education	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
Days of S.A.	18.35	22.95	24.29	22.31	21.38	26.29	30.93	36.6	40.07	41.6	40.34	36.06	30.1
	(54.59)	(64.00)	(68.65)	(67.06)	(68.52)	(75.62)	(82.33)	(90.49)	(95.81)	(98.04)	(98.07)	(92.59)	(81.36)
Admission	0.19	0.19	0.18	0.17	0.16	0.15	0.14	0.14	0.13	0.13	0.13	0.13	0.15
Re-admission	0.00	0.04	0.08	0.13	0.17	0.21	0.25	0.29	0.32	0.36	0.39	0.43	0.22
Disability	0.02	0.02	0.03	0.04	0.04	0.05	0.06	0.07	0.09	0.10	0.12	0.14	0.07
Death	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.03	0.02
Individuals	63,599	63,458	63,268	63,055	62,835	62,609	62,363	62,129	61,879	61,669	61,428	61,173	749,494

NOTE.— The table reports means and (standard deviations) of included variables over the analyzed time period. Incomes are measured in 2001 Swedish Basic Amounts (BA). One BA was equal to approximately € 3,300 in 2001. See Table A.1 for detailed variable definitions.

Table A.3. *Means and standard deviations of the included variables, by sample gender and household status*

	Analysis sample				Comparison sample			
	Men		Women		Men		Women	
	Child	No Child	Child	No Child	Child	No Child	Child	No Child
Age	36.556	36.566	36.851	36.629	36.299	36.29	36.513	36.305
	(2.246)	(2.251)	(2.085)	(2.245)	(2.339)	(2.344)	(2.238)	(2.354)
Earnings	7.05	6.567	4.62	5.676	7.435	6.855	4.83	5.931
	(3.993)	(2.526)	(1.96)	(2.142)	(4.147)	(2.645)	(2.127)	(2.06)
High Earner	0.31	0.243	0.062	0.132	0.353	0.278	0.077	0.156
	(0.463)	(0.429)	(0.24)	(0.338)	(0.478)	(0.448)	(0.266)	(0.363)
Income	5.267	4.68	4.338	4.172	5.412	4.845	4.326	4.241
	(3.82)	(1.72)	(3.858)	(2.014)	(3.03)	(1.637)	(1.64)	(1.386)
High education	0.209	0.171	0.316	0.332	0.241	0.195	0.348	0.355
	(0.406)	(0.376)	(0.465)	(0.471)	(0.428)	(0.396)	(0.476)	(0.479)
Primary Earner	0.072	0.011	0.573	0.09	0.067	0.014	0.61	0.083
	(0.259)	(0.105)	(0.495)	(0.286)	(0.25)	(0.116)	(0.488)	(0.276)
Virtual Income	2.674	0.34	4.963	0.905	2.858	0.375	5.318	0.85
	(2.548)	(1.391)	(7.907)	(2.571)	(3.094)	(1.675)	(4.655)	(2.524)
Days of S.A.	4.127	5.526	8.306	9.968	1.99	2.215	4.515	4.794
	(23.439)	(29.237)	(35.285)	(42.247)	(15.665)	(16.069)	(25.494)	(28.034)
Observations	58,015	12,105	49,670	8,113	75,373	14,204	52,586	8,573

NOTE.— The table reports means and (standard deviations) of included variables for the analysis and the comparison sample for individuals < 40 years of age, by gender and household status. The comparison sample is constructed so that each included individual in the analysis sample is matched with an individual of the same age who were not admitted to a hospital during the sampling period. Incomes are measured in 2001 Swedish Basic Amounts (BA). One BA was equal to approximately € 3,300 in 2001. See Table A.1 for detailed variable definitions.

Table A.4. *Number of individuals in analysis sample, by gender and disease category*

Disease category	Women	Men	Total	Share women
Accident	2,968	5,341	8,309	0.36
Blood	238	88	326	0.73
Cancer	5,085	985	6,070	0.84
Congential	150	127	277	0.54
Digestive	3,697	4,366	8,063	0.46
Ear	443	449	892	0.50
Endocrine	957	593	1,550	0.62
Eye	239	305	544	0.44
Hosp. factors	1,957	1,136	3,093	0.63
Genitourinary	5,084	1,231	6,315	0.81
Heart	1,710	2,899	4,609	0.37
Infection	651	941	1,592	0.41
Mental	1,402	1,783	3,185	0.44
Musculoskeletal	2,184	2,922	5,106	0.43
Nerve system	667	717	1,384	0.48
Perinatal	3	1	4	0.75
Respiratory	1,413	1,900	3,313	0.43
Skin	333	354	687	0.48
Symptoms	3,414	4,826	8,240	0.41
Total	32,595	30,964	63,559	0.51

NOTE.— The diagnosis categories are grouped according to the chapter division of The International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10). Each sampled individual is categorized with respect to the primary diagnosis code registered at the time of the first hospital admission.

Table A.5. *Number of individuals in analysis sample, by gender and industry sector*

Industry	Women	Men	Total	Share women
Agriculture	164	597	761	0.22
Construction	295	3,038	3,333	0.09
Education	3,225	1,486	4,711	0.68
Energy	125	486	611	0.20
Finance	782	594	1,37	0.57
Health	14,779	2,301	17,080	0.87
Services	406	316	722	0.56
Manufacturing	3,407	8,860	12,267	0.28
Mining	20	146	166	0.12
Other	123	288	411	0.30
Other Pers. Service	1,177	1,221	2,398	0.49
Public Administration	2,098	1,948	4,046	0.52
Real Estate and Renting	2,036	2,874	4,910	0.41
Retail and Wholesale	2,476	3,513	5,989	0.41
Transportation	1,482	3,296	4,778	0.31
Total	32,595	30,964	63,559	0.51

NOTE.— The Industry Sector code (SNI) is aggregated into 15 categories covering the full labor market and closely based on the EU standard classification, NACE Revision II. Each sampled individual is categorized with respect to the sector he or she was employed in at the first sampling year.

Appendix B Empirical modeling

This section derives the probability limits of the OLS estimator under different assumptions with respect to the health shock discussed in the paper. In particular, we show that under plausible assumptions the estimator of the gender difference in sickness absence will be conservative.

Gender differences

Assume labor supply and household production are fixed between two time periods, $t = (0,1)$, but that individual health is poorer in the second time period, i.e. $\eta_{i1j} - \eta_{i0j} = \Delta_{ij} < 0$; the change in sickness absence, equation (3.4), over the two time periods is given by

$$s_{i1j} - s_{i0j} = \gamma_j \Delta_{ij} + \delta_f^{\#} c_i F_i \times \Delta_{ij} + \varepsilon_{i1j} - \varepsilon_{i0j},\ i = 1,...,n_j, \tag{B.1}$$

where $n_j = n_{j1} + n_{j0}$ is the the number of men $(j = m)$ and women $(j = f)$ with $(c = 1)$ and without $(c = 0)$ children and with model parameters defined in Section 3. For women we have

$$s_{i1f} - s_{i0f} = \gamma_f \Delta_{if} + \delta_f^{\#} c_i \times \Delta_{if} + \varepsilon_{i0f} - \varepsilon_{i1f},\ i = 1,...,n_f,$$

and for men

$$s_{i1m} - s_{i0m} = \gamma_m \Delta_{im} + \varepsilon_{i1m} - \varepsilon_{i0m},\ i = 1,...,n_m.$$

From the i.i.d. of ε_{itj} we get

$$p\lim \frac{1}{n_f} \sum_{i=1}^{n_f} \left(s_{i1f} - s_{i0f}\right) = \gamma_f^{\#} \Delta_f \equiv \gamma_f^* > 0,$$

where $\gamma_f^{\#} = \gamma_f + \delta_f^{\#} p_{f1},\ p_{f1} = n_{f1}/n_f$ is the fraction of women who have children and $p\lim \frac{1}{n_f} \sum_{i=1}^{n_f} \left(\eta_{i1f} - \eta_{i0f}\right) = (\overline{\eta}_{f1} - \overline{\eta}_{f0}) = \Delta_f < 0$ is the average health shock requiring an admission for women. For men we get

$$p\lim \frac{1}{n_m} \sum_{i=1}^{n_m} \left(s_{i1m} - s_{i0m}\right) = \gamma_m \Delta_m = \gamma_m^* > 0$$

where $p\lim \frac{1}{n_m} \sum_{i=1}^{n_m} \left(\eta_{i1f} - \eta_{i0f}\right) = (\overline{\eta}_{m1} - \overline{\eta}_{m0}) = \Delta_m < 0$ is the average health shock requiring an admission for men.

We can now specify

$$s_{it} = b_0 + b_1 F_i + \gamma_m^* H_{it} + \kappa^{\circ} F_i H_{it} + \varepsilon_{it}, \tag{B.2}$$

where

$$\kappa^{\circ} = \kappa \Delta_f + \gamma_m (\Delta_f - \Delta_m), \tag{B.3}$$

$$b_0 = \phi_m \overline{C}_{sm} + \lambda_{m1}\overline{h}_m + \lambda_{m2} p_{m1}\overline{h}_f + \omega_{m2} p_{m1}\overline{h}_f^p + \omega_{m1} p_{m0}\overline{h}_m^p + \zeta_m p_m \overline{\mu}_m$$

and

$$b_1 = \phi_f \overline{C}_{sf} + \lambda_{f1}\overline{h}_f + \lambda_{f2} p_{f1}\overline{h}_m + \omega_{f1} p_f \overline{h}_f^p + \zeta_f p_f \overline{\mu}_f - b_0$$

with parameters defined in Section 3, $p_j = n_j/n$, $p_{jc} = n_{jc}/n_j$ and $\overline{z}_j$ and $\overline{z}_{jc}$, $j = (m,f)$, $c = (0,1)$ are sample means and z being generic for (C_s, μ, h, h^p).

Assume that men and women have the same pre-admission health (i.e. $\overline{\eta}_{m0} = \overline{\eta}_{f0}$) and that the health requiring a hospital admission is on average the same for men and women.[1] The data generating process for the admission would then imply that $\tau_F = 0$ in

$$H_i = \mathbf{1}(\eta_i < \tau + \tau_F F_i),$$

where $\mathbf{1}(\cdot)$ is the indicator function and τ is the average health threshold of being admitted at a hospital. This implies that $\Delta_m = \Delta_f = \Delta$. Under these two assumptions the ordinary least squares (OLS) estimator in equation (B.2) would be a consistent estimator of the average female to male difference in sickness absence from a negative health shock Δ, that is

$$\kappa^* = \kappa\Delta.$$

Now, consider the situation where women invest more in their health than men. The health change requiring a hospital admission should thus be less severe for women on average, that is $\tau_F > 0$, or

$$\Delta_f - \Delta_m > 0.$$

Under this assumption the OLS estimator would be a biased estimator of κ^*. From the definition of κ° in equation (B.3) it follows that

$$p\lim \widehat{\kappa}^* - \kappa^* = \kappa\left(\Delta_f - \Delta\right) + \gamma_m\left(\Delta_f - \Delta_m\right),$$

where $\widehat{\kappa}^*$ is the OLS estimator. Under the alternative hypothesis, $\kappa < 0$, and $\gamma_m < 0$

$$p\lim \widehat{\kappa}^* - \kappa^* < 0.$$

The estimated relative effect of the health change on sickness absence will thus be conservative if women visit a hospital for less severe health problems than men.

In order to take any pre-admission differences in health between genders into account let

$$\eta_{i1j} = \eta_{i0j} + \omega_{i1j},$$

[1] The hospital admission could be modeled as the combination of two different thresholds; one which governs the individual's decision to visit a hospital and one which governs the physician's decision to admit the visiting individual. We disregard from the latter threshold here, assuming that physicians are not able to perfectly screen the health status of individual patients.

In this specification ω_{i1j} is the health shock from the original level η_{i0j}. Assume

$$\eta_{i0j} = X_i\beta + u_{ij},$$

where X_i is a vector of observed covariates, $\beta \neq 0$ and u_{ij} is i.i.d. This allows us to write the "structural" model (B.1) as

$$s_{i1j} - s_{i0j} = \gamma_j\omega_{i1j} + \delta_f^{\#}c_iF_i \times \omega_{i1j} + \varepsilon_{i1j} - \varepsilon_{i0j},\ i = 1,...,n_j.$$

The implication is that the "difference-in-differences" specification (B.2) is given as

$$s_{it} = b_0 + b_1F_i + \gamma_m^{\star}H_{it} + \kappa^{\star}F_iH_{it} + X_i\beta + \varepsilon_{it}\ i = 1,...,n, \tag{B.4}$$

where

$$\kappa^{\star} = \kappa\overline{\omega}_{f1} + \gamma_m(\overline{\omega}_{f1} - \overline{\omega}_{m1}) \tag{B.5}$$

and $\overline{\omega}_{j1}$ is the average unobserved health shock between the two time periods for the males ($j = m$) and females ($j = f$), respectively. Under the assumption of equal average unobserved health shocks of men and women, the OLS estimator of equation (B.4) would converge to

$$\kappa_x^* = \kappa\overline{\omega}_1,$$

where hence $\overline{\omega}_1 = \overline{\omega}_{f1} = \overline{\omega}_{m1}$. If the unobserved shock requiring a hospital admission is on average less severe for women i.e. $(\overline{\omega}_{f1} - \overline{\omega}_{m1}) > 0$, the OLS estimator $\widehat{\kappa}_x^*$ is biased. It is evident from equation (B.5) that

$$p\lim\widehat{\kappa}_x^* - \kappa_x^* = \kappa(\overline{\omega}_{f1} - \overline{\omega}_1) + \gamma_m(\overline{\omega}_{f1} - \overline{\omega}_{m1}).$$

This will under the alternative hypothesis, $\kappa < 0$, again be biased downwards. If $\widehat{\kappa}_x^* > \widehat{\kappa}^*$, this suggests that women enjoy better pre-admission health than men.

Household investments

Let $w_{itjc} = c_iw_{itj} + (1 - c_i)w_{itj}$ for $w \in (s, \eta, \varepsilon)$. The difference in sickness absence over time for women with children is given by

$$\begin{aligned}\left(s_{i1f1} - s_{i0f1}\right) &= \gamma_f(\eta_{i1f1} - \eta_{i0f1}) + \delta_f^{\#}c_iF_i \times (\eta_{i1f1} - \eta_{i0f1}) \\ &+\varepsilon_{i1f1} - \varepsilon_{i0f1}, i = 1,...,n_{f1},\end{aligned} \tag{B.6}$$

Then, given i.i.d of ε_{itf1},

$$p\lim\frac{1}{n_{f1}}\sum_{i=1}^{n_{f1}}\left(s_{i1f1} - s_{i0f1}\right) = \left(\gamma_f + \delta_f^{\#}\right)\left(\overline{\eta}_{1f1} - \overline{\eta}_{0f1}\right) = \gamma_{f1}^*,$$

where $p\lim \frac{1}{n_{f1}} \sum_{i=1}^{n_{f1}} (\eta_{i1f1} - \eta_{i0f1}) = (\overline{\eta}_{1f1} - \overline{\eta}_{0f1}) = \Delta_{f1}$. Similarly, for women without children the difference is given by

$$\left(s_{i1f0} - s_{i0f0}\right) = \gamma_f(\eta_{i1f0} - \eta_{i0f0}) + \varepsilon_{i1f0} - \varepsilon_{i0f0}, i = 1, ..., n_{f0},$$

where n_{f1} is the number of females without children. Now

$$p\lim \frac{1}{n_{f0}} \sum_{i=1}^{n_{f0}} \left(s_{i1f0} - s_{i0f0}\right) = \gamma_f(\overline{\eta}_{1f0} - \overline{\eta}_{0f0}) = \gamma^*_{f0},$$

where $p\lim \frac{1}{n_{f0}} \sum_{i=1}^{n_{f0}} (\eta_{i1f0} - \eta_{i0f0}) = (\overline{\eta}_{1f0} - \overline{\eta}_{0f0}) = \Delta_{f0}$.

We can now use the same type of "difference-in-differences" specification as for the gender comparison in the analysis in the OLS estimation; that is

$$s_{it} = b^*_0 + b^*_1 c_i + \gamma^*_{f0} H_{it} + \delta^\circ c_i H_{it} + \varepsilon_{it},\ i = 1, ..., n_f, \tag{B.7}$$

where

$$\begin{aligned} \delta^\circ &= (\delta^\#_f + \gamma_f)\Delta_{f1} - \gamma_f \Delta_{f0}, \\ b^*_0 &= \phi_f \overline{C}_{sf} + \lambda_{f1} \overline{h}_f + \omega_{f1} \overline{h}^p_f + \zeta_m \overline{\mu}_m \end{aligned} \tag{B.8}$$

and

$$\begin{aligned} b^*_1 = {} & \phi_f(\overline{C}_{sf1} - \overline{C}_{sf}) + \lambda_{f1}(\overline{h}_{f1} - \overline{h}_f) \\ & + \omega_{f1}(\overline{h}^p_{f1} - \overline{h}^p_f) + \zeta_m(\overline{\mu}_{m1} - \overline{\mu}_m) + \lambda_{m1}\overline{h}_m \end{aligned}$$

where $\overline{C}_{sfc}, \overline{h}^p_{fc}$, and $\overline{h}_{fc}$ are mean values for women with children ($c = 1$) and without children ($c = 0$).

If the health shock that leads to a hospitalization is the same for both groups, i.e. $\Delta_f = \Delta_{f0} = \Delta_{f1}$,

$$\delta^\circ = \delta^* = \delta^\#_{f1} \Delta_f.$$

If the the pre-admission health of women with and without children is the same, but the health shock for women with children rendering a hospital admission is lower than for women without children, that is

$$\Delta_{f1} - \Delta_{f0} > 0,$$

the OLS estimator of δ^* will be biased downwards. From equation (B.8) we get

$$p\lim \widehat{\delta}^* - \delta^* = (\delta^\#_f + \gamma_f)\left(\Delta_{f1} - \Delta_f\right) - \gamma_f \Delta_{f0}$$

and under the assumption, $\Delta_{f1} - \Delta_{f0} > 0$, $p\lim \widehat{\delta}^* - \delta^* < 0$.

Assume $\eta_{i1jc} = \eta_{ijc0} + \omega_{i1jc}$ and that initial health follows $\eta_{i0jc} = X_i\beta + u_i$,. This allows us to write the structural model (B.6) as

$$\left(s_{i1fc} - s_{i0fc}\right) = \gamma_f \omega_{i1jc} + (\delta^\#_f + \gamma_f) c_i F_i \times \omega_{i1jc} + \varepsilon_{i1fc} - \varepsilon_{i0fc}, i = 1, ..., n_{fc},$$

The implication is that the "difference-in-differences" specification (B.7) is given as

$$s_{it} = b_0^* + b_1^* c_i + \gamma_{f0}^* H_{it} + \delta_x^* c_i H_{it} + X_i \beta + \varepsilon_{it}.$$

Now

$$\delta_x^* = (\delta_f^{\#} + \gamma_{f0})\overline{\omega}_{1f1} - \gamma_{f0}\overline{\omega}_{1f0}, \tag{B.9}$$

where $\overline{\omega}_{1fc} < 0$ is the average unobserved health shock in period one for women with ($c = 1$) and without ($c = 0$) children. Under the assumption of equal average unobserved conditional health shocks we have

$$\delta_x^* = \delta_f^{\#}\overline{\omega}_{f1},$$

since $\overline{\omega}_{f1} = \overline{\omega}_{1f1} = \overline{\omega}_{1f0}$. If the unobserved shock requiring a hospital admission is on average less severe for women with children, i.e. $(\overline{\omega}_{1f1} - \overline{\omega}_{1f0}) > 0$, the OLS estimator $\widehat{\delta}_x^*$ is biased. It is evident from equation (B.9) that

$$p\lim \widehat{\delta}_x^* - \delta_x^* = (\delta_f^{\#} + \gamma_{f0})(\overline{\omega}_{1f1} - \overline{\omega}_{1f}) - \gamma_{f0}\overline{\omega}_{1f0}.$$

which, under the alternative hypothesis that $\delta_{f1}^{\#} < 0$, will again be biased downwards. If $\widehat{\delta}_x^* > \widehat{\delta}^*$, this supports the idea that women with children have better pre-admission health than those without.

Assume the household differences in the magnitudes of the health shocks between men and women are proportional, so that

$$\Delta_{f1} - \Delta_{f0} = \pi\left[\Delta_{m1} - \Delta_{m0}\right],$$

where $\Delta_{m1} = (\overline{\eta}_{m11} - \overline{\eta}_{m10})$, $\Delta_{m0} = (\overline{\eta}_{m01} - \overline{\eta}_{m00})$, and where $\pi = \gamma_m/\gamma_f < 1$ under the alternative hypothesis, $|\gamma_f| > |\gamma_m|$. Under this assumption men with and without children can be used to control for the potential difference in the health threshold for visiting a hospital for women with and without children. The implication of this assumption is that the difference in health threshold should be greater for men with and without children than the corresponding difference for women with and without children. Based on the information contained in Figure 5.4 we believe this assumption to be plausible. The OLS estimator of δ^* using a triple difference model is subsequently computed from

$$\begin{aligned} s_{it} &= b_0 + b_1 F_i + b_2 c_i + b_3 F_i c_i + \gamma_{m0}^* H_{it} + \gamma_{m1}^* H_{it} c_i \\ &+ \gamma_{f0}^* F_i H_{it} + \delta^* F_i H_{it} c_i + u_{it} \end{aligned}$$

where γ_{m1}^* and γ_{m0}^* is the response of men with and without children.

Essay 2: A matter of life and death? Hospital distance and quality of care: Evidence from emergency room closures and myocardial infarctions*

1 Introduction

Over the twenty-year period between 1987 to 2007 more than half a million Swedish residents — a country with approximately nine million inhabitants — were at some point registered as having suffered an acute myocardial infarction (AMI). At the end of that period about half of these individuals were deceased, most of them with AMI as either the primary or as a contributing cause of death. Overall AMI incidence in Sweden over the same period exceeded 800,000 cases, making AMI one of the leading causes of hospitalization as well as the leading cause of death in Sweden at the time (Socialstyrelsen, 2009). In other words, about 12 percent of the Swedish population is expected to suffer an AMI at some point in their life (Nationellt register för hjärtstopp, 2011). Far from unique in this respect, Sweden shares these morbidity and mortality patterns with most other countries in the Western world. For example, 500,000 annual deaths in the U.S. are the result of an AMI (American Heart Association, 2012).

The relatively high death rates for AMI arise primarily from two specific disease characteristics; the lack of indication signals — or the unexpectedness — of the disease (more than two-thirds of Swedish AMIs occur in the home) and the critical time aspect for AMI treatment. In the event of a cardiac arrest, a common manifestation of the infarction, the brain may suffer irreversible damage after only five minutes due to the lack of oxygen. After fifteen minutes death is almost unavoidable regardless of any resuscitation attempts made (Pell *et al.*, 2001; GUSTO Investigators, 1993). These two characteristics together imply that professional medical assistance may often be unavailable and out of reach when the life-threatening condition occurs. Hence, many AMI patients die before they reach a medical care facility. For example, in the U.S. about 60 percent of all AMI deaths occur outside a hospital (American Heart Association, 2012).

*The author thanks Per Johansson, Arizo Karimi, Martin Karlsson, Tobias Laun, Johan Vikström and seminar participants at Uppsala University, CINCH-Essen, IFAU-Uppsala and EALE 2013. Financial support from the Swedish Council for Working Life and Social Research (DNR 2004-2005 and 2009-0826) is also gratefully acknowledged.

Health care centralization trends have emerged recently in many countries. In countries with more deregulated health care markets (e.g. the U.S.), centralization has been driven primarily by increased competition in the health care sector in which hospitals have either merged into large multi-hospital units or been ousted by competition from more efficiently driven hospitals (Dranove *et al.*, 1996; Succi *et al.*, 1997; Evans-Cuellar and Gertler, 2003). In countries with an obligatory and, mainly, public provision of health care (e.g. Sweden) increasing health care costs and public budget deficits have been, along with general technological progress and innovations in health care, a driving factor behind the structural changes. Examples of such changes are increased reliance on outpatient care, and on paramedic and emergency ambulance services (Landstingsförbundet, 2002; Sveriges Kommuner och Landsting, 2008). Hence, irrespective of the institutional context, the long-run trend in the organization of inpatient health care has been a considerable increase in the centralization of resources.

One significant feature of these centralization patterns has been the tendency of an increase in the number of rural hospital closures and a corresponding growth in the size of hospitals in urban areas. While potentially reducing overall costs, these centralizations may also entail a number of adverse effects on the quality of health care — in particular with respect to the geographical access to health care. In this context it is interesting to note that Swedish health care authorities have justified the centralization measures taken with the argument that emergency hospitals, while traditionally important for health care equity policies, are less important today due to recent innovations in emergency medical treatment (Sveriges Kommuner och Landsting, 2004).[1] For this reason it is, besides the importance of assessing the magnitude of health care accessibility effects more generally, also a relevant research question for health care policy to empirically probe the validity of this argument.

In this paper I empirically assess the impact of geographical access to health care on AMI survival for individuals who suffered an AMI in Sweden between 1990 and 2010. Previous contributions on this topic have typically found that ambulance response time (i.e. the time it takes from the emergency call until the ambulance has arrived) increases the chance of surviving an out-of-hospital AMI (Bachmann *et al.*, 1986; Piette and Moos, 1996; Norris, 1998; Pell *et al.*, 2001). However, this conclusion mainly stems from evidence based on case studies, i.e. studies using data on single hospitals and/or data culled at one particular point in time. Accordingly, there is an obvious risk that results inferred from these studies suffer appreciably from limitations associated with both the external and internal validity of any estimated parameters. In most cases, location data on both individuals and hospitals is likely to be subject

[1] For example, over the last few decades some therapeutic progress has been made, including the introduction of specific MI wards, mobile defibrillators, more effective treatment of cardiac arrest and the introduction of drugs such as beta blockers, thrombolytic agents, aspirin, ACE inhibitors and lipid-lowering drugs (Julian, 1961; Dellborg *et al.*, 1994; Herlitz, 2000).

to spatial sorting across time in that agents' choice of residence is based on factors related to AMI survival probabilities, such as individual health and the quality of nearby hospitals. First, hospitals in more market-oriented health care systems are likely to be strategically located with regard to underlying patient characteristics and competition aspects. For example, profit-maximizing hospitals are unlikely to be located in impoverished areas where the patient population has poor general health (Dranove *et al.*, 1996; Succi *et al.*, 1997). Furthermore, individuals are likely to take access to health care into account when they decide where to live. In particular, individuals in poor health would, *ceteris paribus*, choose to reside closer to a hospital, compared to individuals in good health. Clearly, unless the analyst is able to sufficiently control for unobserved individual health, these sorting mechanisms will confound any estimated effects of distance.

In the empirical analysis of this paper, I utilize detailed nationwide register data spanning over a twenty-year period. This makes it possible to account for both cross-sectional and time variation in AMI survival rates and to control for observed individual heterogeneity. In addition, I make use of an "exogenous" change in distance to estimate average impact of the distance to a hospital on surviving an AMI.

A second problem hampering the assessment of the impact of distance to hospital on health in many studies is the lack of out-of-hospital data. Using only inpatient data when quantifying the effect of distance means that patients who die before reaching hospital will be left out of the analysis. Clearly, if geographical access to health care has an impact on survival probability, omitting patients that die en route to hospital will underestimate the true effect of distance, since patients admitted to hospital who live farther away from a hospital will on average be in better health relative to patients admitted to hospital who live close by (Gillum, 1990; O'Neill, 2003). I avoid this sample selection problem by supplementing the Swedish national inpatient registry with the Swedish national causes of deaths registry, consisting of detailed information on all deaths that occurred in Sweden for all years of study.

In order to generate plausibly "exogenous" changes in distance, I make use of a number of Swedish emergency hospital closures over the studied time period. In the beginning of the 1990s, Sweden had a large geographical spread of emergency hospitals across the country. However, the economic crisis of the 1990s resulted in large public deficits and, as a reaction to this; aggregate health care spending was cut by more than ten percent. A large portion of these cost savings were derived from centralization measures — in particular the closure of a number of emergency hospitals. These closures, plausibly unrelated to individual AMI survival probabilities due to the public nature of health care provision, entailed an implicit change in the distance to an emergency hospital for patients residing in the catchment areas of a closed emergency hospital. Utilizing variation in individual distances to hospitals generated from the policy-induced closures, I am able to circumvent endogeneity

problems arising from self-selection by estimating AMI survival probability as a function of the current geographical distance to an emergency hospital — *conditional* on pre-closure distance.

I find that an increase in distance significantly predicts a lower AMI survival probability for patients residing in the catchment area of a closed emergency hospital. Specifically, the estimates suggests that increasing the geographical distance to hospital from within a ten-kilometer radius to between a 51 to 60-kilometer radius would result in a decrease in expected AMI survival probability of 11.5 percentage points, corresponding to a 15 percent lower survival chance at sample mean survival rates. Furthermore, this effect is primarily driven by an increased risk of out-of-hospital mortality among affected patients. Much smaller effects are found when estimating the impact of distance based on actual distances to hospital, indicating that selective residential sorting is likely to greatly dilute the effect of distance. Finally, I find that the effect is concentrated to the first year after the closures, indicating that no long-run elevated AMI mortality from the closures seems to have occurred.

The results from this study may to some extent be contrasted to to the volume-outcome literature in which the centralization of health care may increase the quality of health care, due to e.g. scale effects and learning-by-doing (Luft *et al.*, 1987; Maerki *et al.*, 1986; Hamilton and Ho, 1998). According to this literature, centralization increases health care quality and is hence considered desirable. However, the disease context is clearly crucial as to which of these effects is likely to dominate. In particular, while Thiemann *et al.* (1999) finds a positive association between hospital volume and survival of AMI patients, it is likely that any positive quality effects from centralization in this context should be more counteracted by the negative effects on survival, arising from a decrease in geographical access to health care, than for planned surgery where the situation is less acute (e.g. organ transplants and cancer surgery).

The remainder of the article begins with a brief summary of the Swedish health care system in section two. Section three includes a presentation of the data and the sampling methodology. Section four offers a careful review of the empirical approach, in particular with respect to the various identification problems encountered. Section five presents the results from estimations and section six contains a short summary along with some concluding remarks.

2 The Swedish health care system

In contrast to e.g. the U.S., health care in Sweden is highly regulated. The vast majority of Swedish hospitals are owned and run by the public sector.[2] The Swedish health care system is organized and financed by 21 independent

[2]In fact, the first and hitherto only private hospital was established as late as in 1999.

regions, Stockholm being the largest (with about 2 million inhabitants) and Gotland the smallest (with about 60,000 inhabitants). Health care is the single most important responsibility for the regional administration; for instance, in 2012 on average 82 percent of the county budgets were on health care spending.

The regional administrations are governed by political councils elected in national elections every four years. Besides following a few general guidelines set by the national government — e.g., that health care should be provided to all Swedish citizens — the regional authorities have high levels of discretion in organizing health care. This particular institutional setting implies that political representatives of the county councils and bureaucrats, rather than competition among providers, largely determine the number, size, location and coverage of hospitals in each region.

Another consequence of the highly regulated health care sector in Sweden is that patients have little choice as to which hospital they are admitted to in an emergency situation. As the vast majority of health care in Sweden is funded by taxes, there are no individual contracts between patients and hospitals.[3] Instead, depending on the geographical position of a patient's home, he or she will be directed to a specified hospital nearby when health care is needed. This institutional setting ensures that each patient has a designated "home hospital" each year, which can be identified by using aggregated historical admission data for each municipality and linking this information to the patients' registered municipality of residence.

The time period studied in this article, i.e. the last two decades 1990-2010, was a period of strong centralization in the Swedish health care sector. These centralization measures were deemed necessary by regional authorities in order to increase efficiency and to cover public deficits caused by the economic turbulence in Sweden in the beginning of the 1990s.[4] In total, government spending on health care decreased by 11 percent, from 8.8 to 7.7 percent of GDP, between 1990 and 2000. A significant share of these savings were de-

[3]When seeking health care in Sweden a small fee is normally paid up front by the patient. In Stockholm county this fee currently (2013) ranges from 100 SEK (≈10 EUR) when e.g. visiting a physiotherapist to 400 SEK (≈40 EUR) when visiting an emergency room. However, when a patient has paid a total of 1,100 SEK (≈110 EUR)in health care fees in one year, he or she receives a "free card" and health care is free for the remainder of the year. A similar payment system exists for pharmaceuticals in which the patient's share of the drug cost decreases with the total amount spent. In 2013, the maximum amount paid by the patient was 2,200 SEK (≈220 EUR). See e.g. http://www.vardguiden.se/Sa-funkar-det for more information.

[4]The Swedish 1990s economic crisis took place between 1990-1994 and was a combined banking, financial and housing market crisis which is said to have been primarily caused by an unfortunate deregulation of the Swedish credit markets in 1985 (Wohlin, 1998). The financial deregulation led to currency and housing speculation bubbles which deflated in 1991 and resulted in a severe credit crunch and widespread bank insolvency. The cause and development of the Swedish 1990s crisis had much in common with the U.S. subprime mortgage crisis of 2007-2008.

rived from structural changes in health care organization within counties — in particular the closure of a number of emergency hospitals across the country (Landstingsförbundet, 2002).

Importantly, due to the institutional features of the Swedish health care sector, the hospital closures should be unrelated to the health characteristics of the underlying population in the hospitals' catchment areas. Moreover, as each individual patient's designated hospital is known at each point in time, these policy-induced closures can be used in order to compute the shift in geographical distance to the new home hospital among patients whose emergency hospitals were closed.

3 Data and sampling

The data used in this article is primarily based on administrative registers from the Swedish National Board of Health and Welfare, covering information on all Swedish citizens for all the years of study, i.e. 1990-2010. These registers include the Swedish National Patient Register (NPR), consisting of detailed information on all recorded hospitalizations in Sweden, and the National Causes of Death Register (NCDR), consisting of all recorded deaths that occurred in Sweden for individuals with a permanent residence in the country.[5,6] Specifically, the NPR includes individual-level data, for each hospital, on date of admission and discharge, whether the patient were admitted from home or from another clinic, a set of patient characteristics, medical data on diagnoses classified according to the ICD standard[7] and any surgical procedure(s) undertaken during the hospital visit. In addition, the NCDR includes the date, place and primary and contributing causes for each death in the data.

The population of interest consists of all Swedish residents who had an AMI between 1990 and 2010. I sample all the records of hospitalizations and deaths caused by ischemic heart diseases with a primary ICD-10 diagnosis code of I.21 or I.22, corresponding to an *acute* myocardial infarction or re-infarction. I also collect additional information from each hospitalization, such as patient age, gender, residence, specific hospital and clinic as well as any hospitaliza-

[5]The population consists of all deaths that were reported to the Swedish Tax Agency, including all individuals registered as Swedish residents at the time of death. Hence, registered citizens who died outside Sweden (e.g. vacationers) are included while unregistered citizens who died in Sweden are not.

[6]The number of deaths recorded in the NCDR is in practice equivalent to all deaths that occurred in the relevant population. The number of unrecorded deaths in the NCDR in e.g. 2007 amounted to 0.84% (773) of all deaths.

[7]The diagnoses are made by physicians and classified according to the World Health Organization's International Statistical Classification of Diseases and Related Health Problems (ICD-10). ICD-10 is a seven digit coding of diseases and signs, symptoms, abnormal findings, complaints, and external causes of injury or diseases. See e.g. http://www.who.int/classifications/icd/en.

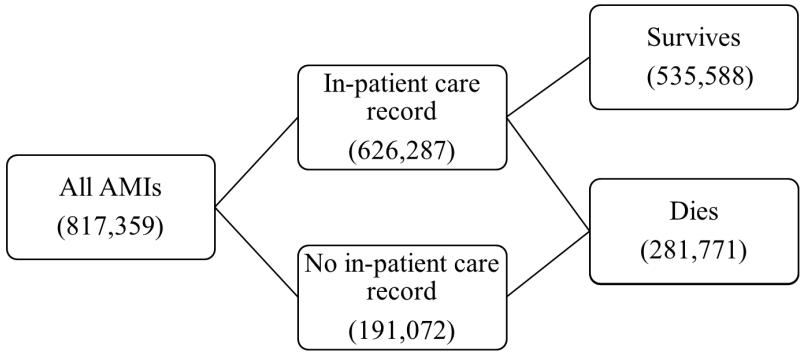
tion and AMI histories for each patient since 1987. The date of death is added to this data from the NCDR (if the individual died at some point).[8] As the data also contains individual identifiers, I am also able to merge the data with other population registers from Statistics Sweden to add further patient characteristics. Crucially, one such characteristic is detailed geographical coordinates for each individual's registered place of residence measured according to the RT-90 standard.[9] These coordinates are subsequently applied to compute the geographical distance from the registered place of residence of each AMI patient included to his or her designated home hospital for each year.

Figure 3.1 illustrates the total number of AMIs in the data broken down into relevant categories. As can be seen, out of approximately 817,000 AMIs, about 75 percent (626,000) show up in the NPR as inpatient care records while the remaining quarter (191,000) consists of individuals who died before arriving at a hospital, and hence only show up in the NCDR. In total, about 65 percent (535,000) of the AMI population survive the AMI while about 35 percent (281,000) die, either before (68 percent) or after (32 percent) being admitted. Clearly, ignoring out-of-hospital mortality will greatly underestimate total AMI mortality in Sweden during this period.

[8] As the main outcome of the empirical analysis is the probability of surviving an AMI, the following population breakdown is important; *i)* patients who survived until they were admitted to a hospital, survived the AMI and were discharged, *ii)* patients who survived until they were admitted to a hospital but died while in hospital and *iii)* individuals who died before reaching a hospital and hence were not admitted. I assume that all AMI patients need inpatient care and hence that there are no patients who survived the AMI but were not admitted. Clearly, as the goal of the empirical analysis is to investigate the effects of the distance to hospital on AMI mortality, excluding out-of-hospital AMI deaths will entail an endogenous sample selection under the alternative hypothesis of the existence of an effect of distance. Therefore, the inclusion of all three categories, using data from both the NPR and the NCDR, is essential in order to establish inference to the population of interest.

[9] Coordinates in "Rikets koordinatsystem" (RT-90) are computed using the Gauss conformal projection or the *Transverse Mercator map projection*. In contrast to the Standard Mercator projection, the transverse projection takes into account that the world is shaped as an ellipsoid and uses complicated calculations and so-called geodetic datums in order to deliver improved accuracy positioning measurements. According to the Swedish Ordnance Survey, the RT-90 measurements cover approximately 3800 triangular points over the country with a relative distance accuracy of 1-2 ppm (mm/km).

Figure 3.1. Acute Myocardial Infarctions in Sweden,1987-2005



NOTE.—Data source: Swedish National Board for Health and Welfare. An AMI is defined as an ICD-10 coding of either I.21 or I.22. The total number of AMIs are obtained by combining the inpatient registry (NPR) and the national causes for death registry (NCDR).

3.1 Home hospitals, emergency room closures and referral hospitals

In order to compute patient hospital distances, I define a "home hospital" for each individual and calendar year based on his or her place of residence. I select this home hospital using historical data on AMI hospitalizations and municipality of residence from the NPR and, for each municipality and year, I select the hospital to which most of the inhabitants in the municipality are admitted (i.e. the modal hospital). For most municipalities this procedure is not a problem. However, a few municipalities do not have a clearly defined home hospital for all the years concerned and for this reason AMI patients residing in these municipalities are taken out of the analysis.[10] Rather than using the actual hospital a patient visits to compute hospital distance, I use the distance to the designated home hospital. In most cases, but not always, these are the same. This classification is used for several reasons: First, I need to assign a counter-factual hospital for AMI patients who died before reaching a hospital. Second, patients observed to be treated at other hospitals than their designated home hospital are likely to be unrepresentative with regard to the distance they actually traveled — e.g. because they were in another region when the AMI occurred. Third, the Swedish institutional setting makes the home hospital definition very reliable — more than 80 percent of all admissions in the sample occurs at the home hospital.

[10]The dropped municipalities are: Salem, Håbo, Boxholm, Ödeshög, Vaggeryd, Hultsfred, Mönsterås, Aneby, Osby, Kungsbacka, Tanum, Färgelanda, Herrljunga, Örkelljunga, Svedala, Falkenberg, Lerum, Grästorp, Vansbro, Leksand and Jokkmokk. These municipalities constitute about seven percent of the total number of Swedish municipalities and much less of the total AMI population. Moreover, none of these municipalities are located in regions where an emergency hospital closure occurred.

The home hospital definition is used to compute the changes in distance due to hospital closures in two steps: First, in order to identify individuals who were affected by an emergency hospital closure and, subsequently, to compute the new distance to hospital for these individuals by defining a new home hospital (the referral hospital) and the new geographical distance to this hospital.[11] The distance to the new home hospital after closure is subsequently used in our empirical application to estimate the parameters of interest — contingent on the pre-closure distance to the old home hospital.

Emergency hospital closures are defined by the change in the number of AMI admissions they receive across two consecutive years.[12] I find a total of sixteen closures between 1990-2010.[13] The closures identified in the data are also validated from other sources such as official documents, local media coverage and prior research on emergency hospital closures in Sweden.[14] Obtaining information about the closures was not difficult as the hospital closures typically generated great stirs in public opinion.

Figure 3.2 (and Figures A.1—A.6 in the Appendix) present the monthly number of visits for each hospital that was closed and the corresponding referral hospital over the period of study. The panel on the left of each individual closure plot displays the unadjusted raw number of admissions while the panel on the right displays a six-month moving average of admission frequency. The figures show that the referral hospitals almost absorb the full reduction of admissions of the hospitals that were closed.[15]

[11] I use the same strategy to define the referral home hospitals as the other home hospitals, i.e. using historical admissions in the NPR, I infer which hospital patients living in closure-affected municipalities are referred to after a closure.

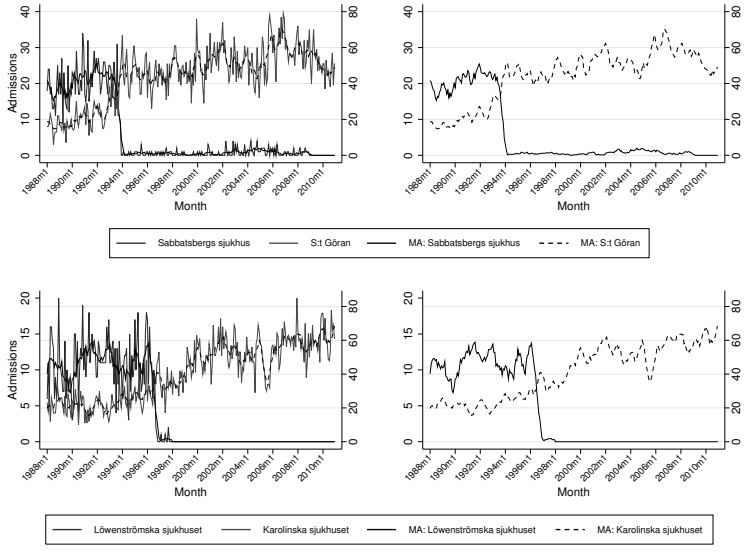
[12] Specifically, a hospital is classified as closed if the number of AMI admissions between two years decreases by more than 90 percent

[13] The closed emergency hospitals are Löwenströmska, Nacka, Finspång, Simrishamn, Landskrona, Strömstad, Falköping, Kristinehamn, Säffle, Sala, Fagersta, Sandviken, Söderhamn, Härnösand, Boden and Luleå hospitals

[14] Lindbom (2013) investigates protests movements in relation to the hospital closures over the same time period. Moreover, Landstingsförbundet (2002) discusses Swedish emergency hospital closures between 1992 and 2000.

[15] Note that the hospitals are plotted on different axes.

Figure 3.2. Number of visits at closing hospitals and their referral hospitals over time



NOTE.— Data source: Swedish National Board for Health and Welfare. The plots on the left show the monthly number of AMI visits at hospitals with closing emergency rooms (blue lines and left y-axis) and referral hospitals (red lines and right y-axis). The plots on the right show the corresponding six-month moving averages of the same plots (three leads and three lags).

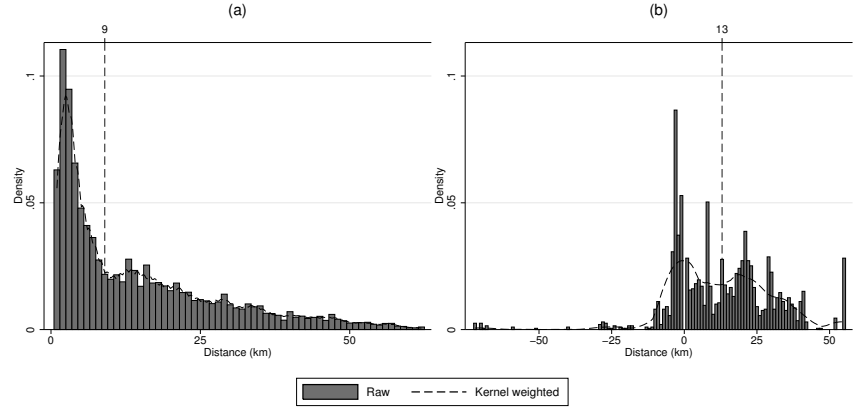
The panel on the left in Figure 3.3 shows the distribution of distance from a home hospital in the data aggregated over all the years. Approximately 95 percent of the population lives within a sixty-kilometer radius of their home hospital with a median distance of nine kilometers.[16] As the distance distribution is highly right skewed, I trim the upper five percentiles of the distribution in order to have a more homogeneous sample and to avoid introducing estimation problems from extreme outliers. This restriction mainly affects individuals living in the rural parts of northern Sweden.[17] The panel on the right in Figure 3.3 shows the corresponding distribution of the changes in distance generated from the emergency hospital closures. These changes in distance have reasonably good coverage over the support of the baseline distance distribution in the panel on the left.[18]

[16] I adopt the metric system as length measurement in this article. One English mile is approximately 1.61 kilometers.

[17] Since the inhabitants of this region are typically older and have a lower level of education than the overall Swedish population it is likely that these individuals also have lower underlying AMI survival probabilities. Hence omitting them would, if anything, give a lower bound on the estimates.

[18] See also Table A.1 for some descriptive statistics of the sample.

Figure 3.3. Distribution of distance and changes in distance to home hospital



NOTE.— Data source: Swedish National Board for Health and Welfare. Hospital distance is obtained by computing the distance from an individual patient's registered residence to his or her home hospital. The panel on the left shows the residential distance distribution to the home hospital for the analysis sample of AMI patients for the period 1990-2010. The panel on the right shows the distribution of the changes in distance generated by emergency room closures as defined in the data section. The dashed vertical lines indicate the median of the distribution. The dashed smoothed lines are kernel density estimates using a standard Epanechnikov kernel with a bandwidth of 3.9.

4 Empirical approach

Let D be the geographical distance from a patient's home to his or her designated (home) hospital and let y be a binary variable indicating whether an AMI patient survived the infarction or not. Specifically, y is coded as one if an individual survives a certain follow-up period after the AMI (occurred) and as zero if the individual died during this period. The empirical focus of this paper is to evaluate the impact of D on y.

There are several problems associated with empirically isolating an effect of the distance to a hospital on mortality risk. The main difficulty is, most likely, that the choice of where to live in relation to a hospital will depend on the health of the individual. Any such sorting of individuals with respect to health would bias any effect of distance downwards as individuals with poor health are more likely to take access to health care into consideration when choosing place of residence.[19] In addition to identification problems arising from the optimizing behavior of individuals, there are also other problems related to the organization of health care and the size and health characteristics of the population in the catchment areas. Average AMI survival rates at any given hospital might vary over time and with both the geographical position of the hospital and general hospital quality. Hospitals located in rural areas admit patients with on average both longer distances to the hospital and poorer health

[19] An upward bias could occur if individuals choosing to live further away from a hospital care in general less about their health relative to people living closer to a hospital due to e.g. heterogeneous health preferences. I do not rule out this possibility in the estimations but consider it less likely from a hypothetical point of view.

characteristics (e.g. older and with a lower level of education). In addition, the preparedness levels for emergency situations may vary between hospitals (e.g. the number of turnkey ambulances) as a consequence of the geographical size of the catchment area.

It is possible to control for heterogeneity across hospitals and common trends by including fixed-effects for these factors in a regression model. Moreover, as the data contains a number of individual health and socioeconomic characteristics these can also be added to the model in order to adjust for individual-level heterogeneity of the patient population within catchment areas. For an individual i experiencing an AMI at calendar time t with home hospital h the effect of distance on survival could hence be estimated using the following regression model:

$$y_{iht'} = \alpha + D_{iht}\beta + X_{it}\gamma + \lambda_h + \lambda_t + \nu_{iht}, \tag{4.1}$$

where λ_h and λ_t are hospital and time fixed effects and $t' \geq t$. The effect of distance, β, would be identified in this model if the individual error ν_{iht} is uncorrelated with the distance measure. Given that the health of individual patients is partly unobservable, residential sorting within catchment areas is likely to exist also after including X_{it}, hence invalidating this independence assumption.

To further address the problem of residential sorting, I utilize variation in individual distances to hospital generated from emergency hospital closures. Using this variation in distance to home hospital across time, I am able to identify the impact of distance on AMI survival among patients who lived in a closed hospital's catchment area at the time of the closure.

Specifically, I initially assume that the variation the closures generate can be used by estimating (OLS) the following model

$$y_{iht'} = \alpha + D_{iht}\beta + D_{iht-j}\beta_2 + X_{it}\gamma + \lambda_h + \lambda_t + \eta_{iht}, \tag{4.2}$$

where D_{iht-j} is the distance for an AMI patient at time t, j years before the AMI occurred.[20] Here, $\Delta_{D_{iht-j}} = D_{iht} - D_{iht-j}$ is the change in distance to the home hospital between the years t and $t-j$. For all patients living in the catchment area of a non-closing hospital, these distances are the same, i.e. $D_{iht} = D_{iht-1}$. These latter patients do not contribute to the identification of

[20] Variation in distance to home hospital may hypothetically emerge from two different sources; closures and migration. Specifically, consider the following distance-generating functions for time periods t and $t-j$, $d_t(coord_{it}, coord_{ht})$, $d_{t-1}(coord_{it-j}, coord_{ht-j})$, where the first argument in the functions is the patient's residential coordinates and the second argument is the coordinates of the patient's home hospital. Now, given that a patient in the year of the closure ($t-j$) does not migrate between the two time periods (i.e. $coord_{it} = coord_{it-j}$) only a switch of home hospital may result in a distance to hospital change. Hence, under the assumption that individuals do not selectively migrate between the two time periods, the change in distance should be unrelated to individual AMI survival probabilities, conditional on the pre-closure distance.

the effect of distance but are still included as they increase the precision of the estimated control variable parameters.

The primary justification for the identification strategy is that individuals cannot immediately adapt to the changing health care environment caused by the decisions of regional authorities to close certain hospitals. The total number of data observations experiencing a change in distance will vary depending on the length of the time window between the closure and the AMI. However, extending the time window to increase the number of patients that are affected also increases the risk of endogenous reactions to the closures, such as selective migration, and may bias the estimation results. Hence, the credibility of the assumption of no endogenous response decreases with the lag j.

In model (4.2) the distance and the lagged distance are both included linearly. This specification is highly restrictive since the outcome y is a binary variable. To increase the validity of the regression model, I therefore relax the linearity restriction by instead including a set of indicator variables for each ten-kilometer distance. Specifically, I estimate the following model (OLS)

$$y_{iht'} = \alpha + I^m_{iht}\beta^m + I^m_{iht-j}\beta^m_2 + X_{it}\gamma + \lambda_h + \lambda_t + \eta_{iht}, m = 1, ..., M, \qquad (4.3)$$

where

$$I^m_{iht-j} = 1((m-1)\times 10) < D_{iht} \leq m \times 10), m = 1, ..., M \text{ and } j = 0, 1.$$

As the emergency room closures also generated distance *cuts* to their home hospital for some patients, it is possible to investigate the symmetry of the effect of distance. One way of investigating effect symmetry is to regress the effect of a positive change and a negative change separately and test whether the coefficients differ. Specifically, I estimate the following model

$$y_{iht'} = \alpha + \delta_1(\Delta_{D_{it}} \times I^-_\Delta) + \delta_2(\Delta_{D_{it}} \times I^+_\Delta) + X_{it}\gamma + \lambda_h + \lambda_t + \eta_{iht}, m = 1, ..., M \qquad (4.4)$$

where $I^+_\Delta = 1(\Delta_{D_{it}} > 0)$ and $I^-_\Delta = 1(\Delta_{D_{it}} < 0)$. In addition, I estimate the restricted version of the model in which $\delta_1 = \delta_2 = \delta$. To test the symmetry of the estimated effect, I simply perform a standard Wald test of equality of δ_1 and δ_2.

Finally, since the dependent variable in the models is dichotomous, the linear probability models are an approximation of a true but unknown data generating process. In an attempt to test the validity of the model approximation, all the results below are also estimated using logit regression models. The results remain qualitatively unchanged by this particular change in specification.

5 Results

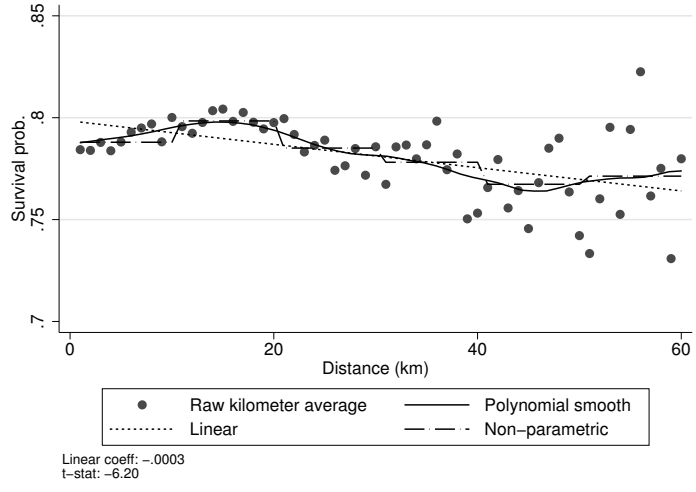
This section presents results from the estimation of the models discussed in the previous section. I begin with a presentation of the main results and sub-

sequently proceed with presenting the results from a number of extensions to corroborate the mechanisms of the estimated effects.

5.1 Descriptive results

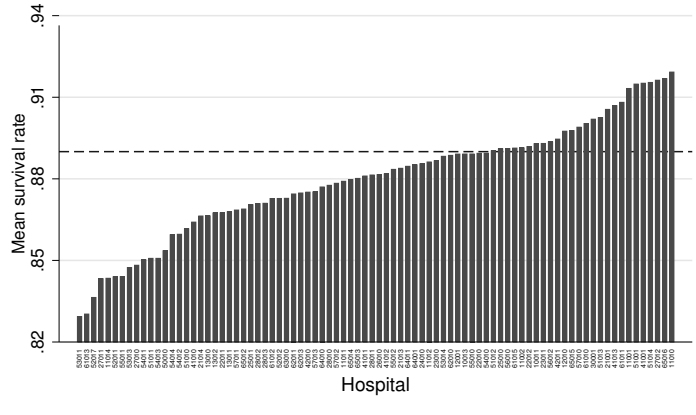
I first present a descriptive analysis of the observed distance-survival relationship in the data. Figure 5.1 plots correlations of distance to home hospital and AMI survival rates for different parameterizations. Specifically, the gray dots indicate the average survival rate for each kilometer to hospital while the dotted, dashed and solid lines illustrate the relationship under a linear regression model, a locally smoothed and a kernel weighted parameterization, respectively. The figure suggests a negative, albeit weak, correlation between distance and AMI survival with slightly higher survival rates for individuals living closer to their home hospital. The estimate from the linear model, reported below the plot, suggests a decreased survival probability of 0.03 percentage points for each additional kilometer a patient resides from his or her home hospital. With a mean survival rate in the analysis sample of about 78 percent, this is clearly a small difference. However, the upward sloping survival trend at the lower end of the distance distribution raises some doubt about whether the plotted relationship can be interpreted causally. For example, Figures 5.2 and 5.3 show substantial heterogeneity in survival rates both across hospitals and over time. If these factors are correlated with the distance to the home hospital, any estimated effect of distance will be confounded unless they are accounted for.

Figure 5.1. Correlations of distance to home hospital and survival probability from an AMI under various parametric assumptions



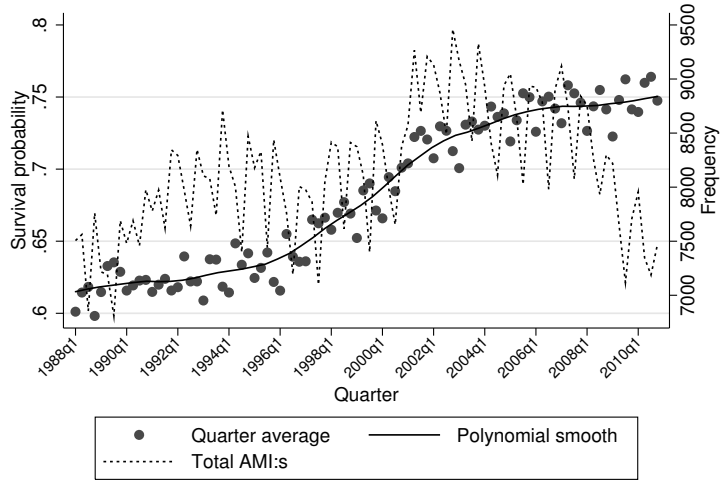
NOTE.— Data source: Swedish National Board for Health and Welfare. The figures display the observed correlation between distance to home hospital and survival probability for the sample of AMI patients used in the empirical analysis under different parametric assumptions. The dots indicate the raw kilometer average while the lines show the relationship for different models; the dotted line shows the linear relationship, the dashed line the non-parametric relationship with a dummy indicator for each ten kilometers and the solid line shows a kernel density estimator using a standard Epanechnikov kernel with a bandwidth of 3.9.

Figure 5.2. Survival probability by home hospital



NOTE.— Data source: Swedish National Board for Health and Welfare. AMI survival probability for each hospital is measured as the share of individuals who were the subject of an AMI living in the hospitals catchment area and were discharged from the hospital alive. Individual hospitals are shown on the x-axis in ascending order with respect to survival probability aggregated over the period 1990-2010. The horizontal dashed line indicates hospital average survival probability in the sample of hospitals.

Figure 5.3. AMI frequency and average survival rates, 1990-2010



NOTE.— Data source: Swedish National Board for Health and Welfare. The figure plots (on the left y-axis) average survival rates as a raw quarterly average and as a smoothed kernel density estimate using an Epanechnikov kernel with a bandwidth of 3.8. The quarterly number of AMIs over the period is plotted on the right y-axis.

5.2 Main results

Table 5.1 presents the main results from the estimation of the effect of distance for different models using the full analysis sample (scaled with a factor of ten for presentation reasons). The first through third columns include only the observed *current* distance to the home hospital, i.e. the observed distance in the year the AMI occurred. The first column reproduces the linear estimate of the distance-survival correlation from Figure 5.1, while the second and third columns include covariate adjustments for a number of health-related characteristics and hospital and calendar time fixed effects, respectively. The estimated distance coefficient remains approximately the same in all specifications, implying relatively small variations in average AMI survival rates over different distances to home hospital, despite covariate adjustments.

The fourth column of Table 5.1 also includes the *lagged* distance for patients in the year before they were the subject of an AMI, corresponding to equation (4.2) with $j = 1$ from the empirical section. The coefficient on current distance now increases in magnitude by a factor of four while the lagged distance coefficient is estimated to be slightly lower and with opposite sign. Comparing over specifications, note that netting out the predicted effect for individuals with the same distance in both periods reproduces, as expected, the distance coefficient displayed in column (3). The estimated current distance coefficient is now interpreted as the marginal effect for an AMI patient of in-

creasing the distance to his or her home hospital by ten kilometers. Hence, this estimate shows a difference in AMI survival probability of about 15 percent for individuals at the lower and upper support of the distance distribution — i.e. zero and sixty kilometers — at mean survival rates.

Table 5.1. *Estimated effects of distance on AMI survival probability from emergency room closures: Different estimators*

	Estimator					
	$\hat{\beta}_{OLS}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{D}$	$\hat{\beta}_{NPD}$	$\hat{\beta}_{NPD}$
Current distance	-0.004***	-0.003***	-0.005***	-0.021***		
	(0.001)	(0.001)	(0.001)	(0.005)		
Lagged distance				0.016***		
				(0.005)		
Current Distance Dummies						
11-20 km					0.015	0.002
					(0.020)	(0.020)
21-30 km					-0.036*	-0.041**
					(0.020)	(0.020)
31-40 km					-0.064***	-0.051**
					(0.024)	(0.023)
41-50 km					-0.073**	-0.086***
					(0.032)	(0.031)
51-60 km					-0.109**	-0.115**
					(0.047)	(0.046)
Lagged Distance Dummies						
11-20 km					-0.002	0.001
					(0.020)	(0.020)
21-30 km					0.033*	0.031
					(0.020)	(0.020)
31-40 km					0.055*	0.036
					(0.024)	(0.023)
41-50 km					0.053	0.065**
					(0.032)	(0.031)
51-60 km					0.087*	0.095**
					(0.048)	(0.047)
Covariates		✓	✓	✓		✓
Fixed effects			✓	✓		✓
Observations	331,515	331,515	331,515	331,515	331,515	331,515

NOTE.— The table reports point estimates (standard error) of the effect of distance on survival probability from an acute myocardial infarction for different estimators as explained in the empirical section and using the full sample of all AMIs over the time period 1990-2010. Geographical coordinates are obtained by linking the patient/death data to the population register. Distance is obtained by computing the distance from an individual patient's registered residence to his or her home hospital. For more information see the data section. The current distance variable is defined as the residential distance in kilometers from an individual's home hospital in the current year while lagged distance corresponds to the same distance in the previous year. The last three columns — $\hat{\beta}_D$ and $\hat{\beta}_{NPD}$ — estimate the effect of distance using variation in the distance to an individual's home hospital arising from closures of emergency rooms as explained in the data section. The last two columns include a number of distance dummies for each ten kilometers instead of the linear specification. Included covariates are gender, age, the number of previous hospitalizations (AMIs) and the number of years since the last hospitalization (AMI). Fixed effects include hospital and calendar year dummies. Standard errors are estimated using a robust covariance matrix. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

One theoretical prediction for the health effects of geographical access to health care is that any such effect should be monotonous over the distance to hospital. The last two columns of Table 5.1 evaluate this prediction by relaxing the assumption of linearity of the effect by replacing the continuous distance measure with a set of dummy variables for each ten-kilometer distance (with the closest distance group, 0-10 kilometers from the hospital, as reference category). The results from estimating the model without and with the full set of controls are reported in the right and left of these columns respectively. The estimation result, irrespective of the inclusion of controls, shows a remarkably clear monotonous pattern on AMI survival probability of experiencing a change in distance to home hospital.[21] The estimated coefficients are highly significant and the pattern corresponds quite well with a linear specification, except for distances between 11 and 20 and 21 and 30 kilometers where there seem to be a large shift in AMI survival. In other words, this finding suggests that there is a critical threshold in the distance where the risk of AMI mortality increases dramatically.[22] Thus, the conforming of the results to the theoretical prediction with respect to the pattern of the effect of distance provides strong evidence for both the empirical design and the causal interpretation of the estimated effect.

Under the more restrictive assumption of additive separability between distance to hospital and health, the difference of the coefficients of Table 5.1 can be given a causal interpretation. Table 5.2 tabulates all possible combinations of these differences for given lagged and current distances under the additional assumption of homogeneity of the effect of distance across lagged distance. For instance, the table reports that the estimated probability of surviving an AMI for a patient who experienced an increase in the distance to home hospital from 11-20 kilometers to 41-50 kilometers is 8.6 percentage points lower compared to an individual who lived close to his or her home hospital in both periods. These effects are also graphically presented using a contour plot in Figure 5.4. Specifically, the brighter (darker) areas of the plot show for which combinations of lagged and current distance AMI survival probabilities decrease (increase). Going from the upper-left corner (illustrating the effect of an *increase* in geographical distance of 50 kilometers) to the lower-right corner (illustrating the effect of a *decrease* in geographical distance of 50 kilometers)

[21] It is interesting to note that including health controls in the last column does not change the results qualitatively. This finding suggests that the endogeneity between the changes in distance and pre-closure distance may not be a severe problem in this application.

[22] This threshold is empirically plausible since according to Nationellt register för hjärtstopp (2011) if medical assistance is not received within 15 minutes after suffering from a cardiac arrest, death is almost certain. Doing a back-of-the envelope calculation assuming that an ambulance has an average speed of 100 km/h it will take emergency medical personnel about 15 minutes to travel a distance of 25 kilometers, which is exactly in the middle of the empirical threshold where the effect of distance kicks in.

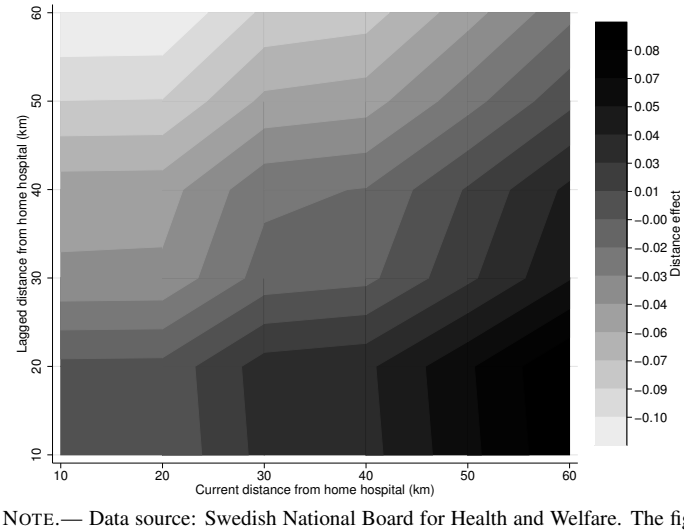
the figure shows a clear monotonous and symmetric pattern of the effect of distance.

Table 5.2. *Estimated effects of distance on AMI survival probability for different pre-closure hospital distances*

	Lagged Distance (km)					
Current distance (km)	0-10	11-20	21-30	31-40	41-50	51-60
0-10	0.000	0.001	0.031	0.036	0.065	0.095
11-20	0.002	0.003	0.033	0.038	0.067	0.097
21-30	-0.041	-0.041	-0.010	-0.005	0.024	0.054
31-40	-0.051	-0.050	-0.019	-0.015	0.014	0.044
41-50	-0.086	-0.086	-0.055	-0.050	-0.021	0.009
51-60	-0.115	-0.114	-0.083	-0.079	-0.050	-0.020

NOTE.— The table shows the estimated effect derived from the last column in Table 5.1 of experiencing a change in home hospital distance from a distance indicated in a given column to a distance indicated in a given row. For instance, an individual who lived between 21 and 30 kilometers from his or her home hospital that subsequently closed the following year which led to an increase in distance to the new home hospital of between 41-50 kilometers had an estimated 3.9 percentage points lower probability of surviving an AMI compared to individuals who lived closer to their home hospital in both time periods. The effect on the diagonal is the estimated effect for individuals who did not experience a change in distance. Geographical coordinates are obtained by linking the patient/death data to the population register. Distance is obtained by computing the distance from an individual patient's registered residence to his or her home hospital. For more information see the data section. See Table 5.1 for estimation details. See also Figure 5.3 for a graphical illustration of the effect.

Figure 5.4. Contour plot of the estimated effects of distance



NOTE.— Data source: Swedish National Board for Health and Welfare. The figure shows a three-dimensional contour plot of the estimated effect from Table 5.1 and Table 5.2. The darker areas in the plot correspond to a lower probability of survival while a brighter area corresponds to a higher probability of survival. The figure can be interpreted as showing the estimated effect of going from a given distance to home hospital in time period $t-1$ indicated on the y-axis to a given distance to home hospital in time period t indicated on the x-axis. See the data section for a definition of a home hospital, the computation of distance to home hospital and an explanation of the sample used in the analysis and the empirical section for an explanation of the estimated effects.

Finally, I have also estimated the model from equation (4.4) to statistically test the symmetry of the effect of distance. The result from this exercise is shown in Table 5.3. The first column of the table reports the estimated coefficients for the change in distance and an indicator variable for a negative change interacted with the change in distance. Similarly, the second column reports results from regressing AMI survival on the *absolute* change in distance interacted with a dummy variable for a positive and a negative change respectively. Since I cannot reject the hypothesis that the coefficients are the same for any conventional statistical significance levels ($p = 0.7990$), this suggests that the magnitude of the effect of distance is the same, regardless of whether an individual experienced an increase or a decrease in the distance to hospital.

Table 5.3. *Estimated effects of distance on AMI survival probability: Symmetry of the effect*

	(1)	(2)
Δ_D	-0.014***	
	(0.004)	
$I[\Delta_D < 0] \times \Delta_D$	0.004	
	(0.015)	
$I[\Delta_D > 0] \times Abs[\Delta_D] = b_1$		-0.014***
		(0.004)
$I[\Delta_D < 0] \times Abs[\Delta_D] = b_2$		0.010
		(0.014)
Test $b_1 = -b_2$		
χ^2-statistic (1 df)		0.06
p-value		0.7990
Observations	331,515	331,515

NOTE.— The table reports point estimates (standard error) from a linear regression model including the full sample of AMI patients as explained in the empirical section over the time period 1990-2010. Geographical coordinates are obtained by linking the patient/death data to the population register. Distance is obtained by computing the distance from an individual patient's registered residence to his or her home hospital. For more information see the data section. The $I[\cdot]$ functions are indicator functions that evaluate to one if the arguments within the brackets are true and zero otherwise. The lower part of the table displays the statistics from a Wald test on parameter equality between the effects of distance from a positive and a negative change in distance, (b_1 and b_2) respectively. Included covariates are gender, age, the number of previous hospitalizations (AMIs) and the number of years since the last hospitalization (AMI). Fixed effects include hospital and calendar year dummies. Standard errors are estimated using a robust covariance matrix. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

5.3 Extensions and robustness checks

The specific outcome studied so far has been the probability of surviving until being discharged from a hospital after suffering an AMI. Table 5.4 presents the results for other survival definitions using the same analysis sample and the preferred empirical specification from the last column of Table 5.1. The first column of the table reproduces the main results while the second column reports the results for the probability of surviving the initial phase before being admitted to a hospital, i.e. the out-of-hospital phase. The last four columns

show the result when the outcome is defined as a binary indicator for whether the patient was alive after one day, one month, a hundred days and one year from the AMI, respectively. As a complement to this analysis, Figure A.7 plots the distribution of deaths in the sample for the first hundred days after the AMI occurred. Day one is excluded in the figure for scaling reasons as the majority of all deaths occur on the first day of the AMI.

The results from the estimation reveal interesting effect mechanisms; first, comparing the first two columns of the table, it is clear that most of the effect on survival seem to arise from an increased probability of out-of-hospital mortality (about 75 percent).[23] This finding is not unexpected since a longer geographical distance to hospital will increase both the time it takes to reach the patient and the time it takes to transport him or her to the hospital. Furthermore, the last four columns of Table 5.4 investigate whether the estimated effect is primarily driven by patients in very poor health, in which the additional distance is simply "the straw that broke the camel's back" — i.e. a harvesting effect — by comparing results from different survival time horizons after the AMI. Interestingly, the pattern in the last four columns of 5.4 shows that distance to hospital slightly *increased* the probability of surviving more than one month, compared with the effect of surviving only the first day. Importantly, this result suggests that the estimated effect is not due to harvesting, in which case we would rather see a substantial effect just after the AMI and thereafter a diminishing and even reversed sign of the effect for the more long-term outcomes.

[23]The estimated coefficients are much smaller in magnitude and barely statistically significant when using the probability of in-hospital mortality as the outcome. However, the monotonous effect pattern remains unchanged.

Table 5.4. *Estimated effects of distance on AMI survival probability from emergency room closures: Different survival outcomes*

	Survival Outcome					
	Hospitalization		Survival time			
	Baseline (AMI=1)	OOH Survival (AMI$\neq$ 2)	Survives $>$ 1 day	Survives $>$ 30 days	Survives $>$ 100 days	Survives $>$ 365 days
Current Distance Dummies						
11-20 km	0.002	0.001	-0.003	0.001	0.003	-0.003
	(0.020)	(0.018)	(0.019)	(0.020)	(0.020)	(0.021)
21-30 km	-0.041**	-0.031*	-0.030	-0.041**	-0.053***	-0.070***
	(0.020)	(0.018)	(0.019)	(0.020)	(0.020)	(0.020)
31-40 km	-0.051**	-0.057***	-0.054**	-0.055**	-0.064***	-0.075***
	(0.023)	(0.021)	(0.022)	(0.023)	(0.024)	(0.024)
41-50 km	-0.086***	-0.052*	-0.061**	-0.089***	-0.085***	-0.098***
	(0.031)	(0.029)	(0.030)	(0.032)	(0.032)	(0.033)
51-60 km	-0.115**	-0.090**	-0.104**	-0.166***	-0.158***	-0.159***
	(0.046)	(0.044)	(0.044)	(0.047)	(0.047)	(0.048)
Lagged Distance Dummies						
11-20 km	0.001	-0.006	0.005	0.002	-0.001	0.004
	(0.020)	(0.018)	(0.019)	(0.020)	(0.020)	(0.021)
21-30 km	0.031	0.019	0.021	0.032	0.043**	0.060***
	(0.020)	(0.018)	(0.019)	(0.020)	(0.020)	(0.020)
31-40 km	0.036	0.042**	0.041*	0.039*	0.048**	0.058**
	(0.023)	(0.021)	(0.022)	(0.023)	(0.024)	(0.024)
41-50 km	0.065**	0.029	0.039	0.067**	0.062*	0.074**
	(0.031)	(0.029)	(0.030)	(0.032)	(0.032)	(0.033)
51-60 km	0.095**	0.068	0.082*	0.147***	0.137***	0.137***
	(0.047)	(0.044)	(0.044)	(0.047)	(0.048)	(0.048)
Observations	331,515	331,515	331,515	331,515	331,515	331,515

NOTE.— The table reports point estimates (standard error) of the effect of distance on survival probability from an acute myocardial infarction as explained in the empirical section and using the full sample of all AMIs over the time period 1990-2010. Geographical coordinates are obtained by linking the patient/death data to the population register. Distance is obtained by computing the distance from an individual patient's registered residence to his or her home hospital. For more information see the data section. The current distance variable is defined as the residential distance in kilometers from an individual's home hospital in the current year while lagged distance corresponds to the same distance in the previous year. Outcomes are defined as indicator functions for being alive when discharged from the hospital following the infarction or surviving until admitted (in the first two columns) and as being alive after a certain time after the AMI occurred (in columns 3-5). Reported coefficients in each column are a number of distance dummies for each ten kilometers. Included covariates are gender, age, the number of previous hospitalizations (AMIs) and years since the last hospitalization (AMI). Fixed effects include hospital and calendar year dummies. Standard errors are estimated using a robust covariance matrix. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

Another extension of the main results is to investigate whether the estimated effects of distance vary over the time span between a hospital closure and an AMI. In the empirical section, I discussed potential hospital closure coping behaviors of both AMI patients and the health care administrations, i.e. strategies that these agents may have conducted to counteract any percieved effects of distance subsequent to the closures. For instance, patients with relatively poor health who experienced reduced access to emergency health care may have decided to move closer to the new home hospital. Another possibility

is that health care authorities may ex post have invested in emergency health care over time (e.g. by building additional ambulance stations). Both these potential coping behaviors would then serve to diminish the effect on survival over calendar time from the hospital closure.

Table 5.5 presents estimation results for AMI patients living in a region in year t where an emergency hospital closure occurred $t - j$ years earlier, with $j = 1, ..., 5$ and where $j = 1$ has been the baseline case studied so far. The sample size is different as the five first years of the sampling period, i.e. 1987-1992, are dropped from the analysis. These five years are excluded in all the specifications in the table in order to facilitate a comparison of the results. In each column the header indicates the specific number of years from closure and the reported results are based solely on variation in distance for patients who experienced a shift in distance to home hospital for this particular number of years following closure.

The results from the estimation are striking; there is only a clear effect of distance for the first year after a hospital closure. At each subsequent leading year, the effect is smaller in magnitude and statistically insignificant when measured with similar precision. This pattern indicates that long-run effects of distance from the closures on AMI survival are unlikely to prevail — perhaps as a consequence of various coping strategies among the agents involved. This result is reassuring for policy-makers since, besides from the initial shock, the hospital closures have not entailed a long-lasting elevated AMI mortality pattern.[24]

[24] A back-of-the-envelope analysis might bring some further insights regarding the cost-benefit trade-off of the closures. In particular, I ran a regression of the survival measure including a dummy variable for being affected by a hospital closure on the right hand side (along with the other covariates) and subsequently related the estimated closure coefficient to the average survival rates and AMI incidence in the relevant population. The effect of being affected by a hospital closure reduced the average survival probability with an estimated two percentage points, i.e. from 0.79 to 0.77 at mean survival rates. As the annual average number of AMIs is about 20,000, this estimate suggests that about 320 extra deaths would have occurred had the closures affected the full AMI patient population. However, as the underlying population of the relevant catchment areas is only ten percent of the total AMI population in a given year, the closures caused only 32 additional deaths. Hence, the total of 16 closures in the data meant an additional two deaths per closure. Assuming that the value of a statistical life is about €2 million, the closures could thus be deemed cost-effective if the cost savings were more than €4 million per closed hospital.

Table 5.5. *Estimated effects of distance on AMI survival probability from emergency room closures: Short and long-term effects*

	Time Horizon (years from closure)				
	One	Two	Three	Four	Five
Current Distance Dummies					
11-20 km	-0.002	-0.013	-0.037**	-0.040**	0.031
	(0.019)	(0.019)	(0.018)	(0.018)	(0.019)
21-30 km	-0.036*	-0.024	-0.017	-0.000	-0.030
	(0.019)	(0.019)	(0.018)	(0.019)	(0.020)
31-40 km	-0.038*	-0.001	-0.016	-0.004	-0.010
	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)
41-50 km	-0.082***	-0.052	-0.034	-0.021	-0.027
	(0.031)	(0.032)	(0.030)	(0.029)	(0.029)
51-60 km	-0.116**	0.013	-0.061	0.087**	0.080
	(0.046)	(0.044)	(0.049)	(0.044)	(0.050)
Lagged Distance Dummies					
11-20 km	0.004	0.014	0.039**	0.042**	-0.029
	(0.019)	(0.019)	(0.018)	(0.018)	(0.019)
21-30 km	0.025	0.013	0.007	-0.010	0.019
	(0.020)	(0.019)	(0.018)	(0.019)	(0.020)
31-40 km	0.023	-0.015	0.001	-0.011	-0.005
	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)
41-50 km	0.062**	0.033	0.014	0.002	0.008
	(0.032)	(0.032)	(0.030)	(0.029)	(0.030)
51-60 km	0.093**	-0.037	0.037	-0.110**	-0.103**
	(0.046)	(0.044)	(0.049)	(0.044)	(0.050)
Observations	285,883	286,030	286,020	286,120	285,988

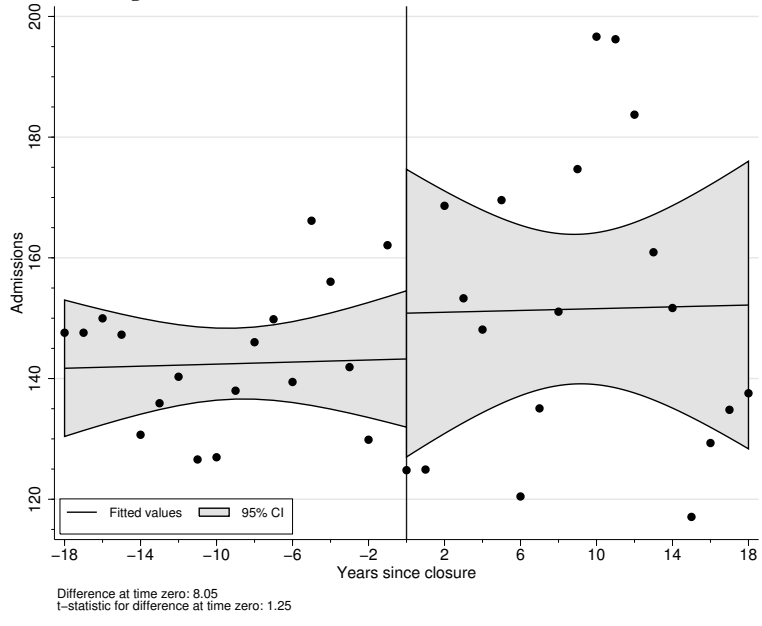
NOTE.— The table reports point estimates (standard error) of the effect of distance on survival probability from an acute myocardial infarction as explained in the empirical section and using the full sample of all AMIs over the time period 1990-2010. Geographical coordinates are obtained by linking the patient/death data to the population register. Distance is obtained by computing the distance from an individual patient's registered residence to his or her home hospital. For more information see the data section. The current distance variable is defined as the residential distance in kilometers from an individual's home hospital in the current year while lagged distance corresponds to the same distance in the previous year. Outcome is defined as an indicator function for being alive when discharged from a hospital following the infarction. Each specification pertains to a specific time horizon from an emergency room closure (the number of lagged years). Reported coefficients in each column are a number of distance dummies for each ten kilometers. Included covariates are gender, age, the number of previous hospitalizations (AMIs) and the number of years since the last hospitalization (AMI). Fixed effects include hospital and calendar year dummies. Standard errors are estimated using a robust covariance matrix. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

A potential problem caused by sampling only individuals observed to have an AMI is that the closures may have endogenously changed the population at risk for an AMI by affecting the health behavior of the population in the catchment area of the closed hospital. Specifically, admissions for other reasons than AMI may change a patient's general perception of his or her health risks and induce a more proactive behavior. In this respect, the closures may have affected general prevention behavior in the population through the reduced access to health care and, hence, the population at risk for an AMI. This

sample selection might, in turn, cause downward bias to any estimated effects of distance for the full population.

To evaluate whether the closures affected the population at risk for an AMI, I investigate the impact of the closures on AMI incidence rates in the catchment areas of the closing hospitals. I conjecture that, if the closures did not affect the population at risk for an AMI, we would not expect any change in AMI incidence rates around the time of the hospital closure. Figure 5.5 shows the empirical relationship between the number of hospital admissions in municipalities affected by a hospital closure in years from the time of the closures, after adjusting for general calendar time trends in AMI incidence. The dots in the figure indicate average numbers across municipalities and the solid line plots the piece-wise linear relationship allowing for a discontinuity in the year of the closure (indicated by the vertical line). The figure reveals a small increase in AMI incidence after, as compared to before, the hospital closures. However, the change is not significantly different from zero at any conventional levels of statistical significance. Moreover, the relationship shows no indication of a discontinuous jump in AMI incidence around the time of the closure. From this exercise, I conclude that severe sample selection from endogenous health prevention responses seem to be unlikely.

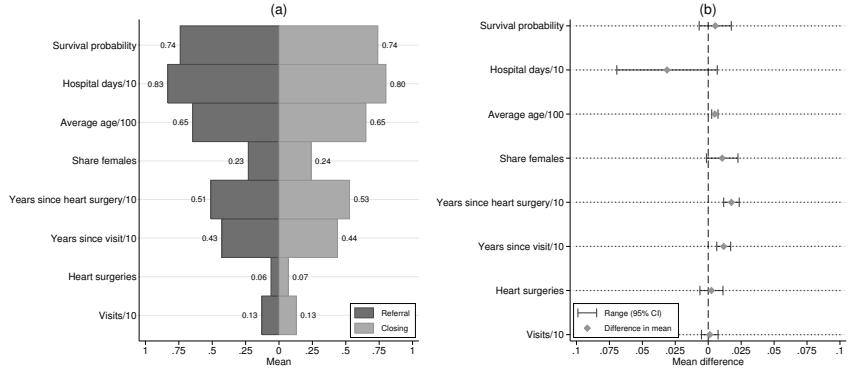
Figure 5.5. Effect of the closures on AMI incidence



NOTE.—Data source: Swedish National Board of Health and Welfare. The figure shows the relationship between the average number of admissions in a closing hospital's catchment area over time since the closure occurred, adjusting for calendar time trends in AMI incidence. The dots show the average values for each particular time period and the solid line pertains to a piece-wise linear relationship allowing for discontinuity at the time of closure, indicated by the vertical line. The shaded area marks the 95% confidence interval of the linear estimate.

Finally, in order to interpret the estimated effects of distance as externally valid for the full population of AMI patients, a necessary assumption is that the regions where emergency hospital closures occurred are comparable to non-closing regions. One potential threat to this external validity of the results would be that patient populations affected by the closures are on average different compared to the unaffected patient populations in terms of health characteristics. In order to investigate this concern, Figure 5.6 shows the average values of a number of aggregate health characteristics for closing and referral hospital catchment areas (left panel) and their difference along with a 95-percent confidence band (right panel) for the years prior to the hospital closures. The results are reassuring; both types of regions have, on average, similar health characteristics, indicating that regions where closures occurred are observationally unrelated to underlying patient population health characteristics.

Figure 5.6. Aggregate health indicators in closing and referral hospital catchment areas



NOTE.— Data source: Swedish National Board of Health and Welfare. See the data section for a definition of a home hospital, closing hospital and referral hospital. The left panel of the figure shows the average values for a number of health indicators for each type of region and the right panel shows the cross-regional mean difference for each of these indicators (point estimate and 95 percent confidence band). Some variables have been scaled to make the plot readable.

6 Summary and concluding remarks

Ischemic heart disease, with acute myocardial infarction (AMI) as one of its more serious manifestations, is the most common cause of death in Sweden as well as in most of the Western world. Since infarctions often occur relatively unexpectedly and rapid medical assistance is fundamental for recovery, the probability of surviving an AMI is highly dependent on a well-functioning health care system which can provide quick access to health care in emergency situations. This is particularly important in relatively sparsely populated coun-

tries like Sweden where distances to medical care facilities with emergency room capacity vary greatly between individual residents.

While the question of the adequacy of emergency health care in terms of providing sufficient coverage over longer geographical distances should be of a high policy relevance, it is nevertheless a difficult task to empirically establish a causal relationship between health care quality and a patient's proximity to a hospital. Apart from time trends and hospital heterogeneity in health care quality, which might confound simple correlations between the distance to a hospital and health outcomes, more sophisticated inferential problems exist relating to the endogenous sorting of the place of residence of individuals with respect to health and distance to hospital. In particular, if individuals choose to live close to a hospital for reasons related to unobserved health factors, any estimated effects of distance on health outcomes would be biased. Moreover, a lack of data on out-of-hospital mortality, to which any hypothetical effect of distance is likely to be correlated, may complicate the interpretation of the results based on only inpatient information as individuals living further away from a hospital are less likely to be admitted.

In this paper, I evaluate the existence and magnitude of the impact of geographical access to health care on health using AMI patients as the empirical application. I circumvent both the problem of missing mortality data and the residential sorting of individuals by; *i)* adding nationwide information on AMI deaths from the Swedish national causes of death registry to supplement the national inpatient registry; and *ii)* utilizing geographical variation in distance to hospital arising from a number of emergency hospital closures during a period of strong centralization of the publicly administered Swedish health care sector. In Sweden, virtually all inpatient health care is publicly provided and financed, implying that competition effects on the number and placing of hospitals in the country will be negligible. Moreover, as individuals are directed to a specific hospital based on their place of residence, I can utilize variation over time regarding which hospital patients are directed to in order to obtain exogenous shifts in individual distances to hospital. As I include the full AMI population over a twenty-year period, i.e. both admitted patients and patients who die before reaching a hospital, the empirical design accounts for both of the presumably most serious identification problems in evaluating the health effects of geographical access to health care.

Using data on more than 300,000 AMI cases and 16 emergency hospital closures over the period 1990-2010, I find a substantial, significant and monotonously decreasing effect of the proximity to an emergency hospital on AMI survival probability. In particular, patients who experienced an increase in the distance to their home hospital of between 51 and 60 kilometers ran an estimated 11.5 percentage points (or 15 percent) lower risk of surviving the AMI than patients who lived within ten kilometers of their home hospital during both periods (at mean survival rates). This effect is primarily driven by an increased risk of out-of-hospital mortality. Moreover, much smaller effects

are found when estimating the effects of distance based on actual distances to hospital, indicating that selective residential sorting is likely to dilute the effect of distance. When varying the time window between the closures and the time the AMI occurred, I find that the effect is only statistically significant in the first year after the closures. Perhaps reassuring for policy makers, this therefore indicates that the closures only had a short-run effect which might later have been counteracted with various types of coping behavior among both individuals (migration) and health care authorities (increased investments in emergency care). Finally, as a number of patients experienced a cut in hospital distance due to the closures, I also investigate the symmetry of the effect of distance. I find that the effect is indeed reversed for patients who experienced a shorter distance to hospital and I cannot reject that the effect is symmetric.

To conclude, in times when health care expenditure increased in most Western countries, Sweden went in the opposite direction and reduced its health care spending by approximately 11 percent between 1990 and 2000. In the European Union, only Finland reduced its total health care expenditure during the same period. Most of the cost savings were derived from structural changes in the health care sector; from inpatient to outpatient care and from the agglomeration and centralization of many care services — in particular emergency surgical procedures. These tendencies were perhaps necessary given the public sector budget deficits, a consequence of the economic depression in Sweden at the time, but the question remains whether the reduction in health care expenditure came at the cost of a decrease in access to health care among individuals living in more remote parts of the country. The results in this paper provide some evidence for the notion that geographical access to health care does have an impact — albeit only temporarily — on the survival rates of AMI patients, and hence that health care centralization may have important side effects that should be taken into account. Perhaps more importantly, this effect of distance may be more persistent in other countries with more unregulated health care sectors due to the strategical positioning of profit-maximizing hospitals. Specifically, hospitals in these markets may abandon geographical areas in which aggregate incidence rates of costly emergency health care is higher thus creating a "health care desert" similar to the phenomenon of food deserts recognized in many countries.

References

AMERICAN HEART ASSOCIATION (2012). *Heart Disease and Stroke Statistics – 2013 Update*. Annual report, American Heart Association, http://my.americanheart.org/professional/General/AHA-Heart-Disease-and-Stroke-Statistics-2013-Update_UCM_445937_Article.jsp.

BACHMANN, J., MCDONALD, G. and O'BRIEN, P. (1986). A Study of Out-of-Hospital Cardiac Arrests in Northern Minnesota. *The Journal of the American Medical Association*, **256** (4), 477–483.

DELLBORG, M., ERIKSSON, P., RIHA, M. and SWEDBERG, K. (1994). Declining Hospital Mortality in Acute Myocardial Infarction. *European Heart Journal*, **15** (1), 5–9.

DRANOVE, D., DURAC, A. and SHANLEY, M. (1996). Are Multihospital Systems More Efficient? *Health Affairs*, **15** (1), 100–104.

EVANS-CUELLAR, A. and GERTLER, P. J. (2003). Trends In Hospital Consolidation: The Formation Of Local Systems. *Health Affairs*, **22** (6), 77–87.

FEINSTEIN, J. S. (1993). The Relationship Between Socioeconomic Status and Health: A Review of the Literature. *The Milbank Quarterly*, **71** (2), 279–322.

GILLUM, R. F. (1990). Geographic Variation in Sudden Coronary Death. *American Heart Journal*, **119** (2), 380–389.

GUSTO INVESTIGATORS (1993). An International Randomized Trial Comparing Four Thrombolytic Strategies for Acute Myocardial Infarction. *New England Journal of Medicine*, **329** (10), 673–682.

HAMILTON, B. and HO, V. (1998). Does Practice Make Perfect? Examining the Relationship Between Hospital Surgical Volume and Outcomes for Hip Fracture Patients in Quebec. *Medical Care*, **36** (6), 892–903.

HERLITZ, J. (2000). One Hundred Years' of Progress in the Treatment of Acute Myocardial Infarction. *Läkartidningen*, 97, 3369-3372, http://ltarkiv.lakartidningen.se/2000/temp/pda21465.pdf.

JULIAN, D. G. (1961). Treatment of Cardiac Arrest in Acute Myocardial Ischaemia and Infarction. *The Lancet*, **278** (7207), 840 – 844.

LANDSTINGSFÖRBUNDET (2002). *Svensk Hälso- och Sjukvård Under 1990-talet: Utvecklingstendenser 1992-2000*. Tech. rep., http://webbutik.skl.se/sv/artiklar/svensk-halso-och-sjukvard-under-1990-talet-utvecklingstendenser-19922000.html.

LINDBOM, A. (2013). *Waking Up the Giant? Hospital Closures and Electoral Punishment in Sweden*. mimeo, The Institute for Housing and Urban Research (IBF), Uppsala, Sweden.

LUFT, H. S., HUNT, S. S. and MAERKI, S. C. (1987). The Volume-Outcome Relationship: Practice-Makes-Perfect or Selective-Referral Patterns? *Health Services Research*, **22** (2), 157–182.

MAERKI, S. C., LUFT, H. S. and HUNT, S. S. (1986). Selecting Categories of Patients for Regionalization. Implications of the Relationship Between Volume and Outcome. *Medical Care*, **24** (2), 148–158.

NATIONELLT REGISTER FÖR HJÄRTSTOPP (2011). *Årsrapport 2011*. Annual report, Svenska Rådet för Hjärt-lungräddning, http://www.hlr.nu/sites/hlr.nu/files/attachment/Rapport%202011.pdf.

NORRIS, R. M. (1998). Fatality Outside Hospital from Acute Coronary Events in Three British Health Districts, 1994-5. *British Medical Journal*, **316** (7137), 1065–1070, United Kingdom Heart Attack Study Collaborative Group.

O'NEILL, L. (2003). Estimating Out-of-Hospital Mortality Due to Myocardial Infarction. *Health Care Management Science*, **6**, 147–154.

PELL, J. P., SIREL, J. M., MARSDEN, A. K., FORD, I. and COBBE, S. M. (2001). Effect of Reducing Ambulance Time on Deaths from Out-of-Hospital Cardiac Arrest: Cohort Study. *British Medical Journal*, **322** (7299), 1385–1388.

PIETTE, J. and MOOS, R. (1996). The Influence of Distance on Ambulatory Care Use, Death, and Readmission Following a Myocardial Infarction. *Health Services Research*, **31** (5), 573–591.

SOCIALSTYRELSEN (2009). *Myocardial Infarctions in Sweden 1987-2007*. Official Statistics of Sweden: Statistics — Hälso- och Sjukvård, http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/17858/2009-12-7.pdf.

SUCCI, M. J., LEE, S. Y. and ALEXANDER, J. A. (1997). Effects of Market Position and Competition on Rural Hospital Closures. *Health Services Research*, **31** (6), 679–699.

SVERIGES KOMMUNER OCH LANDSTING (2004). *Utvecklingen i Svensk Hälso- och Sjukvård : En Redovisning av hur Sjukvården Använder sina Resurser*. Tech. rep., Swedish Association of Local Authorities and Regions (SALAR).

SVERIGES KOMMUNER OCH LANDSTING (2008). *Utvecklingen i Svensk Hälso- och Sjukvård : Struktur och Arbetssätt för bättre resultat*. Tech. rep., Swedish Association of Local Authorities and Regions (SALAR).

THIEMANN, D. R., CORESH, J., OETGEN, W. J. and POWE, N. (1999). The Association Between Hospital Volume and Survival After Acute Myocardial Infarction in Elderly Patients. *The New England Journal of Medicine*, **340** (21), 1640–1648.

WOHLIN, L. (1998). Bankkrisens Upprinnelse. *Ekonomisk Debatt*, **26** (1), 21–30.

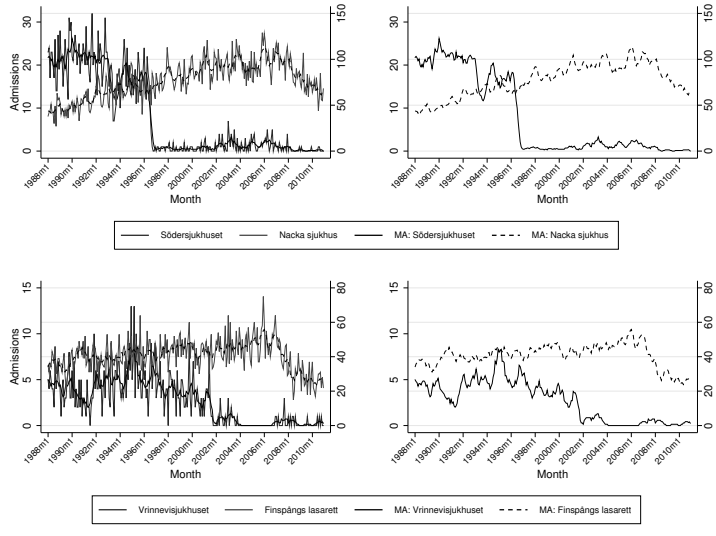
Appendix A Tables and figures

Table A.1. *Descriptive sample statistics*

Variable	Group averages		Group difference	
	No distance change	Distance change	Mean difference	t-value
Heart surgeries	0.137	0.141	0.004	0.279
	(0.537)	(0.525)	(0.012)	
Years since hospital visit	5.841	6.231	0.390	5.422
	(3.001)	(3.104)	(0.072)	
Years since heart surgery	8.549	9.360	0.811	13.862
	(2.450)	(1.615)	(0.058)	
Female	0.313	0.298	-0.015	-1.318
Age	71.12	70.75	-0.37	-1.389
	(1.113)	(1.123)	(0.027)	
Days in hospital	6.728	6.408	-0.320	-1.822
	(6.607)	(5.198)	(0.176)	
Hospital distance in j	14.175	26.085	11.909	36.869
	(13.516)	(12.659)	(0.323)	
Hospital distance in $j-1$	14.175	14.076	-0.100	-0.309
	(13.516)	(11.163)	(0.323)	
Survived AMI	0.773	0.760	-0.013	-1.300
OOH AMI death	0.174	0.194	0.020	2.215
IH AMI death	0.054	0.047	-0.007	-1.309
Observations	329,756	1,759	331,515	-

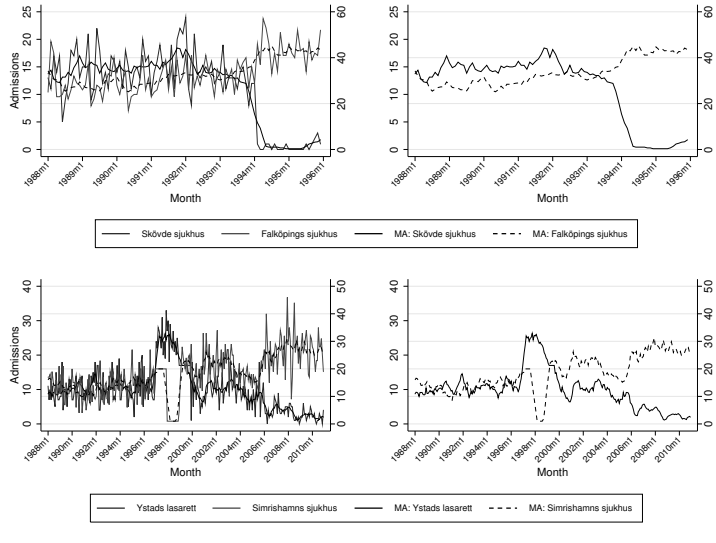
NOTE.—The table reports estimated means, mean differences and (standard deviations) of included covariates for sampled AMI patients who did or did not experience a change in distance from an emergency hospital closure respectively. The variables are; the historical number of heart surgeries, number of years since the last hospital visit, years since the last reported heart surgery, the individual's gender and age, the historical number of days in hospital since 1987, the observed distance from an individual's registered residence to his or her designated home hospital in time period j and $j-1$ where j indicates the year of the hospital closure respectively, and finally the proportion of patients who survived, died outside and inside a hospital respectively. The last two columns report the difference in group means and the result from a standard t-test of equality of means across the groups.

Figure A.1. Visits at closing hospitals and their referral hospitals over time



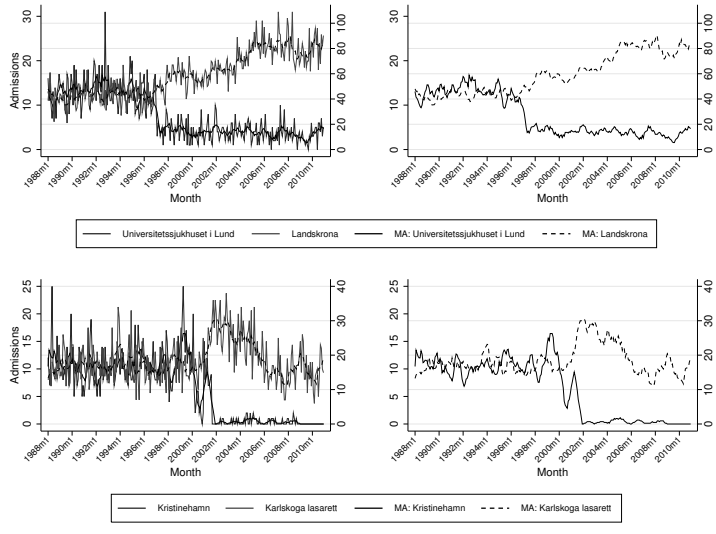
NOTE.— Data source: Swedish National Board for Health and Welfare. The plots on the left show the monthly number of AMI visits at hospitals with closing emergency rooms (blue lines and left *y*-axis) and referral hospitals (red lines and right *y*-axis). The plots on the right show the corresponding six-month moving averages of the same plots (three leads and three lags).

Figure A.2. Visits at closing hospitals and their referral hospitals over time



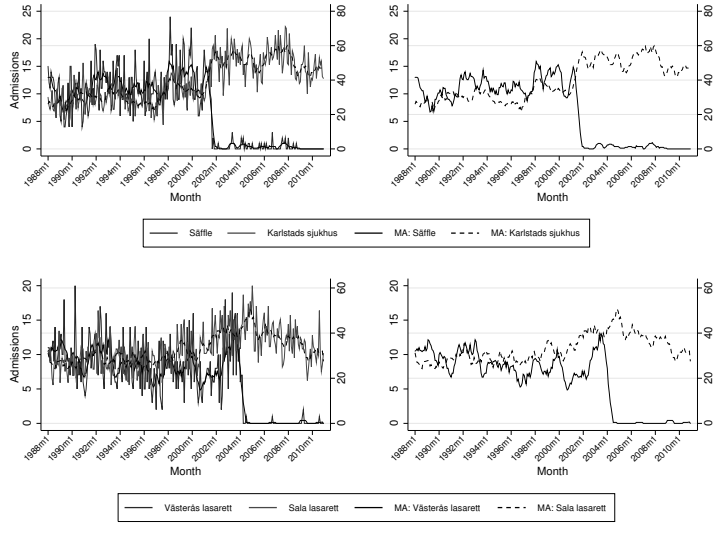
NOTE.— Data source: Swedish National Board for Health and Welfare. The plots on the left show the monthly number of AMI visits at hospitals with closing emergency rooms (blue lines and left *y*-axis) and referral hospitals (red lines and right *y*-axis). The plots on the right show the corresponding six-month moving averages of the same plots (three leads and three lags).

Figure A.3. Visits at closing hospitals and their referral hospitals over time



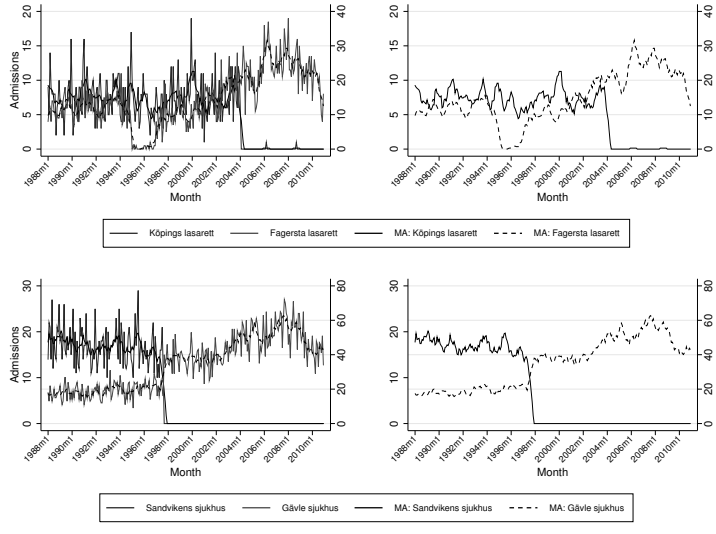
NOTE.— Data source: Swedish National Board for Health and Welfare. The plots on the left show the monthly number of AMI visits at hospitals with closing emergency rooms (blue lines and left *y*-axis) and referral hospitals (red lines and right *y*-axis). The plots on the right show the corresponding six-month moving averages of the same plots (three leads and three lags).

Figure A.4. Visits at closing hospitals and their referral hospitals over time



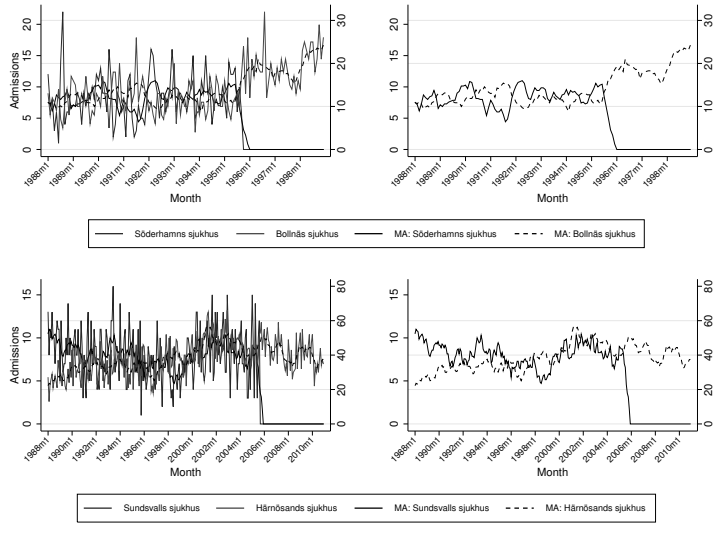
NOTE.— Data source: Swedish National Board for Health and Welfare. The plots on the left show the monthly number of AMI visits at hospitals with closing emergency rooms (blue lines and left *y*-axis) and referral hospitals (red lines and right *y*-axis). The plots on the right show the corresponding six-month moving averages of the same plots (three leads and three lags).

Figure A.5. Visits at closing hospitals and their referral hospitals over time



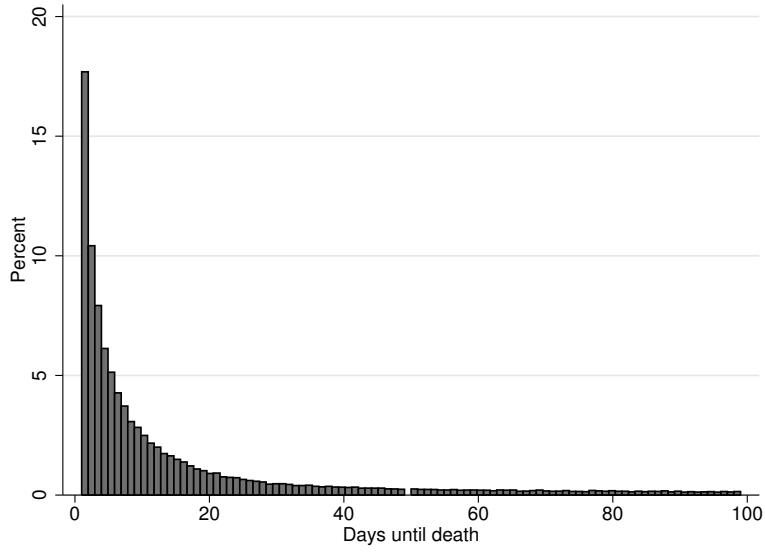
NOTE.— Data source: Swedish National Board for Health and Welfare. The plots on the left show the monthly number of AMI visits at hospitals with closing emergency rooms (blue lines and left y-axis) and referral hospitals (red lines and right y-axis). The plots on the right show the corresponding six-month moving averages of the same plots (three leads and three lags).

Figure A.6. Visits at closing hospitals and their referral hospitals over time



NOTE.— Data source: Swedish National Board for Health and Welfare. The plots on the left show the monthly number of AMI visits at hospitals with closing emergency rooms (blue lines and left y-axis) and referral hospitals (red lines and right y-axis). The plots on the right show the corresponding six-month moving averages of the same plots (three leads and three lags).

Figure A.7. Distribution of deaths by days after an AMI



NOTE.— Data source: Swedish National Board of Health and Welfare. The figure shows the distribution of the observed number of deaths in the analysis sample of AMI patients excluding individuals that die on the same day as the AMI occurred (due to scaling issues). The number of AMI cases ending in death on the same day as the AMI occurred is approximately 191,000 or 58 percent of the total number of deaths.

Essay 3: Learning-by-doing in a high-skill profession when stakes are high: Evidence from advanced cancer surgery*†

1 Introduction

Learning-by-doing is believed to be an important source of productivity growth (Arrow, 1962; Lucas, 1988). The intuition for this relationship is simple — the more individuals or organizations perform a specific task, the better they become at doing it. Evidence for learning-by-doing has been obtained by linking greater accumulated production volumes to reductions in unit labor costs or, more directly, to quality improvements in activities ranging from Kibbutz farming, ship building, car manufacturing, airplane flight control, nuclear plant operation reliability, iron works and even pizza making (Wright, 1936; Lundberg, 1961; Darr *et al.*, 1995; Benkard, 2000; Jovanovic and Nyarko, 1995; Thornton and Thompson, 2001; Thompson, 2001; Levitt *et al.*, 2012).

While it is widely believed that greater production facilitates learning-by-doing effects in almost all sectors of the economy, surprisingly few contributions have been able to pin down a credible causal relationship of this phenomenon. Difficulties of identifying the causal chain from production volume to productivity arise mainly from issues related to empirical investigation of the relationship. The perhaps most obvious inferential problem is that conventional measures of experience and tenure are likely to be correlated with unobserved factors which, in turn, are associated with productivity.[1] Moreover, most existing evidence for learning-by-doing stem from competitive markets

*Co-authored with Petter Lundborg and Johan Vikström.

†The authors thank Jerome Adda, Tabea Bucher-Koenen, Tomas Buchmueller, Christina Felfe, Martin Huber, Per Johansson, Martin Karlsson, Tobias Laun, Nikolaj Malchow-Moller, Paolo Pertile and seminar participants at Uppsala University, MEA-Munich, IFAU-Uppsala, University of St. Gallen, University of Southern Denmark, Lund University, 12th Journées LAGV in Aix-de-Provence, The Economics of Disease workshop in Darmstadt, Maastricht, University of Duisburg-Essen, CBS-Copenhagen, RWI-Essen, CHEGU-Gothenburg, ESPE 2013 and EALE 2013. Financial support from the Swedish Council for Working Life and Social Research (DNR 2004-2005 and 2009-0826) is gratefully acknowledged. Vikström acknowledges support from the Tom Hedelius and Jan Wallander foundation.

[1]In a recent overview of the literature, Thompson (2012) states that: "The most obvious danger in estimating organizational learning curves is that the conventional measure of experience, cumulative output, is correlated with variables known to be associated with higher productivity but that are often not available to the researcher." Moreover, "Unsurprisingly, since the much greater part of the empirical learning curve literature predates the wide use of instrumental vari-

where organizational changes facilitating learning-by-doing mechanisms are likely to be intentionally implemented and exploited by profit-maximizing agents, causing observed relationships of volume and productivity to include effects from competition such as selective sample attrition. However, this may be less of a problem in the public sector in which competition is restricted and where regulations control market entries and exits — implying that efficiency-enhancing organizational changes are more likely to remain unexploited. The health care sector is in many countries an example of such a heavily regulated market and, at the same time, a high-stake sector in which efficiency gains can be counted in lives saved — underlining the importance of identifying the determinants of productivity.

Apart from general difficulties in empirically assessing learning-by-doing mechanisms, the specific case of the health care sector involves additional identification problems; first, differences in outcomes of high- and low-volume hospitals may be explained by *selective referral* where high-quality hospitals have a higher volume simply because they attract more patients — i.e. a reverse causality problem.[2] Second, individuals with different health characteristics may self-select into areas where the size of nearby hospitals varies, such as in rural and urban areas. Third, local health shocks, such as localized outbreaks of contagious diseases, may affect both hospitalization incidence rates and health outcomes in the area, possibly inducing a spurious relationship between cumulated number of treatments and their subsequent degree of success.[3] For policy purposes, while being observationally equivalent, these competing explanations clearly have very different implications than conclusions drawn from a learning-by-doing mechanism.

A large number of studies in the medical literature have found evidence of a positive volume-outcome relationship, accounting for selection by controlling for observable patient characteristics such as age, gender, and reported health conditions.[4] A notable contribution is Birkmeyer *et al.* (2003) who, using a nationwide Medicare sample, found a positive relation between volume and survival in cancer and cardiovascular surgery. In order to also account for

---

able techniques (Angrist and Krueger, 2001), the body of literature offering reliable support for the standard formulation is smaller than is generally supposed." Already Wright (1936) noted that a negative relationship may also arise if more tooling and standardization of procedure means that the firm's ability to use less skilled labor increases.

[2]Another source of selective referral could be that more serious cases are treated at larger hospitals because these hospital have typically more specialized treatments.

[3]More generally, a negative relationship between a firm's accumulated production volume and production costs could also result from a negative cost shock which induces the firm to hire more inputs and increase output.

[4]Halm *et al.* (2002) review 135 medical studies on the volume-outcome relationship for surgical procedures of which about 70 percent find a significant and positive correlation. In a summary article, Kizer (2003) concludes that the strongest relations between volume and outcome have been found for AIDS treatment, surgery for pancreatic cancer, esophageal cancer, abdominal aortic aneurysm, and congenital heart disease.

unobserved heterogeneity across hospitals, some researchers have utilized longitudinal data to estimate fixed-effects models (Hamilton and Hamilton, 1997; Hamilton and Ho, 1998).[5] However, the limitation of these types of models is that they only account for time-invariant unobserved heterogeneity across groups, whereas e.g. selective referral is likely to be driven by trends in hospital quality. To address this empirical problem, a few U.S-based economic studies have used the number of hospitals within a certain radius as an instrument for hospital volume, applying instrumental variables techniques (Gaynor *et al.*, 2005; Gowrisankaran *et al.*, 2006).[6] However, since market structure in more deregulated health care systems is an endogenous outcome of a competitive process, one potential problem is the selective placement of hospitals with regard to underlying patient characteristics of the catchment area and the current and expected future size and quality of nearby competing hospitals.[7]

With the limitations of the existing literature in mind, we provide new evidence of learning-by-doing by studying the causal effect of production volume on quality outcomes in a high-skill activity where stakes are high; advanced cancer surgery. In order to deal with endogeneity we introduce a novel instrument, generating quasi-experimental variation in the number of cancer surgeries performed in Swedish public hospitals over the last two decades. Specifically, the proposed instrument exploits regional variation in closures and openings of hospital cancer clinics over time, affecting the surgical volume at nearby hospitals in ways that we argue are unrelated to individual surgery outcomes. Importantly, since in-patient care in Sweden is organized by the public sector, changes in the number of regional hospitals are unlikely to be primarily driven by aspects of competition, but rather caused by political concerns, unrelated to hospital quality and underlying population health indicators — a conjecture which is also supported by supplementary data analyses.

For the purpose of our analysis, we have collected detailed data on *all* cancer surgeries performed at Swedish hospitals between 1998-2007 — in total more than 100,000 episodes. The data includes rich and detailed information about type of surgery, date of hospital admission and discharge, post-surgery

[5] In neither of these studies were there any evidence of a causal volume-outcome relationship left after accounting for time-invariant heterogeneity in treatment quality, hence supporting a selective referral explanation.

[6] Gaynor (2006) finds positive effects of volume on survival following heart surgery using the number of CABG-offering hospitals operating in a specific radius around the hospital of treatment as an instrument for surgical volume. Gowrisankaran *et al.* (2006) use hospital distance interacted with hospital and patient characteristics as an instrument and find positive effects of volume on survival after heart surgery and repair of abdominal aortic aneurysm.

[7] Other studies have used the total number of hospital beds as an instrument for hospital volume (Luft *et al.*, 1987; Hughes *et al.*, 1988; Farley and Ozminkowski, 1992; Norton *et al.*, 1998). However, in presence of selective referral it is likely that high-quality hospitals should both be larger and have more beds. Finally, Tsai *et al.* (2006) and Kahn *et al.* (2009) use distance to a high-volume hospital for each patient as an instrument for the probability of getting treated at a high-volume hospital.

complications and co-morbidities for each patient. Furthermore, we link the admission data to individual-level register data on mortality and socioeconomic data for the entire Swedish population. From this dataset, we subsequently select and include the three most common types of cancers in Sweden — and in most of the Western world — in our analysis; breast, prostate, and colorectal cancers.

Cancer surgery is particularly well suited for studying learning-by-doing for a number of reasons; first, the intuition for a causal volume-outcome is simple: practice makes perfect. Cancer surgery is in many aspects a complex procedure; surgical removal of tumors often require high individual (and organizational) skills, where even small mistakes might result in serious consequences for the patient's immediate and long-term health. A surgeon who has performed a large number of surgeries may have more finely developed skills and is better able to deal with potential complications during surgery as well as coping with heterogeneity in the patient population. Moreover, hospitals performing large numbers of surgeries may also have better outcomes due to greater accumulated staff experience.[8]

Second, stakes are high in cancer surgery and critical decisions made by surgeons may have lethal consequences for the patient — incentives to learn are therefore strong. Third, analyses of treatment effectiveness in cancer surgery have vast policy relevance. The three cancer types we study in this paper amounts to more than half of the annual total number of cancer surgeries performed at Swedish hospitals and constitute some of the major causes of death in the adult population. Clearly, evidence of learning-by-doing has important policy implications for e.g. health care concentration policies.

Fourth, the unique features of the data implies that we can focus directly on quality improvements in terms of survival following surgery.[9] Moreover, we are able to study follow-up surgeries and re-admission probabilities as indicators of complications due to mistakes made from the initial cancer surgery. These indicators are also useful in the sense that they comprise less dramatic health events, in contrast to cancer mortality. Such data on direct mechanisms has, with few exceptions, been lacking in the previous literature.

Fifth, we are able to study effects from the complexity of the surgery by comparing estimation results from separate cancer types. This allows us to test the hypothesis of whether more learning takes place for more advanced, as compared to more trivial, tasks. As most of the general learning-by-doing literature has focused on learning in the manufacturing industry, which pre-

[8]This type of organizational learning was proposed already by Arrow (1962), who stated that "...it is the very activity of production which gives rise to problems for which favorable responses are selected over time.".

[9]Focusing directly on quality improvements helps to establish learning-by-doing compared to focusing on costs, since any negative relationship between costs and accumulated production can also be explained by the ability to use less and less skilled labor as more tooling and standardization of procedure is introduced (Thompson, 2012).

dominantly consists of manual labor, we can provide more insight into the learning process by estimating separate effects by complexity of the surgery.

Sixth, as cancer surgeries are normally planned months ahead we need not worry about potential confounders, such as emergency access to health care, in our analyses. This would be more problematic for more acute types of health events, e.g. acute myocardial infarctions, where any gains from learning due to centralization of in-patient care may be offset by reduced health care access for patients living further away from a hospital.

Finally, analyzing learning in the Swedish context of publicly provided in-patient health care helps us to rule out effects of mergers on health care quality that are not mediated through learning-by-doing. In a competitive system, consolidation and mergers leading to reduced competition may negatively affect incentives to improve quality in order to attract patients.[10] This effect thus works in opposite direction to any learning-by-doing effects. However, in a publicly provided health care system, we are able to rule out such effects from competition.

Our results indicate rapid gains from increased operation volume. In our preferred IV-specification, the results suggest a death rate elasticity with respect to surgical volume of about 0.22. Relating the size of this effect to the variation in hospital volume across hospitals in Sweden, we find that increasing the volume of a given hospital from the 25th to the 50th percentile of the hospital volume distribution would imply an estimated decrease of the death rate with about 4.1 percentage points, or 18.7 percent. We also find, in line with our expectations, that learning effects increase with the complexity of the procedure; learning effects are greater for prostate and intestine cancers than for breast cancer. Furthermore, we find that higher surgical volume reduces the probability of another, subsequent, cancer surgery. Given that multiple cancer surgeries for the same patient may indicate e.g. subsequent metastases that the surgeon might have missed when performing the initial surgery, we could interpret this result as a surgeon learning effect.

We consider a number of threats to our IV-design. First, even if changes in volume are exogenous, one could argue that the estimated effect of volume on mortality runs not only through learning-by-doing but also through changes in patient characteristics. In particular, the exclusion restriction would not be satisfied if the additional patients who are referred to a non-closing hospital have a different underlying mortality risk than other patients treated at the same hospital. We deal with this potential problem by *i)* including a battery of individual-level health indicators, such as medical history and socioeconomic factors in order to capture heterogeneity in health, and *ii)* by exploiting an institutional feature of the Swedish health care system in which individuals are

[10] Since US health care consumers face little differences in out-of-pocket expenditures across hospitals, it has been suggested that incentives are strong for hospitals to compete for patients on quality dimensions rather than on financial dimensions (Gaynor, 2006).

assigned hospital based on their geographical location and not by choice. We are thus able to keep patient population constant by studying the outcomes of only those who belonged to the hospitals' catchment area before the change in volume, which ensures that the studied population remains the same in terms of both observed and unobserved factors. Results are not affected to any important extent when we impose this restriction.

The exclusion restriction would also fail if the closures we exploit affect patient outcomes in ways other than through volume or patient characteristics. For instance, if an increase in the size of a hospital also implies improved access to specialists and equipment, then improved patient health outcomes may be a consequence of these improvements rather than an effect from surgeon learning. However, if our estimated effects are mainly driven by general improvements in staff and equipment accessibility, we would expect patient outcomes to change also for other types of surgeries in the same hospital, even in absence of an increased surgical volume for these surgery types.[11] We test for such effects of structural changes by exploiting that surgical volume for other surgery types than cancer in remaining hospitals may not change when cancer clinics close at nearby hospitals. In this case, we expect no first-stage on volume for these surgeries from the hospital closures. However, if other changes than volume affect both types of surgeries, we would still expect a reduced form effect on the non-cancer surgeries. Utilizing information on cardiovascular surgery, we find that closures of nearby cancer clinics do not affect the number of ischemic heart surgeries performed in remaining hospitals. Moreover, we cannot reject the hypothesis that the reduced form effect for the heart surgeries are distinguishable from zero, hence lending support for the exclusion restriction.

Even if distance to the nearest hospital should, hypothetically, be less important for cancer surgeries, as they are typically planned ahead, we collect data on exact distances from patient residence to all included hospitals and use this measure as an additional control in our regression model. As expected, changes in the distance to the nearest hospital cannot explain our findings. Finally, we also address the possibility that the estimated effects reflect changes in surgeon characteristics when cancer clinics are closed. Exploiting linked employer-employee data from administrative registers, we show that this is an unlikely explanation for our findings.

To shed further light on the implications of learning-by-doing, we consider non-linearities of the estimated volume-quality relationship with respect to the existence of a learning threshold, as is often found in the general learning-by-doing literature. Specifically, surgeons in low-volume hospitals may gain important experience from each additional surgery, while surgeons in high-

[11] Note, however, that such a pattern could also arise if there are substantial productivity spillovers between clinics at the same hospital and must therefore not necessarily reflect a violation of the exclusion restriction.

volume hospitals might already have had the opportunity to gain experience through a large number of surgeries.[12] In order study non-linear effects in more detail, we complement the IV-analysis with a hospital fixed-effects analysis, relying on within-hospital changes in volume over time. Our estimation results suggest that the volume effect is more prominent at lower levels of surgical volume, whereas practically no volume effect at all is visible at volumes above 140 annual surgeries. This finding hence supports the hypothesis that learning-by-doing is relatively more important at lower levels of surgical volume.

To sum up, we believe that this paper makes several contributions to the volume-outcome literature. First, we contribute to the general literature on learning-by-doing by providing new and compelling evidence of a causal relation between production volume and product quality, as well as on the mechanisms through which this effect is mediated. Furthermore, we contribute with important implications for health care policy, as our results imply that health care consolidation, in the context of a publicly provided health care sector, may lead to important health gains — at least in the context of planned surgery. This is an important finding, not the least with respect to the particular context we study, as a considerable number of hospitals operate below the empirical learning threshold we identify in the data.

The remainder of this article is organized as follows: Section two outlines the basic characteristics of the cancer surgeries we study, focusing on the scope for learning-by-doing. Section three describes the data we use for estimation. Section four includes relevant information on the institutional features of the Swedish health care setting together with a discussion of why this particular context could provide us with plausibly exogenous variation in hospital volume. Section five presents the main results from estimation of the effect of surgical volume on mortality along with a number of supplementary robustness checks. Section six discusses potential mechanisms underlying the estimated results, including a number of empirical analyses in order to assess the validity of these mechanisms. Finally, section seven concludes with a brief discussion of the policy implications of our findings.

2 Cancer surgery and Learning-by-doing

We focus on the three most common cancer types of the Western world, i.e. breast, prostate and colorectal cancers. Prostate (breast) cancer is the most common type of cancer among men (women) in Sweden accounting for 37 (30) percent of cancer incidence among men (women) in 2003. Colorectal can-

[12] One may also argue that increased hospital volume below a certain point could have negative impacts on hospital quality. This would occur if a too high volume leads to less efficient information channels, less efficient decision making, to exhausted medical staff and greater coordination problems.

cer is the second most common type of cancer both among men and women constituting about 11 percent of all cancer diagnoses the same year (Socialstyrelsen, 2007). In general, cancer surgery is a common type of operation in Swedish hospitals which normally requires high-skilled medical staff in order to be performed.

Breast cancer is commonly diagnosed via so-called triple assessment consisting of a physical examination, mammography/ultrasound and a biopsy. The biopsy usually includes performing a so-called sentinel node biopsy (SNB) where the sentinel node —the first breast lymph node to which cancer cells are most likely to spread — is diagnosed for cancer. After diagnosis surgical treatment of the cancer may be performed through either a partial, full or a modified radical mastectomy depending on factors such as the size and spread of the tumor and diagnostic results from the SNB. A partial, or breast-conserving, mastectomy is preferred due to its relatively low invasiveness while a full or radical mastectomy is performed for more complicated and aggressive cancer types. We include all three types of surgery in our analysis.

Surgical treatment of prostate cancer normally involves removing the entire prostate via a radical prostatectomy. We include the two most common types of radical prostatectomy; the retropubic and the endoscopic prostatectomy. The procedures differ by the type of strategy the surgeon uses in order to access the tumor — in the retropubic prostatectomy the surgeon makes a large incision in the abdomen (so-called open surgery) while in the endoscopic prostatectomy several small incisions are made in order to insert surgery instruments into the body (so-called minimally invasive surgery). The latter is also called laparoscopic prostatectomy due to its common use of a laparoscope for visualization and can be done either manually or with the assistance of a robot. We also include two additional palliative surgery procedures in the analysis directed towards slowing the disease progression; Orchidectomy which involves surgically removing one or both testicles in order to reduce the amount of testosterone in the body which increases cancer progression, and transurethal resection of the prostate (TURP) in which a part of the prostate gland is planed away using a resectoscope in order to facilitate urination for end-of-life patients.

Approximately two-third of all colorectal cancers in Sweden are colon-situated while one-third are located in the rectum. We include the three most common colorectal surgeries from each type in the analysis. The specific procedures of each colorectal cancer type is typically chosen depending on where in the intestine the cancer is situated. For rectal cancer, a low anterior resection (LAR) is performed if the cancer is situated in the upper third part of the rectum while an abdominoperineal resection is typically performed if the tumor is located closer to the anus. While the former procedure usually allows for a so-called anastomosis in which the bowels are sewn together again after the surgery, the latter normally result in a colostomy — an opening in the skin of the abdomen to allow body wastes out. A surgical resection of the

rectosigmoid colon, know as the Hartmann's procedure, is sometimes also performed in acute cases when there is no possibility of reconnecting the bowels after surgery. In latter years, the procedure known as total mesorectal excision (TME) has become popular in surgeries relating to the lower two-thirds of the rectum, replacing abdominoperineal resection surgery due to both lower cancer recurrence rates but also the possibility of avoiding a permanent colostomy as the procedure involves a reconstruction of the rectum. For colon cancer the typical surgical procedure is to perform a colectomy in which a part of the colon is removed and then re-attached; specifically, we include the right and left hemicolectomy which refer to the resection of the ascending (right) or the descending (left) colon, respectively, and the sigmoidectomy which involves resection of the sigmoid colon.

Individual and organizational learning-by-doing

Even with general treatment guidelines surgeons have high discretion in deciding which type of surgical procedure to apply and how to apply it for each patient, for example in deciding how much tissue to be removed (or marginal to keep). Clearly, as both cancers and humans are very complex biological entities this involves making important trade-off decisions, not only before but also during surgery, as available information regarding the tumor is incomplete and constantly updating. Hence, there exists obvious scope for individual skill in executing successful treatments, not only in terms of survival but also in other aspects of post-operative patient health such as complications from surgery and tumor recurrences.

From the above discussion it is not far-fetched to argue that individual skill might be related to surgical experience — as cancers and human beings are heterogeneous by nature, a higher number of executed surgeries should increase the probability that a similar case will show up again which in turn may reduce initial information deficiencies regarding the characteristics of the disease.

Similarly, organizational learning-by-doing may arise from learning and experience at the organizational level. Surgeon(s) performing surgery is assisted by a team of nurses, other specialists (e.g. gynecologists, radiologists, plastic surgeons and orthopedists) and need proper equipment and space. If the organizational environment at the clinic or hospital is poor, for example if surgery teams are badly composed, this will also inhibit the surgeon's individual performance. Hence, greater experience may potentially lead to improved cooperation in such teams as the organization learns over time who works well with whom. All in all, we believe that there is clear potential in cancer surgery for both individual and organizational learning-by-doing in isolation of each other as well as possible interaction effects between the two processes on patient outcomes.

3 Data

Our primary data sources consist of the Swedish National Patient Register (NPR), containing population-wide information on all in-patient care in Sweden, and the National Causes of Death register (NCDR), containing information about all deaths for individuals who had a permanent residence in Sweden. The NPR contains individual-level data on date and hospital of admission and discharge, the nature of the admission such as the length of stay and whether it was acute or planned as well as rich medical information including main and co-diagnoses (through the International Classification of Diseases, ICD) and information about any medical procedures made in relation to the hospitalization (through the National Classification of Surgical Procedures NCSP[13]). The NCDR includes information on the date, place and the underlying cause of death. For both registers we have access to data until 2011.

As a major revision of the NCSP in 1997 made comparisons of procedures over time very difficult we chose to sample cancer surgeries from 1998. Furthermore, we set 2007 as our last sampling year in order to have at least a four year follow-up period for each patient to evaluate surgical quality. As mentioned previously, we focus on breast, prostate and colorectal cancers and select the three most common surgical procedures within each category. Specifically, as some of the surgery procedures may also be performed for other diseases than cancer (e.g. TURP for prostatic hyperplasia) we restrict our sample to individuals that have a cancer diagnosis at the time of hospitalization.

We use as our main outcome survival four years after surgery constructed using exact dates of hospitalization and death. We also vary this measure allowing for long- and short-run survival outcomes. Using the characteristics of the NPR we additionally construct a number of other quality-related health outcomes such as re-admission prevalence, multiple cancer surgeries and number of days spent in hospital.

Table 3.1 presents sample statistics of the three included cancer types and the corresponding surgery procedures. Our sample consists of in total 109,761 observations of which 48 percent are breast cancer surgeries, 28 percent are colorectal cancer surgeries, and the remaining 24 percent are prostate cancer surgeries. The table also indicates a slight increase in the number of surgeries and a substantial decrease in four year mortality over time. On average 39 percent of the patients in our sample die within four years after having undergone surgery. The four year mortality rate is highest for colorectal cancers (56 percent) and lowest for breast cancer (28 percent). Finally, breast cancer patients are on average substantially younger than colorectal and prostate cancer patients.

[13] NCSP consists of fifteen main chapters containing operations organized by organ system, four sub-chapters, and a chapter with additional codes. The Nordic Medico-Statistical Committee (NOMESCO) published the first printed edition of the NOMESCO Classification of Surgical Procedures (NCSP) in 1996.

Table 3.1. *Estimation sample statistics*

	Frequency	Fraction of sample	Mean age	4 year death rate
Full sample	109,761	1.00	66.26	0.39
1998	9,707	0.09	66.38	0.58
1999	9,914	0.09	66.73	0.55
2000	10,450	0.10	66.69	0.51
2001	10,800	0.10	66.39	0.47
2002	11,225	0.10	66.63	0.43
2003	11,778	0.11	66.22	0.38
2004	12,119	0.11	65.87	0.33
2005	12,568	0.11	66.02	0.28
2006	12,492	0.11	65.83	0.25
2007	8,708	0.08	65.99	0.21
Breast cancer	52,363	0.48	62.03	0.28
Partial mastectomy	25,953	0.24	59.75	0.18
Full mastectomy	10,022	0.09	66.22	0.37
Radical mastectomy	16,399	0.15	63.08	0.39
Intestine cancer	30,626	0.28	70.90	0.56
Right hemi-colectomy	10,903	0.10	73.15	0.58
Left hemi-colectomy	2,086	0.02	70.28	0.53
Sigmoidectomy	5,199	0.05	70.22	0.55
Rectum resection (LAR/TME)	7,342	0.07	68.29	0.47
Hartmann's operation	1,679	0.02	74.41	0.78
Abdominoperineal resection	3,473	0.03	69.03	0.59
Prostate cancer	26,772	0.24	69.22	0.43
Retropubic prostatectomy	11,525	0.11	62.64	0.09
Endoscopic prostatectomy	2,245	0.02	61.53	0.03
Transurethal resection	10,015	0.09	75.79	0.78
Orchidectomy	2,920	0.03	78.59	0.89

NOTE.— Definition of surgeries are discussed in section three of the paper and classified according to the Nordic Classification of Surgical Procedures, NCSP: Partial mastectomy (HAB40), full mastectomy (HAC20), radical mastectomy (HAC22), right hemicolectomy (JFB30), left hemicolectomy (JFB43), sigmoidectomy (JFB46), rectum resection (LAR/TME) (JGB00), Hartmann's operation (JGB10), abdominoperineal resection (JGB30), retropubic prostatectomy (KEC00), endoscopic prostatectomy (KEC01), transurethal resection (KED22) and orchidectomy (KFC10).

4 Empirical strategy

This section concerns the empirical approach we apply in order to test the surgery learning-by-doing hypothesis. We first discuss the descriptive relationship in our cancer surgery sample and the potential pitfalls in interpreting this relationship as a causal one before moving on to motivate the instrument we utilize to avoid these inferential problems. In the last subsection, we briefly explain how the instrument is implemented in the empirical model we apply to estimate the volume effect.

4.1 Descriptive background

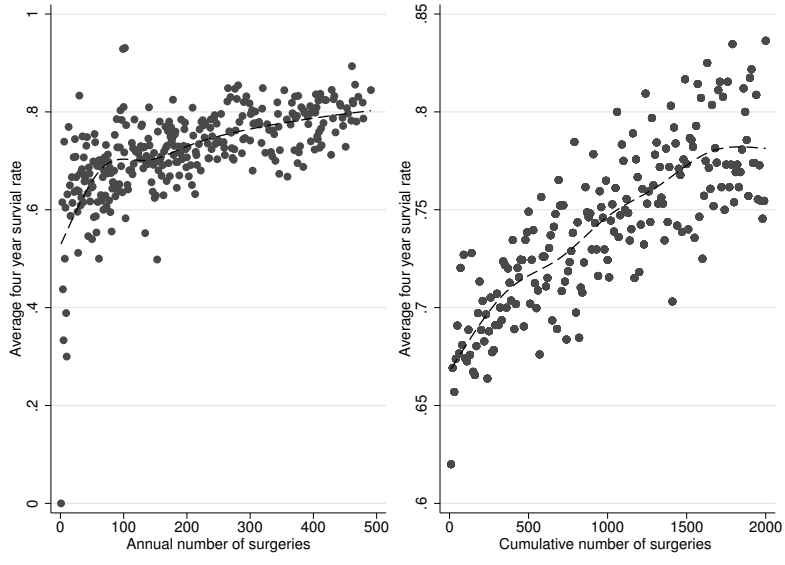
Before presenting our empirical strategy in more detail, we explore the variation we use in our data descriptively. The left panel of Figure 4.1 plots the

raw correlation between the annual number of cancer surgeries and four year survival rates for each included hospital and year. Each dot in the figure corresponds to the average survival rate for a given hospital-year combination. The smoothed average, estimated from a local polynomial regression, indicated by the solid line in the figure, shows that survival rates are higher for hospitals which perform more annual surgeries, which is in accordance with a learning-by-doing hypothesis. Clearly however, the positive volume-outcome correlation visible in the descriptive figure could also be explained by other mechanisms such as patient or staff selection from e.g. selective referral or selection in the underlying case-mix of patients in the catchment areas of small and large hospitals.

The right panel of Figure 4.1 describes our data from a more traditional learning-by-doing outset. Specifically, the figure evaluates the existence of a learning curve by plotting the relationship between *cumulated* surgical volume and survival rates.[14] Each dot in the figure corresponds to the average survival rate as a function of cumulated volume in bins of ten surgeries pooled over all hospitals and years. Interestingly, we see from the figure that average survival probability increases substantially with cumulative hospital volume, again indicative of a learning-by-doing mechanism. However, while interesting, this analysis does not consider the influence of other confounding factors such as time trends from general surgical improvements, serially correlated health or cost shocks affecting both the number of surgeries and survival rates, and changes over time in hospital quality and selective referral.

[14] Accumulated volume is defined as accumulated volume from 1998 and onwards due to the changes in NPR coding that occurred in 1997 (see the data section).

Figure 4.1. The observed relationship between surgical volume and cancer survival



NOTE.— The left panel of the figure relates the hospital annual number of cancer surgeries to the probability of being alive four year after a cancer surgery for all individuals in the analysis sample. The right panel similarly plots the average four year survival rate by the cumulative number of surgeries performed at each hospital since 1998 in the sample — i.e. by order of hospital admission date. The dots represent surgical volume averages in bins of ten and the dashed line represent the estimated relationship using local polynomial smoothing techniques.

Thompson (2012) concludes in his overview of the learning-by-doing literature and the problems of identifying such effects that; "The solution, of course, is to find settings in which cost shocks do not induce changes in input use, or to find instruments for cumulative output". In the remainder of the paper we continue our empirical investigation by applying such an instrument and hence shifting focus from discussing general volume-outcome associations towards making causality claims.

4.2 The instrument

In order to motivate our instrument we first provide relevant background information on the Swedish health care system. In contrast to the U.S., the health care sector in Sweden is highly regulated and the vast majority of hospitals are owned and run by the public sector. In fact, the first and still the only private hospital were started as late as in 1999. Moreover, in-patient health care is organized and mainly financed at the regional level. Sweden is divided into 21 regions, with Stockholm as the largest (with about 2 million inhabitants) and Gotland the smallest (about 60,000 inhabitants). Organizing health care is the

single most important responsibility for the regional authorities. For instance, in 2012 about 82 percent of the regional budgets were used for health care.

The regions in Sweden are run by a political council which is elected every fourth year, held on the same day as the national and municipality elections. Besides following some general laws and guidelines set by the government, e.g. that health care should be provided to all citizens, the regional authorities are more or less free to organize the health care in their region. In effect, since the hospitals are almost exclusively run by the public sector, the members of the regional board thus have a great influence in deciding the specific organizational features of the regional health care — in particular the degree of centralization.

We exploit the fact that the substantial regional autonomy of the Swedish health care system leads to very different responses to a general health care centralization trend taking place in the early 2000s. Specifically, the central government passed a law in 2000 stating that the regional councils were no longer allowed to run budget deficits. This law led many of the regions to discuss and propose centralization measures in order to cut costs and increase efficiency but, due to different outcomes from the political negotiations, these proposals were only implemented — and to different extents — in some of the regions (Lindbom, 2013; Larsson Taghizadeh, 2009). In this paper we utilize closures of cancer surgery clinics generating discontinuities in the number of cancer surgeries at remaining nearby hospitals within the same region across time. Applying these closures ensures that we utilize volume changes derived from political processes rather than surgical volume swings caused by regional health shocks, patient sorting and case-mix selection in hospital catchment areas. Moreover, as we have access to very detailed data stretching over a long period of time, we are able to perform extensive robustness checks to validate that the closures provide us with exogenous variation for our purpose.[15,16]

Using the NPR we define a cancer surgery clinic as closed (opened) if the total number of surgeries is two or fewer (greater) during three consecutive years. With this definition we obtain seven closures and two openings during our analysis period.[17,18] To supplement the data we also validate the closures/openings using information obtained from local newspapers and political protocols.

[15]We are also able to rule out competition effects that are important in some other health care markets (see e.g. Bloom *et al.* (2011)). The reason is that patients in Sweden are normally referred to a designated hospital (usually the closest one) so that competition between hospitals should be scarce.

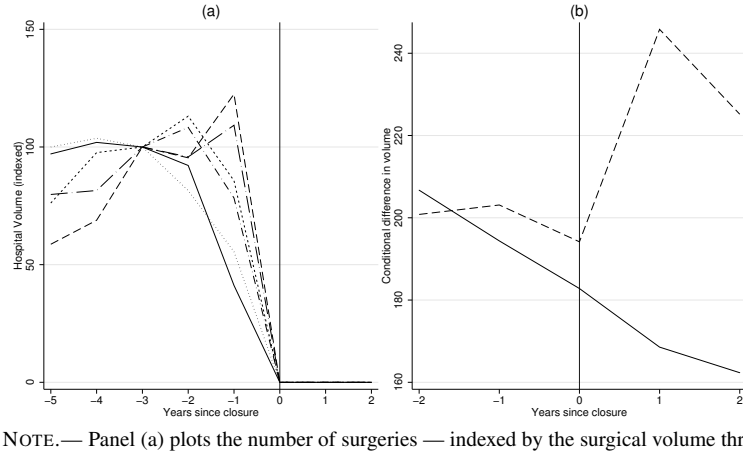
[16]We consider only closures of cancer surgery clinics — and not the entire hospital — ruling out potential merger effects as documented by e.g. Gaynor (2006).

[17]The two openings consists of two formerly closed clinics that were later re-opened.

[18]The closed clinics are Sophiahemmet (2006), Vrinnevisjukhuset (2003), Motala lasarett (2003), Trelleborgs lasarett (2005), Kristinehamns sjukhus (2000), Säffle sjukhus (2000) and Piteå Älvdals sjukhus (2005) and the opened clinics are Vrinnevisjukhuset (2007) and Motala lasarett (2007).

In order to illustrate the variation we exploit for estimation the left panel of Figure 4.2 displays an indexed measure of the number of cancer surgeries performed at the closed clinics before and after the closure. Clearly, there are striking drops in the numbers at the closure years, indicating that the closures occurred quite unexpectedly. Moreover, the right panel of Figure 4.2 illustrates how the closures affected remaining nearby clinics in terms of surgical volume — i.e. the first step of our proposed IV strategy. Specifically, the figure plots the average hospital volume for remaining hospitals in regions with closures (dashed line) and for hospitals in regions without any closures (solid line), respectively, after adjusting for time and hospital fixed effects. For non-closing regions the year of closure represents the year a clinic *potentially* could have closed in the region — i.e. a placebo closure.[19] The closures evidently implied a substantial increase in the number of surgeries at the nearby remaining hospital in the region (on average about twenty percent) in contrast to non-closing regions where, as expected, volume was unaffected.

Figure 4.2. Hospital closures and surgical volume in remaining regional hospitals



NOTE.— Panel (a) plots the number of surgeries — indexed by the surgical volume three years before closure (=100) — for each *closing* cancer clinic by years from the closure. Panel (b) plots the average surgical volume for all *remaining* clinics in each county with a closure (dashed line) and the corresponding average surgical volume for unaffected clinics by years from the closure, adjusted by hospital and calendar time fixed effects. Unaffected regions are used as a *potential* closure for each year a change occurred in an affected region.

Our identification strategy would be invalidated if the political processes leading to the closures and openings were influenced by demographic changes such as urbanization, business cycle effects or underlying trends in public health and/or in the population mix. In order to explore this potential threat to identification, Figure 4.3 presents the level values of certain key variables for

[19] Since the first and last closure in our data occurred in the beginning and the end of the sampling period, respectively, the difference can only be displayed two years before and after the potential closure year.

regions with at least one opening or closure and unaffected regions before and after the change takes place, respectively. The figures reveal that regions affected by a hospital closure or opening on average have larger populations and higher average morbidity than the corresponding unaffected regions. However, no important group differences in unemployment, share of cancer hospitalizations or cancer mortality is distinguishable.

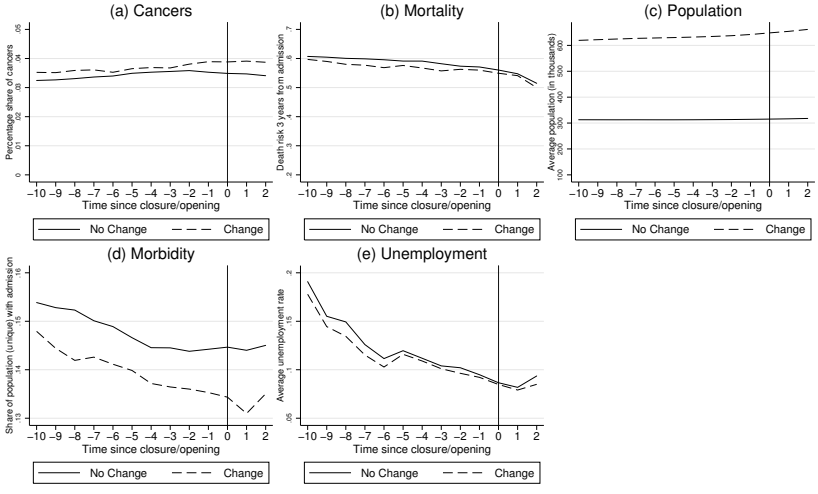
Due to these group differences in observable characteristics we include hospital and calendar time fixed effects and regional linear time trends[20] in the analysis. Including these controls will ensure that the variation provided by our instrument will be exogenous unless the timing of the openings/closures is correlated with underlying non-linear time trends from unobserved factors. As is evident from Figure 4.4, showing a plot of the residuals obtained from estimating linear regressions including this set of controls, the difference between the groups is now indistinguishable for all variables.[21] Hence, this result lends some credibility for our empirical strategy.

The remaining threat of our instrumental variable approach concerns the potential existence of any direct effects of the closures on our outcomes of interest — i.e. whether the closures affected the surgical quality in remaining hospitals through other mechanisms than through changes in hospital volume. Any such effect would invalidate the exclusion restriction of our model setup. To investigate the plausibility of this potential identification problem we collect additional data and perform a series of specification checks. In particular, we test if the closures and openings systematically affected the case-mix in the remaining hospitals, which would be problematic if the additional patients treated at the remaining hospitals had different underlying health characteristics than other patients. For this reason, we run regressions where we exclude the additional patients from our analysis and only focus on the patients that belonged to the hospital catchment area *before* the increase in volume. Importantly, in this case-mix adjusted analysis we exclude individuals treated at closed clinics both before and after the closure. Hence, systematic hospital closures arising from e.g. different health care quality trends across closed and remaining hospitals are by definition not a problem in this analysis since closed units are entirely excluded from the analysis.

[20] As a robustness analysis we have also estimated models with linear hospital trends instead of regional trends, resulting in practically unchanged main estimates.

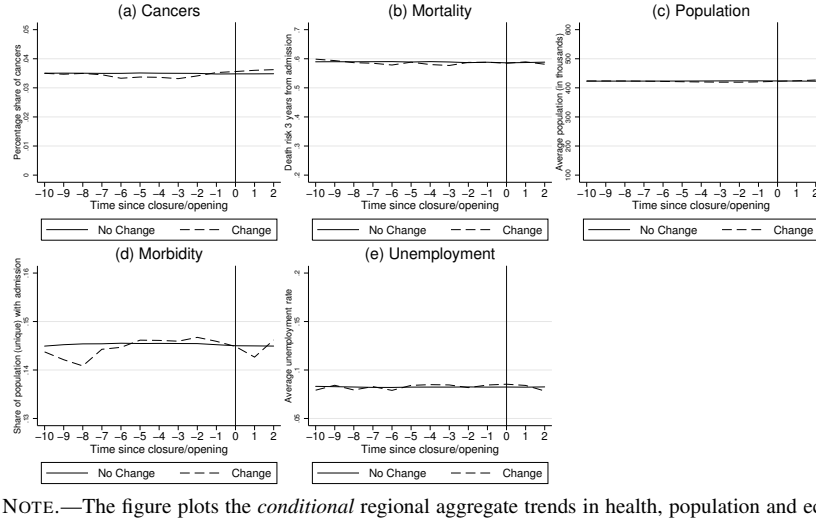
[21] For ease of comparison we have added the variable means to the residuals.

Figure 4.3. Aggregate trends in health and socioeconomic statistics in regions with and without clinic opening/closure



NOTE.—The figure plots the *unconditional* regional aggregate trends in health, population and economic conditions for regions that had an opening or closure (dashed line) and for unaffected regions (solid line) over time since the opening or closure occurred. Unaffected regions are used as a *potential* opening or closure for each year a change occurred in an affected region. The vertical line at zero indicates the year of the closure or opening.

Figure 4.4. Aggregate trends in health and socioeconomic statistics in regions with and without clinic opening/closure



NOTE.—The figure plots the *conditional* regional aggregate trends in health, population and economic conditions (after adjusting for calendar year and hospital fixed-effects and regional linear trends) for regions that had an opening or closure (dashed line) and for unaffected regions (solid line) over time since the opening or closure occurred. Unaffected regions are used for each year a change occurred in an affected region. The vertical line at zero indicates the year of the closure or opening.

Moreover, to assess the exclusion restriction further we; *i)* track all surgeons at hospitals with a cancer clinic closure in order to investigate to which extent the composition of surgeons changed in the remaining hospitals after the closure; *ii)* test for effects from potential organizational changes coinciding with the closures and openings using data on other types of surgeries unaffected by the cancer clinic closures; *iii)* collect data on exact distances from each patient's place of residence to all hospitals in order to examine whether hospital distance affects surgery outcomes; and *iv)* examine whether hospital surgery staff change their procedures (by e.g. changing the type of surgery) as a response to the additional inflow of new patients.

4.3 Econometric model

We specify a linear probability model for the probability of surviving at least four years after cancer surgery. Formally, our baseline model for survival of individual i undergoing surgery s at hospital h in calendar year t is

$$y_{ihst} = \lambda_t + \lambda_h + \lambda_s + \log(volume_{ht})\gamma + X_i\beta_X + \varepsilon_{ihst}. \tag{4.1}$$

Here, $volume_{ht}$ is the number of total cancer surgeries performed at hospital h in calendar year t. The coefficient of interest is γ measuring the effect of an additional surgery on surgical quality, i.e. probability of survival. This baseline model controls for general time trends through calendar time fixed effects λ_t, general differences between different hospitals through hospital fixed effects λ_h, and differences across type of surgery through surgery fixed effects λ_s. A set of individual characteristics X_i, such as gender, age and education level and pre-surgery health factors, such as the number of previous hospitalizations and time since last hospitalization, is also included in the model to account for individual heterogeneity in survival probability. Finally, we also include region-specific linear time trends in the model.[22,23]

We initially follow the traditional medical literature and estimate (4.1) by OLS. We then proceed by estimating FE-IV models applying the first-stage equation

$$\log(volume_{ht}) = \alpha_t + \alpha_h + \alpha_s + \gamma\delta D_{rt}^{closure} + \beta_X X_i + \varepsilon_{ihst}, \tag{4.2}$$

where our instrument, $D_{rt}^{closure}$, is implemented as an indicator variable assuming the value one if a cancer surgery clinic in region r closed in year t or ear-

[22]The health variables are constructed using the data on hospitalizations presented in the data section and the socioeconomic variables are constructed using data from Statistics Sweden on the entire Swedish working age population. For older individuals we use the latest available information.

[23]As a robustness check we also specify surgical volume in levels.

lier.[24] This captures the idea that closures generate a shift in surgical volume at remaining clinics within the region. Initially, we restrict δ to be a scalar and thus have the same effect for all regions. We then utilize the variation generated by the instrument in a more flexible way by allowing the volume shifts from the closures to vary across regions by specifying δ as a vector; δ_r. This, unrestricted, version of the instrument captures the idea that effects on remaining hospitals' volume may differ depending on the size of the closing clinic.

The model in (4.1) is a restricted version of the standard power law formulation of learning-by-doing models.[25] In our setting the power law model implies that current surgical quality, Q_{ht}, at hospital h in time period t is related to cumulative prior surgical volume, E_{ht}, through the power law specification, $Q_{ht} = BE_{ht}^{\beta}$, where β measures the rate of learning and B is a constant. Note that we measure quality (survival) and not production cost so that learning-by-doing implies that β is assumed positive. In this model quality increases with a constant factor each time volume is doubled, so that the quality effect from one additional surgery is lower at higher volumes. Our baseline model is a restricted version of this model where only current volume and not cumulative volume affects current quality. Moreover, our log specification assumes that learning is diminishing with respect to volume, i.e. $\beta < 1$.

In order to investigate the learning dynamics in more detail we replace current volume in (4.1) with cumulated volume in a subsequent analysis. We explore experience in the last year, last two years and full cumulated volume from 1998 and onwards.[26] In these analyses we use lagged versions of our instrument in order to instrument for cumulated volume. The intuition is that the volume shifts induced from the closures accumulates over time as the surgeons gain additional experience for each year since the closure occurred. However, this elaborate analysis stretches the limit of what can be estimated using the variation created by closures and openings, which is also why we mainly focus on the effects of current volume in our result section.

5 Main results

This section presents the main results from analyzing the relationship between hospital volume and post-surgical health outcomes for our sample of cancer patients. Specifically, after an initial analysis of the volume-outcome relationship in the data we investigate the mechanisms underlying the results with

[24] That is, $D_{rt}^{closure} = \mathbf{1}[t \geq t_{closure}|r]$ where $t_{closure}$ is the year of closure in region r. Similarly, if a clinic is re-opened the closure dummy takes the value zero in the re-opening year and all subsequent periods.

[25] See e.g. Thompson (2012) for a discussion of various organizational learning models.

[26] As the code classification changed in 1997 we cannot track volume back further than this year.

regard to the dynamics of the estimated effect, the presence of scale effects and morbidity outcomes. Finally, we discuss the sensitivity and interpretation of the estimated results from a number of supplementary robustness checks.

5.1 The volume-outcome relationship

We begin this section by presenting results on the raw volume-quality relationship, estimated from a simple bivariate regression of hospital volume on four year survival. As seen from column (1) of Table 5.1, there is a strong and significant volume-quality association in our data. In particular, the volume estimate implies that a hospital performing twice as many cancer surgeries as another hospital has on average about 5.7 percentage points higher four-year survival rates.

Due to inferential problems arising from potential reversed causality, time trends and other selection effects we instead turn to our proposed IV analysis in order to empirically evaluate the learning-by-doing hypothesis. Column 2 of Table 5.1 presents the first-stage estimates for our instrument when the closure effect is restricted to be homogeneous across regions. The estimated coefficient on the closure indicator from equation (4.2) is statistically significant and positive indicating that the closure of a cancer surgery clinic increases the average number of surgeries at remaining nearby clinics in the region. The standard errors are clustered at hospital level since we estimate individual level effects using variation at the hospital level.[27] F-statistics for instrument relevance are reported at the bottom of the table. As can be seen, for the restricted IV specification the F-statistic is below ten and hence does not satisfy the usual criterion for a strong instrument.

Next, consider the baseline IV estimate using the restricted IV model presented in column (3) of Table 5.1. The estimated coefficient suggests a sizable, however not statistically significant, effect of surgical volume on the four-year survival rate. To improve on inference by increasing precision we switch to the unrestricted version of our instrument in column (4). The more flexible use of the closure instrument increases the first stage F-statistic by a factor of six resulting in a reduction of the standard error of the volume effect point estimate by more than 50 percent. The effect is now highly significant and close to the initial OLS estimates from column (1) of the table. Table 5.1 also presents estimates from additional IV specifications using the unrestricted version of our instrument. As patients treated at different hospitals may be heterogeneous with respect to health we add a set of individual-level covariates to the model in column (5). The inclusion of these factors leaves the IV-estimates practically unchanged, suggesting that patients at remaining and closed hospitals

[27] We have also estimated standard errors clustered at regional level and multi-level standard errors clustered at both hospital and regional level. This exercise shows that clustering at hospital level renders the most conservative inference.

have similar observed characteristics. The latter is also confirmed from sample statistics for patients at remaining and closed hospitals one year before the closures presented in Table A.1.[28]

Table 5.1. *OLS and IV estimates of the effect of surgical volume on four-year post-surgery survival probability*

	(1) OLS	(2) FS	(3) Restricted IV	(4) Main IV	(5) With controls	(6) Case-mix adjusted
log(volume)	0.057***		0.070	0.055***	0.055***	0.048***
	(0.005)		(0.043)	(0.018)	(0.017)	(0.018)
Closure		0.167**				
		(0.077)				
Socioeconomic characteristics						
Age					0.018***	0.018***
					(0.001)	(0.001)
Age squared					-0.000***	-0.000***
					(0.000)	(0.000)
Female					0.059***	0.060***
					(0.005)	(0.005)
Immigrant					0.002	0.003
					(0.003)	(0.004)
Unknown origin					-0.012	-0.009
					(0.009)	(0.010)
High school					0.013***	0.013***
					(0.003)	(0.003)
College					0.023***	0.023***
					(0.003)	(0.003)
Unknown educ.					-0.016*	-0.016*
					(0.009)	(0.009)
Previous admissions						
One					0.005	0.006*
					(0.004)	(0.004)
Two					-0.003	-0.001
					(0.005)	(0.005)
Three					-0.014***	-0.013***
					(0.005)	(0.005)
Four					-0.006	-0.004
					(0.007)	(0.007)
More than four					-0.066***	-0.065***
					(0.005)	(0.005)
Time since last admission						
One to three years					0.039***	0.040***
					(0.004)	(0.004)
Four to six years					0.060***	0.061***
					(0.004)	(0.004)
Seven to nine years					0.068***	0.067***
					(0.005)	(0.004)
More than nine years					0.060***	0.062***
					(0.018)	(0.018)
Time FE	✓	✓	✓	✓	✓	✓
Hospital FE		✓	✓	✓	✓	✓
Linear trends		✓	✓	✓	✓	✓
First-stage F stat.	-	4.7	4.7	29.3	29.3	199.1
Mean survival	0.78	0.78	0.78	0.78	0.78	0.78
Observations	109,761	109,761	109,761	109,761	109,761	105,028

NOTE.— The table reports point estimates and (standard errors) from estimation of various regression models specified in section four of the paper. Outcome variable in all specifications is four-year survival after cancer surgery. Surgical volume is measured in hundreds. Fixed effects in the last two columns also include surgery fixed effects. In columns (2) and (3) the closure instrument is specified as a dummy variable and in columns (3)-(6) as a set of dummy variables. The last column reports estimates from using the case-mix restricted sample excluding everyone with a designated hospital which at some point is either opened or closed. First stage F-statistics, due to Kleibergen and Paap (2006), are reported for columns (2)-(6). Reference categories are male, native, less than high school, no previous admissions and admission the same year. Standard errors are clustered at the hospital level. *, ** and *** denote statistical significance at the 10, 5 and 1 percent levels.

[28]There are some differences with respect to the gender and immigrant mix. Closed hospitals on average have a larger share of female patients and a smaller share of immigrants. These composition differences do not seem to affect the point estimates, however.

In the last column of Table 5.1 we also condition on unobservable changes in the case-mix of cancer patients by re-estimating the model on a restricted sample including only individuals belonging to a remaining hospital's catchment area prior to the closure. The intuition is the following; while we include these additional patients when predicting hospital volume in the first stage equation (our surgical volume measure includes all surgeries) we subsequently exclude them in the structural equation. However, even if excluded, the surgeons still perform surgery on these patients so that any learning effects will be picked up on the outcomes of the included patients in any case.[29] Excluding patients from closing hospitals have little effect on our volume effect and from this we conclude that our main effects are robust to unobservable changes in the patient case-mix.[30]

Is our estimated volume effect also economically significant? The point estimate from our preferred model in the last column of Table 5.1 implies that doubling the per annum number of surgeries at a hospital will on average increase the survival rate by 4.8 percentage points. The average four year mortality rate in our sample is 0.22 so that the death rate elasticity with respect to surgical volume is about 0.218 — which we regard as a sizable effect. Moreover, relating the estimated effect to the sample variation in hospital surgical volume we find that increasing the volume of a hospital at the 25th percentile to the 50th percentile of the volume distribution (i.e. from 70 to 130 surgeries p.a.) would decrease the four year death rate by about 4.1 percentage points, or 18.7 percent.

Our preferred IV estimate is, while not significantly so, slightly smaller than the OLS estimate. When interpreting this difference it is important to keep in mind that some factors may bias the OLS estimate upwards and others may bias the estimate downwards. For example, a selective referral story in which high-quality hospitals have a higher volume simply because they attract more patients would imply the former bias, while selective referral in which more severe cases directed to large hospitals will imply the latter bias. Moreover, sorting of patients into areas with varying hospital sizes will also bias the OLS estimate but in an *a priori* unknown direction. Finally, an additional possibility is that the IV estimate captures a LATE effect, i.e. the effect on a subset of the hospital volume distribution.

5.2 Effect dynamics and learning

In the former section we studied the direct volume-outcome effect in terms of the effect of current volume on current quality. This analysis confirmed that

[29] See section three for a discussion on how patients are assigned hospitals. Descriptive sample statistics show that 86 percent of the cancer surgery patients are treated at their designated hospital.

[30] We have also re-estimated the case-mix adjusted model using volume in levels with essentially the same results.

recent experience has important impact on surgical quality in terms of post-surgery survival prospects. We next study the effect dynamics in more detail within our empirical framework by investigating whether learning-by-doing persists over time, i.e. whether cumulative experience matters. To this end we re-estimate our preferred specification from the last column of Table 5.1 and instrument cumulated volume over one, two and three years using current and lagged closure status during the last one, two and three years as instruments.

Table 5.2. *IV estimates of the effect of cumulated surgical volume on four-year post-surgery survival probability*

Panel A: Cumulated volume			
	(1)	(2)	(3)
log(current volume)	0.048***		
	(0.018)		
log(cum. vol. last 2 years)		0.083*	
		(0.046)	
log(cum. vol. last 3 years)			0.082
			(0.068)
F statistics First-stage			
F-stat. Level	199.1	23.8	25.0
F-stat. Lag 1		14.5	34.2
F-stat. Lag 2			3.4
F-stat. All	199.1	45.9	117.3
Observations	105,028	95,802	86,272
Panel B: Cumulated and current volume			
	(1)	(2)	(3)
log(current volume)	0.048***	0.050**	0.032*
	(0.018)	(0.021)	(0.017)
log(cum. vol. last 1 year)		0.021	
		(0.059)	
log(cum. vol. last 2 years)			0.050
			(0.048)
F statistics First-stage			
Log volume	199.1	183.7	220.4
Cumulated variable(s)		18.7	127.6
Observations	105,028	95,802	86,272

NOTE.— The table reports point estimates and (standard errors) from estimation of various regression models with cumulative hospital surgical volume estimated using instrumental variable techniques discussed in section four of the paper. Outcome variable in all specifications is four-year survival after cancer surgery. In column (1) the volume instrument is specified as a set of dummy variables indicating a closure in the region the same year while in columns (2) and (3) instruments for cumulative volume are both current and two (three) years lagged clinic closures. All models include calendar time, hospital and surgery fixed-effect as well as regional linear trends and are estimated using the case-mix adjusted sample specified in the last column of Table 5.1. First-stage F-statistics are reported for all instruments both separately and jointly. Panel A reports estimates when including current volume in the cumulative volume measure while panel B reports results when current and one and two-years cumulative volume are included as separate regressors. Standard errors are clustered at the hospital level. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

The results from this exercise are given in panel A of Table 5.2. First-stage F-statistics are reported for both the joint inclusion of all instruments and for

each instrument separately. While the joint inclusion of the instruments are always highly significant the three-year lagged instrument is rather weak, resulting in more imprecisely estimated parameters. Nevertheless, the estimated volume effects show an interesting pattern with higher quality effects of cumulated, relative to current, volume. This result hence suggests that cumulative experience may have an additional impact on current surgical quality above and beyond the immediate experience acquired by the surgeon.

To further investigate the relationship between current and cumulative experience and surgical quality, we include both current volume and cumulated volume as two separate endogenous variables in panel B of Table 5.2. While the standard errors are high, we see an interesting pattern from the point estimates reported in the table; as experience is accumulated over several years the estimated proportion of the effect on current experience diminishes in relation to the cumulated experience. This effect pattern is expected in a situation where learning from experience is, at least partially, kept and not forgotten over time, see e.g. Benkard (2000).

5.3 Scale effects

Besides learning, the volume effects we estimate may also arise from scale effects, in particular from the utilization of more advanced equipment. Specifically, if larger surgical volume implies that clinics can take advantage of scale effects — such as that any fixed costs will be distributed over a greater number of surgeries — and cut unit costs of surgery they may invest elsewhere, such as in surgery instruments which potentially could improve surgery outcomes. While, theoretically, such streamlining of the production chain should be more important for more standardized "products" than the advanced treatment of cancer patients, we investigate the scale effect hypothesis by using complementary information on robot assisted surgery.

Surgery robots are a very expensive and advanced type of equipment which could potentially improve surgery outcomes. We conjecture that, if scale effects existed, one indication for the existence of these effects would be the observation of an increased usage of surgery robots in remaining hospital cancer clinics after a clinic closure in a closure region. In our sample, robots are only used in the treatment of prostate cancers and so we estimate the probability that a cancer surgery if performed with the assistance of a surgery robot both for the aggregated sample and separately for the relevant surgery types using our preferred IV specification.[31] The results are reported in Table 5.3 and does not indicate a statistically significant increase in the probability of using a robot after a regional clinic closure.

[31] On average about five percent of all prostate cancer surgeries are performed with the assistance of a robot in our sample.

Table 5.3. *IV estimates for the choice to perform robot assisted surgery*

	All surgeries	Prostate cancer
	(1)	(2)
log(volume)	0.015	0.041
	(0.013)	(0.044)
Observations	105,028	25,693

NOTE.— Outcome is an indicator for the choice to perform robot assisted surgery. All models include calendar time fixed effect, hospital fixed effects, linear regional trends and surgery fixed effect. Case-mix restricted samples excluding everyone with home hospital that is either opened or closed. Standard errors are clustered at the hospital level. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

Another way to empirically separate learning-by-doing and scale effects would be to study the effect dynamics over time in the remaining clinics of a region in which a closure occurred. Hypothetically, volume effects from learning-by-doing should emerge gradually as surgeons accumulate experience while scale effects should be static and only shift quality. While the evidence obtained from estimating cumulative volume effects gave us some initial support for the learning-by-doing hypothesis we now take the analysis one step further by estimating the reduced form of our restricted model allowing for a time-varying effect of the closures in order to capture their impacts one and two years after the closure occurred. If the effect persists several years after the closure occurred we conjecture that it cannot be driven solely by scale effects. Table 5.4 presents the results from estimation; the first column report the baseline reduced form point estimate for all years while column (2) separates this effect into the first and the second year after closure. Clearly, as the point estimate is much higher for the second year after the closure we conclude that it is unlikely that scale effects could explain all of the volume effect we identify in the data.

Table 5.4. *Time-varying reduced form estimates following regional closures of cancer surgery clinics*

	(1)	(2)
Closed clinic in region	0.016**	
	(0.007)	
First year after closure		0.013*
		(0.007)
Second year after closure		0.033**
		(0.014)
Observations	107,743	107,743

NOTE.— The table reports point estimates and (standard errors) from estimation of the reduced form regression from the empirical model discussed in section four using a restricted sample including only observations for a maximum of two years after a regional cancer clinic closure. Outcome is four-year survival after cancer surgery. All models include calendar time, hospital and surgery fixed-effect as well as regional linear trends. Column (1) report the reduced form estimate on post-closure survival rates of a regional cancer clinic closure while column (2) separates this estimate into the first and second year after the closures occurred, respectively. Standard errors are clustered at the hospital level. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

5.4 Complications

While an increased survival probability may perhaps be the most important outcome for cancer patients, other important post-surgery outcomes such as general life quality and subsequent complications are also meaningful quality measures which could be influenced by learning-by-doing. Moreover, since the level of invasiveness of the surgical procedure performed may be inversely related to survival probability, survival prospects and post-surgery life quality may sometimes be traded off against each other. In order to study complications we use data from the hospitalizations records and construct detailed information on re-admissions and subsequent cancer surgeries. If surgeons can lower re-admission and re-operation incidence rates with additional experience, this clearly has an impact on post-surgical health as well as an informal test of the trade-off hypothesis.

Specifically, we estimate our preferred IV model using the probability of a re-admission within one year and a new cancer surgery within three years from the initial surgery as outcomes.[32] The first two columns of Table 5.5 present the results from this exercise. The estimate from column (1) suggests a statistically significant reduced probability of a subsequent cancer surgery within three years from the initial surgery. In particular, if hospital surgical volume is doubled then, on average, the probability of another surgery decreases with about 0.034 percentage points. With a mean sample probability of follow-up cancer surgery of 0.08, this translates into a sizable effect.[33] Finally, the re-

[32] We have also used other follow-up time periods for the re-admission and re-operation measures which gives similar results.

[33] These estimates suffer from a competing risk problem since deceased patients cannot undergo additional surgery. However, since we find that higher volumes increases the survival probability then, if anything, our estimates should be biased towards zero.

admission point estimate also has a negative sign, however indistinguishable from zero at standard levels of significance, which also point in the direction that learning-by-doing not only affects the survival dimension of health care quality but also other aspects of post-surgical life quality.

Table 5.5. *IV estimates of the effect of surgical volume on additional outcomes*

	(1) New cancer surgery	(2) Re-admission	(3) Number of surgeons	(4) Time in hospital
log(volume)	-0.034**	-0.041	2.204	0.344
	(0.017)	(0.027)	(5.289)	(0.401)
Mean of outcome	0.08	0.42	82.87	5.91
Observations	105,028	105,028	80,633	105,028

NOTE.— The table reports point estimates and (standard errors) from estimation of various regression models of hospital surgical volume, estimated using instrumental variable techniques discussed in section four of the paper, on different outcomes. New cancer surgery (re-admission) is defined as an indicator for the event of subsequent cancer surgery (hospitalization) within three (one) year(s) after the initial cancer surgery. Number of surgeons is measured as the total number of surgeons in the hospital in which the patient was treated for cancer and time in hospital is measured as the length in days of the admission spell at the time of cancer surgery. All models include calendar time, hospital and surgery fixed-effect as well as regional linear trends and are estimated using the case-mix adjusted sample specified in the last column of Table 5.1. Standard errors are clustered at the hospital level. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

5.5 Robustness checks

In previous sections we have provided empirical evidence, using closures and openings of cancer surgery clinics as a source of exogenous variation for hospital surgical volume, that the number of surgeries performed at a hospital has a sizable impact on the four-year survival of patients who underwent cancer surgery. Moreover, we also showed that these results are robust to observable and unobservable changes in patient case-mix following closures and openings in the same region. In this section we corroborate further potential concerns against our interpretation of the empirical results as learning-by-doing effects by performing a sequence of additional robustness checks.

Staff composition

Closures of cancer surgery clinics release additional surgeons into the labor market. If a large enough share of these surgeons move to hospitals within the same region the exclusion restriction might be violated because of staff compositional changes. To investigate this concern we gather information on surgeons employed in closing hospitals in order to check to which extent these surgeons migrated to nearby hospitals in the region. To this end we use the Swedish employment register, providing annual information on all hospital employees. Together with information on attained education levels from the Swedish population registry this allows us to identify surgeons at all hospitals

and to follow them over time as they switch employment. Moreover, the data also provides background characteristics of each individual.[34]

We focus on all surgeons who were employed one year before the closures, distinguishing between surgeons who stayed at the same hospital and surgeons who transferred to hospitals within and outside of the closure region at the year of closure, respectively. Transfer statistics of the surgeon groups are presented in the first row of panel A of Table 5.6 in which we see that 69 percent of all surgeons at closing clinics remain at the same hospital after the closure while 18 and 13 percent of the surgeons transfer within and outside of the region, respectively. In total, however, the transferring surgeons only constitute 2.4 percent of the total number of surgeons in the remaining hospitals, implying that the surgeon mix is practically unchanged after the closures.[35] Even so, we compare the characteristics of the surgeons who transferred to nearby hospitals to the pre-existing composition in these hospitals on basis of observed factors in the remaining rows of Table 5.6. From the reported labor earnings in panel A of the table we see that incomes of surgeons transferring within the closure region are higher than the average labor earnings among staying surgeons, implying that more experienced surgeons leave the closing hospital. However, since average earnings of these surgeons are significantly lower than average earnings of surgeons at the remaining hospitals, we conclude that reallocation of highly qualified surgeons cannot explain the observed patterns in surgical quality.

[34]The population registers contain detailed information on type of education. We identify surgeons as individuals with completed medical education specializing in surgery or other surgery related specialties like anesthesia and emergency care. Unfortunately we are not able to uniquely identify cancer surgeons, e.g. oncologists.

[35]In Panel B we examine the transfer of surgeons in three years before the closures. In these years only six percent of all surgeons at hospitals closing down three years later transferred within the region. Hence, this result strengthens the conclusion that there was no excess transfer of surgeons following the cancer clinic closures.

Table 5.6. *Statistics for surgeons at remaining and closing cancer clinics who remain or transfer within or outside closure regions*

Panel A: Statistics one year after closure

	Surgeons at closed clinics one year before closure			
	Stay	Transfer within region	Transfer outside region	Remaining hospitals
% of surgeons at "closed"	0.69	0.18	0.13	
Age	45.8	46.0	46.5	45.4
Labor earnings	606,278	611,720	581,665	634,123
Females %	0.49	0.20	0.30	0.36
Married %	0.63	0.64	0.48	0.74

Panel B: Statistics two years before closure

	Surgeons at closed clinics three years before closure			
	Stay	Transfer within region	Transfer outside region	Remaining hospitals
% of surgeons at "closed"	0.88	0.06	0.06	
Age	45.8	45.0	47.2	43.6
Labor earnings	536,698	520,714	534,520	586,588
Females %	0.37	0.57	0.27	0.36
Married %	0.62	0.71	0.60	0.66

NOTE.— The table samples mean characteristics for surgeons who were employed at hospitals where the cancer clinic remained and surgeons employed at hospitals where the cancer clinic closed who stayed, transferred within or outside the region, respectively. Panel A and panel B report sample means for surgeons one and three year before the closures, respectively. Labor earnings are measured on an annual basis in SEK (1 SEK is approximately equal to €0.10).

Organizational changes

Organizational changes coinciding with the cancer clinic closures may confound the learning effect if these changes affected the health outcomes of treated cancer patients. We evaluate this potential identification problem by using data on other types of surgery which plausibly were not affected by the cancer clinic closures. Specifically, we utilize in-patient data on the most frequently performed heart surgeries arguing that outcomes from these surgeries should be affected by any general organizational changes, but that the mechanics of this effect would not go through any volume changes if only cancer surgery volume was affected by the closures. Thus, we effectively force the reduced form effect to only include direct effects from the closures and not the indirect volume effects from the first stage. Re-estimating the reduced form model using our closure instrument on heart surgery will thus offer an informal test of whether any important organizational changes at the remaining hospitals coincided with the shift in surgical volume.

The last two columns in panel A of Table 5.7 present the first-step and reduced form estimates from estimating our model using the heart surgery sample while the first two columns report the baseline cancer surgery results for comparison. The first step using cancer clinic closures as an instrument

for heart surgery volume is, as expected, insignificant, indicating that the instrument is irrelevant for this particular sample. Moreover, the reduced form estimate is also insignificant and close to zero, implying that there is also no direct effect of the closures on the probability of surviving heart surgery. However, both the first step and the reduced form estimates are highly significant for the cancer sample. Taken together, we interpret the results from this exercise as that any important hospital organizational changes coinciding with the clinic closures — and thus violating our exclusion restriction — seem to be unlikely.

Distance to the hospital

Even if the impact on health outcomes from a patient's proximity to a hospital should not be important for cancer surgeries, which are normally planned months ahead, we nevertheless investigate to what extent distance plays a role for our estimated results. To this end we merge detailed geographical information from Statistics Sweden on the place of residence down to the level of single building blocks for each patient in our sample. Together with constructed geographical data for all hospitals this allows us to calculate the exact distance to each hospital. As we only have population data for the working-age population, i.e. for individuals younger than 65 years, we proxy hospital distance for older patients using the average distance to the hospital for younger patients living in the same municipality.

In Panel B of Table 5.7 we present the IV estimates including hospital distance (in kilometers) as a control variable in the model. We use two different measures of distance; distance to the treating hospital (reported in column (2) of the table) and distance to the nearest hospital (reported in column (3)). For comparison, column (1) reports the baseline estimate from Table 5.1 without controlling for hospital distance. As the point estimates for the volume effect are similar in all three specifications we conclude that hospital distance does not impact the interpretation of our results.

Table 5.7. *Robustness analysis: organizational changes, distance to hospital, access to surgery and type of surgery*

Panel A: First-step and reduced form estimates for cancer and heart surgeries

	Cancer		Heart	
	FS	RF	FS	RF
	(1)	(2)	(3)	(4)
log(volume)	0.167**	0.012*	0.038	-0.002
	(0.059)	(0.006)	(0.057)	(0.003)
Observations	109,760	109,760	139,976	139,976

Panel B: IV estimates adjusting for changes in distance to hospital

	Main	Distance to treating hospital	Distance to closest hospital
	(1)	(2)	(3)
log(volume)	0.048***	0.049***	0.050***
	(0.018)	(0.018)	(0.019)
Distance measure	No	Yes	Yes
Observations	105,028	105,028	105,028

Panel C: IV estimates on timing of surgeries (age of patients)

	Age at surgery
	(1)
log(volume)	0.061
	(0.068)
Observations	105,028

Panel D: IV estimates for choice between less and more invasive procedure

	Partial mastectomy	Laparoscopic prostatectomy
	(1)	(2)
log(volume)	-0.013	0.011
	(0.039)	(0.025)
Observations	50,096	25,693

NOTE.— The table reports point estimates and (standard errors) from estimation of various regression models to assess the robustness of the main results from Table 5.1. Panel A report the first-stage and reduced-form estimates of surgical volume from the cancer and heart surgery samples in which the cancer clinic closure instrument is used for both samples. Outcome is four-year survival after cancer surgery and all models include calendar time, hospital and surgery fixed-effect as well as regional linear trends. Column (1) of Panel B reproduces the results from our preferred specification in column (6) of Table 5.1 while columns (2) and (3) additionally include as control variable the distance to *i)* the *treating* hospital and *ii)* the *closest* hospital. Panel C reports the result from our preferred specification in column (6) of Table 5.1 using patient age at surgery as outcome. Finally, panel D also reports estimates from our preferred IV specification using the subsample of breast (prostate) cancer surgery in column (1) ((2)) and an indicator variable for whether the performed surgery was coded as a partial mastectomy (laparoscopic prostatectomy) as outcome variable. For all specifications in the table, standard errors are clustered at the hospital level and *, ** and *** denote significance at the 10, 5 and 1 percent levels.

Access to surgery and choice of surgery

Finally, if the inflow of additional patients to a remaining surgery clinic somehow qualitatively affects the organizational structure of how cancer surgery is performed in the hospital, our estimated coefficients might simply capture the effect of organizational responses from an increased surgeon workload. We

investigate two such mechanisms; selection in the pool of patients receiving surgery (i.e. if a surgery queue arises) and changes in the type of surgical procedures performed (e.g. from more advanced and time-consuming procedures to simpler and more rapid ones). To evaluate potential patient selection we examine whether the average age of patients is affected by the additional patient inflow. Panel C of Table 5.7 reports the results from re-estimating our preferred IV-model using patient age at surgery as outcome. The point estimate is insignificant and close to zero and hence does not indicate that patients are treated at a later (or earlier) stage due to increased workload.

To evaluate whether surgeons change their choice of surgical procedure as a consequence of the patient inflow we use information on the invasiveness of the surgery; partial mastectomies are less invasive but require more time than full or radical mastectomies. Prostatectomies can either be done using the more invasive open, retropubic, surgery or via the more time-consuming, but minimally invasive, laparoscopic procedure. We estimate our IV model separately for each type of cancer using as outcome an indicator variable for whether the surgery was performed using the less invasive procedure. The results presented in panel D of Table 5.7 show no significant effects of the closures on the choice of procedure. In fact, the point estimates for the different cancer types even have different signs. Hence, we conclude that neither patient selection nor selection in surgery procedure can explain the volume-quality pattern we document in this paper.

6 Extensions and mechanisms

Thus far we have documented that larger surgical volume lead to improved surgical quality outcomes from both increased survival prospects and fewer post-surgery complications among treated cancer patients. Moreover, we have also provided evidence that this effect seem to be mainly driven by surgeon learning-by-doing. In this section we further examine the mechanism(s) behind this effect.

Complexity of the surgery

One interesting hypothesis to investigate is whether learning is stronger for more complex tasks — if more heterogeneous tasks increases the payoff of experience in terms of surgical quality we would expect more learning to take place in such procedures. In order to analyze learning and complexity we estimate the volume effects separately using our IV model for each type of cancer which we categorize by complexity according to the "Birkmeyer top six" (see e.g. Birkmeyer *et al.* (2003)). In this measure, breast cancer surgery is considered to be a less complicated procedure than both prostate and colorectal

cancer surgery. This is also indicated by the significantly higher average four-year mortality for the latter two types of cancer compared to breast cancer.

The results from separate estimation of the different cancer types are presented in Table 6.1 showing a positive, however non-significant (due to decreased sample sizes), volume effect for all three types of surgeries. Even if imprecisely estimated, the lower point estimate for the less complex breast cancer surgery — compared to the more complex intestine and prostate cancer surgeries — indicates a positive relationship between complexity and learning, which would be expected if learning pays off more for more heterogeneous operations. Generalizing this result, as much of the existing learning-by-doing literature consider mostly manual tasks in the manufacturing industry, total productivity derived from learning in the economy may be greater than previously thought.

Table 6.1. *IV estimates of the effect of surgical volume on post-surgery survival by complexity of the surgery*

	Breast (1)	Colorectal (2)	Prostate (3)
log(volume)	0.017	0.066	0.059
	(0.020)	(0.087)	(0.051)
Mean survival rate	0.86	0.62	0.71
First-stage F-stat.	311.9	93.3	55.6
Observations	50,096	29,334	25,693

NOTE.— The table reports point estimates and (standard errors) from estimation of our preferred model specification from column (6) of Table 5.1 on different subsamples of our main analysis sample. Each column corresponds to the estimation results conditional on cancer type; breast, colorectal or prostate cancer. Outcome is four year survival after cancer surgery. All models include calendar time, hospital and surgery fixed-effect as well as regional linear trends and are estimated using the case-mix restricted sample excluding everyone with a designated hospital which at some point is either opened or closed. Standard errors are clustered at the hospital level. *, ** and *** denote significance at the 10, 5 and 1 percent levels.

Non-linear effects

So far we have focused on linear volume effects, mainly for methodological reasons, since the instrumental variable strategy relies on variation capturing a weighted average effect for hospitals of different sizes. Non-linear effects are hence difficult to investigate using these techniques. Instead, to analyze non-linearities we return to a hospital fixed-effects specification, relying on within-hospital variation over time to estimate the volume effects.[36] This way we are able to specify a more flexible model by dividing volume into volume

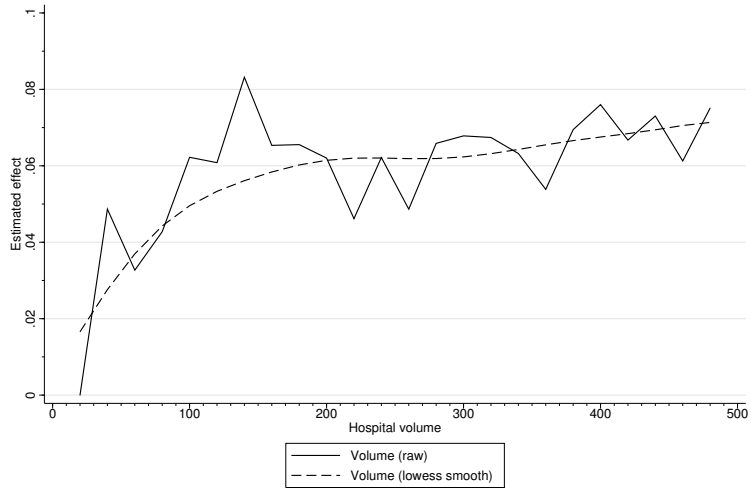
[36]In this empirical setup we maintain the identifying assumption that "dynamic" selective referral is non-existent, in the sense that patients do not sort themselves into different hospitals based on hospital quality trends. However, as we are only interested in the volume-quality relationship over the volume distribution, and not the average volume effect, this may be a more innocuous assumption in this setting.

bins of twenty and estimate separate effects for each bin; i.e. we estimate

$$y_{ihst} = \lambda_t + \lambda_h + \lambda_s + volume_{ht}^{bin}\gamma_{bin} + X_i\beta_X + \varepsilon_{ihst} \tag{6.1}$$

where $volume_{ht}^{bin}$ is a set of indicator variables for each volume bin, where $bin = 1-20, 21-40, ..., 480-500$. The estimated coefficient vector, γ_{bin}, normalized so that the coefficient for the first volume bin is zero, is plotted in Figure 6.1 along with a local polynomial smoothed line. The figure reveals interesting non-linearities of the volume effect, resembling a traditional learning curve. Specifically, the volume gradient is high at low volumes, diminishes over the volume distribution and reaches an empirical "learning threshold" for hospital volumes over 140 surgeries p.a. after which there are practically no volume effects left in our data.[37] As a large proportion of hospitals in our data operates below — as well as above — the empirical threshold we identify, this result lends some policy implications to the efficiency of health care organization — at least in Sweden.

Figure 6.1. Non-linear volume effects on four-year post-surgery survival probability



NOTE.—The figure plots the estimated coefficients of surgical volume on four-year post-surgery survival from a fixed-effects regression model including regional, calendar year and surgery fixed effects, see equation (6.1) in the results section. The non-linear volume effects are estimated by including dummy variables for each twenty volume bin constructed from the analysis data, using the first volume bin as reference category. The solid line pertains to the non-smoothed relationship while the dashed line is obtained by running locally weighted regressions of the estimated volume coefficients on the corresponding dependent variable.

[37] Note that we do not constrain the effect to be diminishing in this analysis by using the log transformation of volume in the regression model.

Individual or organizational learning

Given that we have identified learning-by-doing in surgery there is still a question whether these observed learning effects are primarily due to acquired experience embodied in individual surgeons or derived from experience obtained on the level of organization.[38] This relates closely to the individual vs. firm-specific human capital dichotomy in the labor economics literature in which important sources of productivity are lost as surgeons leave the hospital — if individual human capital is the driving force underlying the learning effects. However, if productivity from learning is primarily derived from the organization level, the productivity loss would mainly be carried by the individual surgeon.

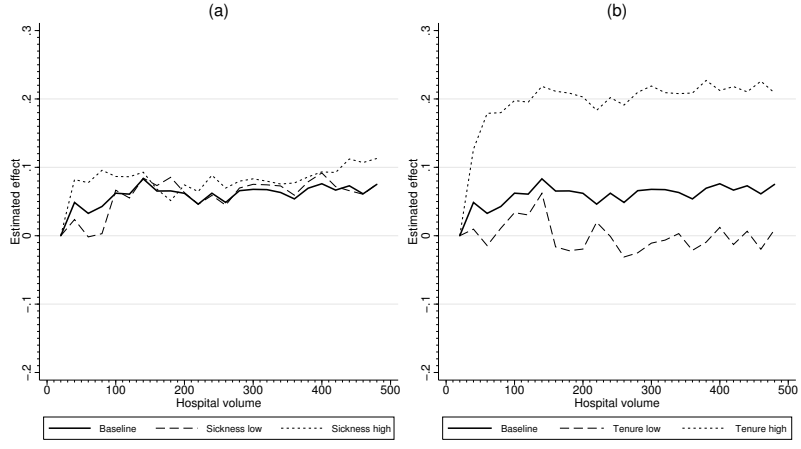
To attempt to separate the individual from the organizational learning in our data we follow a similar reasoning as in Levitt *et al.* (2012). In this article it is argued that, if organizational learning-by-doing is important, learning should be the same at organizations with high and low employee turnover. The intuition is simple; if the entire organization learns from experience, then the fact that employees are constantly replaced at high turnover organizations should not impact the relationship between volume and quality, whereas if experience is mainly embodied in individual employees we should observe less learning at organizations with a high degree of turnover.

We test the turnover hypothesis using supplementary information from employment records which allows us to compute each surgeon's tenure. Based on average surgeon tenure we separate hospitals into groups of low and high turnover and subsequently estimate the fixed-effects model from the previous section separately for each hospital group. The result, shown in Figure 6.2, clearly shows that the volume effects are mainly driven by the low turnover clinics, suggesting that the learning effects seem to mainly derived from individual surgeon experience.

To further investigate individual learning we may draw additional conclusions from the observed volume-per-surgeon ratio at each clinic. In particular, if the number of surgeons remains the same when additional surgeries have to be performed at the remaining clinics as a consequence of the clinic closures, this means that each surgeon has to perform more surgeries — and thus accumulating experience at a higher rate than in pre-closure years. However, if the number of surgeons do increase, so that the volume-surgeon ratio stays the same, then it is unlikely that the volume effect is driven by individual learning-by-doing. Thus, to test this hypothesis we therefore re-estimate our preferred IV model using the hospital number of surgeons at each hospital as outcome rather than four-year survival. The results reported in column (3) of Table 5.5 do not show any indication of an increased number of surgeons at remaining clinics as a response to the additional inflow of patients, hence implying

[38] See the discussion on learning-by-doing in surgery in section three.

Figure 6.2. Non-linear volume effects on four-year post-surgery survival probability by high and low hospital average turnover



NOTE.—The figure plots the estimated coefficients of surgical volume on four-year post-surgery survival from a fixed-effects regression model including regional, calendar year and surgery fixed effects, see equation (6.1). The non-linear volume effects are estimated by including dummy variables for each twenty volume bin constructed from the analysis data, using the first volume bin as reference category. Each line corresponds to a different sample category; the solid line pertains to the aggregated volume effect from Figure 6.1 using the full sample while the dashed (dotted) lines pertain to hospitals with more (less) than median turnover rates, estimated by pooling average surgeon tenure across all hospitals and all years in our sample.

that pre-existing surgeons in these clinics really perform a greater number of surgeries after the closures.

Production costs

Finally, besides improving patient health outcomes it may also be important for health care authorities to have an idea whether health care production costs are affected by learning-by-doing. In particular, increased learning might lead to lower costs for a hospital in terms of fewer per patient days of inpatient care if surgeons and the organization as a whole become more efficient at performing surgery. We analyze this important aspect of health care production costs by estimating our IV model using information on the number of days in hospital for each patient in association with the surgery as outcome. From column 4 of Table 5.5 we see that the number of hospital days does not seem to have changed significantly as a consequence of the volume increase from from the hospital closures. From this we conclude that this particular aspect of health care production costs does not seem to have been affected by increased efficiency from learning.

7 Conclusions

This paper provides new evidence of learning-by-doing by finding a direct causal link between production volume and quality using unique and detailed data on more than 100,000 episodes of advanced cancer surgery. We find that increases in hospital surgical volume significantly improves both patient survival prospects and leads to fewer post-surgical complications. The estimated effects are stronger at lower levels of production and for more complex procedures consistent with a learning-by-doing hypothesis in which experience from treating highly heterogeneous subjects plays a fundamental role for the learning process. We also show that these effects are not driven by selective referral or self-selection, changes in the patient population nor by organizational changes or heterogeneity at the hospital level.

Our paper contributes to several different literatures; first, we relate to the empirical volume-outcome relationship which has generated much interest in the medical literature. Exploiting the specific features of the public health care system in Sweden we are able to overcome many of the empirical difficulties that earlier contributions have struggled with. In contrast to the US health care sector, where market structure — and thus hospital volume — is an endogenous outcome of a competitive process, the specific institutional context of the Swedish health care sector provides us with variation in hospital volume that is unlikely to be driven by such forces. Rather, hospital closures generating shifts in surgical volume are in our application driven by political and bureaucratic considerations and likely to be unrelated to underlying population health.

Second, we contribute to the literature on general learning-by-doing and productivity growth which, while believed to be a key factor determining growth, has been scarcely investigated in a reliable causal sense thus far — in particular with respect to the specific mechanisms through which it operates. Using the plausibly exogenous variation in surgical volume derived from the closures and openings of cancer clinics within Swedish health care regions we provide evidence of rapid learning in a high-skill context with high stakes, in contrast to the heavy focus on the predominantly manual manufacturing industry prevalent in much of the previous learning-by-doing literature. Moreover, focusing on outcomes directly related to quality improvements, such as survival rates, subsequent surgeries and re-admissions we are able to shift the focus of the learning effect from quantity to quality aspects. The evidence we provide on the volume-quality relationship can hence be directly related to the mechanisms through which learning-by-doing arises.

Finally, our paper provides important implications for the organization of health care in publicly provided health care systems. Most previous empirical contributions on the volume-outcome nexus in health care has used as the contextual outset a competitive system in which competition effects and determinants of hospital size are interconnected in complicated ways. We find, in the context of a public health care system, that many of the hospitals in our

data operate at surgery levels far below or above the empirical learning threshold we identify and, as such, there seem to exist a role for policy makers to encourage mergers and consolidation in order to increase efficiency in terms of quality outcomes, at least for more advanced surgical procedures. In fact, such thresholds have already been called for by researchers and policy-makers alike (Epstein, 2002; Shahian and Normand, 2003).

References

ANGRIST, J. and KRUEGER, A. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, **15** (4), 69–85.

ARROW, K. (1962). The Economic Implications of Learning by Doing. *Review of Economic Studies*, **29** (3), 155–173.

BENKARD, L. (2000). Learning and Forgetting: The Dynamics of Aircraft Production. *American Economic Review*, **90** (4), 1043–1054.

BIRKMEYER, J., STUKEL, T., SIEWERS, A., GOODNEY, P., WENNBERG, D. and LUCAS, F. (2003). Surgeon Volume and Operative Mortality in the United States. *New England Journal of Medicine*, **349** (22), 2117–2127.

BLOOM, N., PROPPER, C., SEILER, S. and VAN REENEN, J. (2011). *The Impact of Competition on Management Quality: Evidence from Public Hospitals*. Discussion Paper 983, CEP.

DARR, E. D., ARGOTE, L. and EPPLE, D. (1995). The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises. *Management Science*, **41** (11), 1750–1762.

EPSTEIN, A. (2002). Volume and Outcome — It is Time to Move Ahead. *The New England Journal of Medicine*, **346** (15), 1161–1164.

FARLEY, D. E. and OZMINKOWSKI, R. (1992). Volume-Outcome Relationships and in Hospital Mortality: The Effect of Changes in Volume over Time. *Medical Care*, **30** (1), 77–94.

GAYNOR, M. (2006). *What Do We Know About Competition and Quality in Health Care Markets?* Working Paper 12301, NBER.

—, SEIDER, H. and VOGT, W. (2005). The Volume-Outcome Effect, Scale Economies, and Learning-by-Doing. *The American Economic Review (Papers and Proceedings)*, **95** (2), 243–247.

GOWRISANKARAN, G., HO, V. and TOWN, R. (2006). *Causality, Learning and Forgetting in Surgery*. mimeo 17608, University of Minnesota.

HALM, E. A., LEE, C. and CHASSIN, M. R. (2002). Is Volume Related to Outcome in Healthcare? A Systematic Review and Methodological Critique of the Literature. *Annals of Internal Medicine*, **137** (6), 511–520.

HAMILTON, B. and HAMILTON, V. (1997). Estimating Surgical Volume-Outcome Relationships Applying Survival Models: Account for Frailty and Hospital Fixed Effects. *Health Economics*, **6** (4), 383–395.

— and HO, V. (1998). Does Practice Make Perfect? Examining the Relationship between Hospital Surgical Volume and Outcomes for Hip Fracture Patients in Quebec. *Medical Care*, **36** (6), 892–903.

HUGHES, R., GARNICK, D., LUFT, H., MCPHEE, S. and HUNT, S. (1988). Hospital Volume and Patient Outcomes. The Case of Hip Fracture Patients. *Medical Care*, **26** (11), 1057–1067.

JOVANOVIC, B. and NYARKO, Y. (1995). A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves. *Brookings Papers on Economic Activity, Microeconomics*, pp. 247–305.
KAHN, J., TEN HAVE, T. and IWASHYNA, T. (2009). The Relationship between Hospital Volume and Mortality in Mechanical Ventilation: An Instrumental Variable Analysis. *Health Services Research*, **44** (3), 862–879.
KIZER, K. (2003). The Volume-Outcome Conundrum. *New England Journal of Medicine*, **349** (22), 2159–2161.
KLEIBERGEN, F. and PAAP, R. (2006). Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics*, **133** (1), 97–126.
LARSSON TAGHIZADEH, J. (2009). *Are Client Concerns Invisible and Weak in Sweden? About Protest Movements and their Influence on the County Council's Policy*. Master thesis, Department of Political Science, Uppsala University.
LEVITT, S. D., LIST, J. A. and SYVERSON, C. (2012). *Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant*. Working paper 18017, NBER.
LINDBOM, A. (2013). *Waking up the Giant? Hospital Closures and Electoral Punishment in Sweden*. mimeo, Department of Political Science, Uppsala University.
LUCAS, R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, **22** (1), 3–42.
LUFT, H. S., HUNT, S. and MAERKI, S. (1987). The Volume-Outcome Relationship: Practice-Makes-Perfect or Selective-Referral Patterns? *Health Services Research*, **22** (2), 157–182.
LUNDBERG, E. (1961). *Produktivitet och Räntabilitet*. Stockholm: Norstedt & Söner.
NORTON, E. C., GARFINKEL, S. A., MCQUAY, L. J., HECK, D. A., WRIGHT, J. G., DITTUS, R. and LUBITZ, R. M. (1998). The Effect of Hospital Volume on the In-Hospital Complication Rate in Knee Replacement Patients. *Health Services Research*, **33** (5), 1191–1210.
SHAHIAN, D. and NORMAND, S. (2003). The Volume-Outcome Relationship: From Luft to Leapfrog. *Annals of Thoracic Surgery*, **75** (3), 1048–1058.
SOCIALSTYRELSEN (2007). *Cancervården i Sverige. Kvalitet, Struktur och Aktuella Utmaningar (Cancer Care in Sweden. Quality, Organization, and Recent Challenges)*. Tech. Rep. 2007, Socialstyrelsen.
THOMPSON, P. (2001). How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study. *Journal of Political Economy*, **109** (1), 103–137.
— (2012). The Relationship Between Unit Cost and Cumulative Quantity and the Evidence for Organizational Learning-by-Doing. *Journal of Economic Perspectives*, **26** (3), 203–224.
THORNTON, R. and THOMPSON, P. (2001). Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding. *American Economic Review*, **91** (5), 1350–1368.
TSAI, A., VOTRUBA, M., BRIDGES, J. and CEBUL, R. (2006). Overcoming Bias in Estimating the Volume-Outcome Relationship. *Health Services Research*, **41** (1), 252–264.
WRIGHT, T. (1936). Factors Affecting the Cost of Airplanes. *Journal of Aeronautical Sciences*, **3** (4), 122–128.

Appendix A Tables and figures

Table A.1. *Sample statistics for closed/opened and remaining clinics*

	Closed/opened clinics	Remaining clinics
Patient characteristics		
Age	66.7 (12.8)	65.3 (12.3)
Post-secondary	0.23	0.27
Immigrants	0.09	0.13
Females	0.74	0.58
Hospital characteristics		
# Cancer surgeries	117 (51)	439 (263)
Days of care	9.2 (32.0)	8.6 (22.5)

NOTE.— The table reports shares/averages for different characteristics of the patient population by type of cancer clinic. Remaining clinics are defined as cancer clinics within the region that remain when a clinic within the same region is closed or opened. Standard deviations are reported in (parentheses).

Essay 4: To work or study while studying? Student aid design, spillovers and the efficiency of the tertiary education system*†

1 Introduction

Most OECD countries provide financial aid for college students, with the common goal of increasing college attendance and completion (OECD, 2008). Increasing graduation rates and the rate at which individuals attain higher education in general are declared social objectives in many countries. Long times-to-graduation may involve individual monetary costs by shortening the careers of college graduates, resulting in lower lifetime income. Furthermore, a slower study pace may also include considerable social costs from e.g. reductions of aggregate labor supply and increasing dependency ratios.[1] In particular, this phenomenon may be seen in the light of the increasingly aging populations in many countries and the economic consequences these demographic changes give rise to.

Previous research has typically found evidence that costs of attending college has an impact on the individual enrollment decision.[2] These costs mainly comprise college tuition fees and student aid levels but also the opportunity

*Co-authored with Marie Gartell.

†The authors thank Olof Bäckman, Robert Eriksson, Per Johansson, Håkan Regnér, Anna Sjögren, Johan Vikström, Johnny Zetterberg and seminar participants at the Institute for Futures Studies, IFAU-Uppsala, ESPE 2012, EALE 2012 and the workshop on the Economics of Education and Education Policy in Uppsala. Financial support and data provided by the Institute for the Evaluation of Labour Market and Education Policy (IFAU) are gratefully acknowledged.

[1]For example, Brodaty *et al.* (2008) provide evidence from France where graduates with longer than average time-to-graduation have significantly lower wages and employment rates in their early career. Moreover, Holmlund *et al.* (2008) show that post-graduation work experience is more valuable than pre-graduation work experience, implying that graduation *per se* is important. Brunello *et al.* (2003) investigates the expected times-to-graduation in ten European countries and find that the fraction of students who expect to graduate at least one year later than required ranges from more than 30 percent in Sweden and Italy to almost zero in the UK and Ireland.

[2]For studies of the impact of tuition levels on enrollment decisions, see e.g. Manski and Wise (1983); McPherson and Shapiro (1991a,b); Kane (1994); Rouse (1994); Hoenack (1971); Ehrenberg and Sherman (1987); Moore *et al.* (1991). For studies of the impact of student aid on enrollment decisions, see e.g. Schrøter Joensen (2010); Skyt Nielsen *et al.* (2010); Baumgartner and Steiner (2006); Linsenmeier *et al.* (2006); Dynarski (2002, 2003); van der Klaauw (2002); Reuterberg and Svensson (1994); Fredriksson (1997); Hammarström (1996).

cost of time allocated to e.g. market work. The literature has primarily focused on the extensive study margin (i.e. the decision of whether to attend college or not) and less so on the intensive study margin (i.e. decisions made within a study spell).[3] While the former primarily concerns the extent to which the higher educational system is able to match prospective students to corresponding educational pursuits, the latter relates more to the throughput and study duration of enrolled college students.

There are many factors that could potentially affect the average duration of study within countries; the composition of the student population with respect to ability and other individual characteristics, the opportunity cost of studying, the degree of mismatch between student preferences and educational track, the possibility of part-time studies and the generosity and/or strictness of the student aid system. These, and other, factors are likely to contribute to the widely varying average study durations observed across countries around the world. To assess which of these hypothetical factors are more or less important in this aspect is critical for policy makers in order to design efficient national educational systems.

In this paper we focus on a specific factor which is likely to impact average study durations in most countries; namely the degree of labor supply performed by college students. Part-time work while enrolled in higher education is common in many OECD countries in order for students to top up their incomes from other sources. The average OECD student in 2003 was employed about 27 percent of full-time while studying, see e.g. OECD (2005). While effects from working while studying could be positive for later labor market outcomes (cf. Light (2001); Hotz *et al.* (2002); Häkkinen (2006); Geel and Backes-Gellner (2012)), there are also potential adverse effects associated with it, such as increased risks of dropping out and prolongation of times-to-graduation (cf. Ehrenberg and Sherman (1987); Stinebrickner and Stinebrickner (2003); Bound *et al.* (2012)).

We add to the literature on the effects of college students labor supply on academic achievement by exploiting the contents of a Sweden public student financial aid reform. Specifically, the intervention altered the relative cost of financing college education through taking up student loans and engaging in market work, respectively. As an increased student labor supply potentially decreases time available for studying, any such relative change in the cost of study financing might also carry an indirect effect on the study pace of college students. In this paper, we make use of the intervention in the Swedish stu-

[3]Three studies examining the effect of student aid on study efficiency are; Häkkinen and Uusitalo (2003), who evaluate times-to-graduation for Finnish college students in the 1990s during which time a major student aid reform was implemented in the country. Nielsen Arendt (2013) analyze the impact of the financial aid system on student drop-out and completion behavior using policy variation derived from a Danish student aid reform. Finally, Schrøter Joensen (2010) models the sequential nature of the college-to-employment decisions of individuals within a structural dynamic model framework.

dent financial aid system in order to test this *time-reallocation* hypothesis and thereby contribute with important insights on the extent to which the design of the student aid system may affect a crucial policy parameter, i.e. the average study duration of college students.[4]

The intervention, introduced in 2001, was a comprehensive reformation of the Swedish public student aid system.[5] The major rule changes consisted of *i)* a substantial increase in the exempt amount of income students were able to earn without being penalized with a reduction in the student aid, and *ii)* a significant tightening of the loan repayment rules. The total amount of student aid remained, however, unchanged. Taken together, the relatively more costly student loan and the substantial tax credit on earnings hypothetically incentivized college students to substitute some of their student loans for earnings.

To empirically investigate the time-reallocation hypothesis, we sample students enrolled prior to the intervention and estimate effects on annual earnings and study pace attributed to the reform using variation over time. Specifically, we sample all students who enrolled in an academic program between 1997 and 2000 and follow them from enrollment until they complete the number of academic credits equivalent to a bachelor's degree. To this end, we use detailed Swedish individual-level register data on incomes, academic achievement and information on demographic and socioeconomic factors. Furthermore, previous literature have found evidence that students from a lower socioeconomic background are generally more dependent on — and more sensitive to changes in — the student aid system.[6] We follow this literature and estimate relative effects of the intervention by socioeconomic background in order to investigate whether students from a lower social background responded relatively more to the intervention than students from a higher social background.

We find that the intervention induced students from a low socioeconomic background to increase their relative student earnings with an estimated 25 percent compared to students from a high socioeconomic background. Furthermore, we find that this change coincided with a ten percent average decline in the relative study pace for students from a low socioeconomic background. We interpret these results as supporting the hypothesis that the policy change induced students, in particular from a lower socioeconomic background, to increase their labor supply in a way that indirectly caused their relative study

[4]Declining reported number of hours of studying among college students seem to be a major concern in the U.S. (see e.g. Boston Globe (2010)). In particular, Babcock and Marks (2011) estimate that the effective number of hours of studying declined for the average full-time college student by one-third (from 40 to 27 hours) per week between 1961 and 2003. They attribute this drop mainly to falling achievement standards in post-secondary educational institutions (Babcock and Marks, 2010).

[5]See CSN (2007) for a detailed overview of the 2001 student aid reform and its implementation.

[6]See e.g. Ehrenberg and Sherman (1987); Becker (1993); Card (1999); Eckstein and Woplin (1999); Bettinger (2004); Cameron and Taber (2004).

pace to fall. This interpretation is further reinforced by the results of a number of robustness checks we subsequently perform to corroborate the main results. Hence, we find robust evidence that the design of the student aid system is likely to be an important factor influencing times-to-graduation, at least for more credit-constrained college students.

Since the intervention primarily affected relative, rather than absolute, costs of studying, we see potential to generalize the findings of this analysis. Even if the magnitude of the effects might vary across different educational contexts, the results may still be important for policy purposes as they hint that students indeed react to economic incentives in the student financial aid system. Moreover, since college students in most OECD countries work part-time while enrolled, there exist few reasons to believe that students in other countries would behave in a qualitatively different way.

The remainder of this paper is organized as follows; the next section briefly outlines the main characteristics of the Swedish university system and the financial aid for college studies in Sweden. The empirical approach and the data used in the analysis are discussed in section three. Results from estimation are presented in section four. Finally, a short summary and some concluding remarks are offered in section five.

2 The Swedish system of higher education[7]

The Swedish system of higher education is financed and regulated by the Swedish Government and Parliament. Historically, the educational system has undergone a major expansion in recent decades. The number of full-time equivalent college students increased dramatically from around 160,000 in 1991 to 300,000 students in 2009 and the number of universities and colleges[8] increased from five to 25 during the same time period (HSV, 2001, 2010). The expansion has mainly been driven by a political objective of the Swedish Government stating that at least fifty percent of each birth cohort should be enrolled in higher education before they reach the age of 25.

A national administrative authority handles admissions to all colleges. High school GPA is the primary selection instrument governing entry into the higher education system. However, a national aptitude test and previous work experience may also be taken into account in the selection process. The graduation requirement is a minimum of 120 completed academic credits where

[7] See HSV (2004, 2006, 2007)) for a more detailed description of the Swedish system of higher education.

[8] We use the terms university, university college and college interchangeably throughout the paper.

each credit corresponded to one week of full-time study and 120 credits corresponded to three years of full-time studies.[9]

Two general student types can be distinguished; program and course students. Program students, who constitute approximately two-thirds of the total student population, enroll in a program normally lasting three or more years, while course students register for separate courses that typically last only for one semester. The graduation requirements are, however, the same for both student types.

2.1 Financial aid for college students in Sweden

The overall political motive of the public student financial aid system in Sweden is to encourage participation in higher education and to reduce social stratification in the population of students entering higher education. For these reasons Swedish universities have no tuition fees. Instead, the Swedish Government offers universal financial support for all Swedish citizens. In particular, all students admitted to a Swedish university are eligible for financial support irrespective of their family background, given that the individual intends to study for at least three weeks and fulfill an age restriction. The student aid can be received for a maximum of 240 weeks, equivalent to six years of full time studies (Bill no.1999/2000:10, 1999).

Since the introduction in 1965, the Swedish student financial aid has consisted of a grant which, since 1989, has been supplemented by a student loan with relatively generous repayment rules. The loan component constitutes the bulk of the aid for a full-time student, amounting to about two-thirds of the total aid. Students have flexibility in choosing whether to only receive the grant or to also include part of or the full loan amount over the course of study. All students are also allowed to combine student aid with part-time work up to a specified earnings ceiling (exempt amount) without being penalized with a proportional reduction of the student aid.[10] In order to retain the student aid over the course of study, the student must, for each academic study year, meet a certain completion level on the share of academic credits he or she receives student aid for.[11]

The student financial aid reform, introduced in the fall semester of 2001, included the most comprehensive changes in the Swedish student financial aid system since its introduction (CSN, 2007). In particular, while the total

[9]Since 2007, as a result of the Bologna process, one academic credit in the old system corresponds to 1.5 points in the new system. The time period we study here does not cover 2007 and later years.

[10]The percentage reduction in the student aid is 61 percent of the amount of earnings that exceeded the exempt amount and is withdrawn from the grant and the loan component in proportion to their share of the total aid.

[11]The completion threshold is equal to 75 percent of the registered academic credits in the first year of studies and thereafter 50 percent for each subsequent study year.

amount of financial aid (approximately 7,000 SEK[12] per month) was left unchanged, the reform comprised three major changes; first, the grant share was increased by approximately 25 percent (from 27.8 percent to 34.5 percent of the total amount of student aid). The motivation for this change was to reduce social stratification into higher education by increasing the possibility for low-income groups to enter college. In terms of money, it implied that approximately 500 SEK per month was transferred from the loan to the grant part of the aid.

Second, the exempt amount on earnings increased dramatically from an annual amount of around 55,000 SEK to 90,000 SEK annually, i.e. an increase of about 67 percent. The Government motivated this change by *i)* making it possible for students to attain useful work experience and *ii)* to improve students living standards. This specific change meant in practice a substantial earnings tax subsidy over a wide income range for many students, due to the penalizing of earnings above the exempt amount.

The third and final major change in the student financial aid system involved making the student loan relatively less generous. Specifically, college students in the former system started to repay their loans six months after they received their last aid payment and paid four percent of their annual taxable earnings until the age of 65 when all remaining debt was written off. In the new system, the repayment was instead calculated as an annuity such that the total debt should be repaid within a maximum of 25 years. Hence, the new repayment system implied a substantial increase in nominal interest rates for loan students. The reason for tightening the loan repayment schedule was primarily to reduce the total amount of outstanding debt by making the repayment more coherent with the size of the loan (Bill no.1999/2000:10, 1999).[13]

The new system covered both students who began studying in the fall of 2001 as well as students who were already enrolled in college. However, all student aid received before the fall of 2001 was subject to the previous rules.

2.2 Theoretical aspects of the intervention

The theoretical arguments underpinning the expected effects of the student aid reform on the behavior of college students originates from two basic tenets of production theory; first, a change in the relative price of two production factors should increase the relative utilization of the factor that fell in price; second, additional inputs of time into an activity should (weakly) increase the value of the output from the activity. Translated into the present application,

[12] One Swedish krona (SEK) corresponded to approximately €0.1 in 2001. We use 2001 currencies throughout the paper.

[13] Some minor changes of the student aid system was to reduce the possibilities of being granted an extension of the student aid after the eligible period was ended, increasing the age restriction for eligibility for student aid from 45 to 50 years and to enable students with children to apply for higher benefits.

more time used for studying should increase the annual number of completed academic credits and more hours devoted to market work should increase earnings. However, as time is a scarce resource, a rational economic agent has to make choices of where and how to allocate his or her time endowment in different activities in order to maximize utility. This necessarily means that more time spent in one activity (i.e. performing market work) will be at the expense of another (i.e. studying).

The public student financial aid reform affected the relative cost of financing college education with student loans and with earnings, respectively. In particular, the relatively more costly loans should have decreased the attractiveness of financing studies with loans while the simultaneous increase of the exempt amount on labor market income should have increased the attractiveness of financing studies with earnings. The intervention thus incentivized students with loans to substitute some of these loans for earnings by increasing their labor supply. If this reallocation of time implied less time for studying, then, as additional inputs of time used for studying should improve academic achievement, it might also have entailed an adverse spillover effect on the study pace of affected college students.

The contents of the student aid reform should primarily have impacted the behavior of students who were particularly dependent on the public student aid system for financing their studies. Previous studies have found evidence that students from a lower socioeconomic background are in general more responsive to changes in the costs of attending college than other students (see e.g. Ehrenberg and Sherman (1987); Becker (1993); Card (1999); Eckstein and Woplin (1999); Cameron and Taber (2004); Bettinger (2004)). One potential explanation for this finding might be due to more severe credit constraints among students from a lower socioeconomic background, whereas students from a higher socioeconomic background may have more resourceful parents who could provide additional financing alternatives. From these findings, we would, hence, expect students from a higher socioeconomic background to be less responsive to the contents of the student aid reform relative to students from a lower socioeconomic background.[14]

[14]This argument also has some empirical support in the Swedish context. Specifically, Hammarström (1996) investigates the reasons why high school graduates in Sweden do not enroll in higher education. Her results suggests that individuals from a low socioeconomic background are more dependent on financial aid when deciding whether to pursue higher education. Moreover, a survey conducted by the National Board for Student Aid administered to students with student aid in 2001 and 2003 reveals that approximately 60 percent of all students would "probably not" or "would not" have enrolled in higher education if no student aid existed. For students with both parents having less than high-school education the corresponding figure was 72 percent. For student with parents having at most high-school education the share was 64 percent. Finally, for students with post-secondary educated parents the share was only 42 percent (CSN, 2007).

3 Empirical strategy

The objective of the empirical analysis is to evaluate whether the intervention in the Swedish public student financial aid system of 2001 changed student college financing behavior and, as a consequence thereof, the relative study pace of students from different socioeconomic backgrounds. The main difficulty in empirically assessing these hypotheses is related to the fact that the intervention itself was likely to have endogenously changed the composition of college students. In particular, if the new student aid rules affected enrollment decisions of students from distinct socioeconomic groups differently, the estimated impact of the intervention would be biased. Based on the extensive existing evidence of the effects of college costs on enrollment decisions, we expect this problem to be present in our application.[15]

To circumvent problems from having an endogenously changing student population, we sample only students who were enrolled before the intervention and follow them over their course of study as the student aid reform was implemented. Specifically, we set up and estimate the following linear regression model using OLS,

$$Y_{ics} = \beta_0 + \beta_1 R_{ics} + \beta_2 SEB_{ic} + \beta_3 (SEB_{ic} \times R_{ics}) + X_{ics}\alpha' + (SEB_{ic} \times \lambda_s)\delta' + \lambda + \varepsilon_{ics}, \tag{3.1}$$

where Y_{ics} is the outcome of interest for student i of cohort c in study year s; i.e. the annual number of completed academic credits and annual student earnings, respectively. Furthermore, R_{ics} and SEB_{ic} are binary indicators for whether the student was studying in a post-reform year and whether the student comes from a high socioeconomic background, respectively. X_{ics} is a vector of controls and ε_{ics} is an assumed random error term. We moreover include fixed cohort and study year effects ($\lambda = [\lambda_c, \lambda_s]$) along with a flexible study year trend, interacted with the student socioeconomic background indicator, to control for potential heterogeneity across cohorts and over the course of study across student socioeconomic groups. The source of the variation used to estimate the parameter of interest in the model, β_3, stems from calendar time, as students in different enrollment cohorts are at different stages of their studies at the time the reform was introduced.[16]

The relative effect of the intervention for students from a high and low socioeconomic background is identified by the model under an assumption of common trends, i.e. that the relative trend in Y_{ics} across student socioeconomic groups would have been constant in absence of the reform. We informally evaluate this assumption by substituting the reform indicator in (3.1) for a set

[15] As we do not possess information on college applicants but only for enrolled students we are unable to condition on e.g. the ability composition of students over time.

[16] We have also run regressions including individual rather than cohort fixed effects with qualitatively unchanged results.

of calendar year dummy variables,

$$Y_{ics} = \gamma_0 + \gamma_1 SEB_{ic} + (SEB_{ic} \times \lambda_y)\gamma_y' + X_{ics}\alpha' + (SEB_{ic} \times \lambda_s)\delta' + \lambda + \varepsilon_{ics}, \tag{3.2}$$

where λ_y is defined as the set of calendar year dummies, $y = 1997, ..., 2005$, and $\lambda = [\lambda_c, \lambda_s, \lambda_y]$. If the common trends assumption is valid, we should not expect to see any significant effects from the estimated coefficients of the calendar year-interacted variables for years prior to the intervention.

3.1 Data and sampling

The data used in the analysis is maintained by Statistics Sweden and the National Board for Higher Education and consists of a number of merged administrative records. The specific registers used in the analysis are; the college registration records, containing individual-level data on field of study, education level, college of attendance and the number of registered and completed credits for each semester for all enrolled students; a longitudinal population-based register including information on demographic and socioeconomic characteristics for all Swedish citizens 16-64 years of age; and the Swedish employment register containing information about individual earnings.

To avoid identification problems relating to an endogenously changing student population and to make sure that the students we sample intend to pursue an academic degree, we restrict our sample to students who *i)* registered for a program with a theoretical length of at least 120 academic credits, corresponding to a bachelors degree or three years of full-time studies, and *ii)* enrolled in an academic program between 1997-2000.[17] We follow the students until they obtain 120 academic credits or a maximum of six years, i.e. the maximum number of years a student is eligible for student aid. Finally, we also exclude individuals from the year they have three future consecutive semesters without college registration (i.e. college drop-outs).[18]

We use parental educational levels to define the students' socioeconomic background. Specifically, we categorize students as having a high socioeconomic background, SEB_{High}, if both parents have a post-secondary education, and students as having a low socioeconomic background, SEB_{Low}, if neither parent has a post-secondary education. In the few cases where the parental

[17] Data on enrollment is available from 1995 and forward. However, we choose 1997 as the starting year for our analysis since the number of registered and completed credits for 1995 and 1996 was only coded for the full *academic* year, implying that we are unable to identify which *calendar* year the student actually obtained his or her credits for these years. As the remainder of the variables are coded on the calendar year level we therefore decided to drop these years from the analysis.

[18] As the intervention might have affected drop-out rates we investigate this potential problem as a robustness check in the next section.

educational level changes over time, we use the highest attained educational level observed in the data.

As regression outcomes, we use the number of annually completed academic credits to estimate the effect on study pace. We observe the number of completed academic credits each semester and aggregate these over the full calendar year to correspond with the other included variables. To estimate the effect on earnings, we use annual taxable labor market income from official tax records. We also have information on the level of student aid for each student. However, as we can only observe the total amount of student aid, we cannot separate the respective loan and grant share each student receives.[19]

The set of control variables we include in the analysis are; the specific college the student attended each year, immigrant status, gender, age at first enrollment, GPA from high-school and field of education. We define field of education as the field of the college program that the student enrolled in each year. If the individual has several registered fields of education in the same year, we choose the field in which the individual registered for the greatest number of academic credits. Finally, if the student did not register for any academic credits a given year, the field of education is coded the same as in the previous year. See Table 3.1 for a list and description of the included variables.

[19] As the student aid reform affected the student aid both through the grant (making it more attractive) and the loan (making it less attractive), the effect of the reform on the total level of student aid is theoretically ambiguous. In particular, it will depend on the relative densities of the marginal student aid distributions across student socioeconomic background groups. We investigate this further in the next section.

Table 3.1. *List of included variables*

Variable	Description
Reform	1 for years after 2001
Cohort$_i$	1 if the student has a first registration in an academic program in year i for $i = 1997, .., 2000$
Field$_i$	1 if the field of education is i for $i =$ Social science, Pedagogics Humanities, Science, Technology, Health Care, Other
Female	1 if female
Immigrant	1 if immigrant
Age at first registration	The student's age at first registration
SEB_i	1 if student's socioeconomic background is i for $i =$ Low, High
GPA	Grade point average from high school
Earnings	Work related earnings
Student income	Student aid income
Dropout	1 for all years subsequent to the student has been observed to drop out from college

NOTE.— All variables are defined on the calendar year level. Field of education is defined by the field in which the student register for the greatest number of academic credits each study year. Student socioeconomic background is defined as high (low) if both (neither) parent(s) have a post-secondary education. Earnings and student income pertains to the annual taxable labor market income and student grant and loan incomes, respectively. A dropout is defined as a previously enrolled student with three consecutive semesters without any registered nor completed academic credits.

The population of college students enrolling into higher education in Sweden between 1997-2000 amounted to approximately 271,000. Including only program students exclude 70,000 of these students. We also drop 2,000 individuals who were over 40 years of age when they enrolled, as they were ineligible for student financial aid. Furthermore, we exclude 5,000 students with missing information in the population registers. These individuals were predominantly foreign students and hence also ineligible for Swedish student financial aid. Finally, we keep only students from a high or a low socioeconomic background, excluding another 68,000 students. After applying these restrictions, we are left with approximately 126,000 individuals. Table 3.2 reports summary statistics of the analysis sample.

Table 3.2. *Descriptive sample statistics*

	Sample				
	All students	SEB_{Low}	SEB_{High}	AID_{Low}	AID_{High}
Cohort_{1997}	0.23	0.24	0.23	0.35	0.30
Cohort_{1998}	0.24	0.24	0.25	0.34	0.32
Cohort_{1999}	0.26	0.26	0.25	0.31	0.37
Cohort_{2000}	0.27	0.26	0.27	-	-
Humanities	0.11	0.11	0.11	0.08	0.11
Social Science	0.28	0.27	0.30	0.25	0.27
Science	0.18	0.17	0.21	0.20	0.18
Technology	0.19	0.19	0.20	0.25	0.19
Health	0.18	0.20	0.15	0.18	0.20
Other	0.05	0.06	0.03	0.04	0.05
Female	0.57	0.60	0.51	0.57	0.59
Immigrant	0.12	0.16	0.04	0.07	0.11
Age at first reg.	23.43	24.62	21.00	21.21	23.44
	(5.44)	(6.00)	(2.81)	(3.77)	(5.04)
SEB_{Low}	0.67	1.00	0.00	0.62	0.68
SEB_{High}	0.33	0.00	1.00	0.38	0.32
GPA	13.7	12.9	15.2	15.1	13.6
	(5.3)	(5.3)	(5.3)	(5.2)	(5.4)
Earnings	44,681	48,737	36,281	36,632	37,545
	(52,463)	(56,124)	(43,030)	(41,916)	(41,313)
Student income	35,414	35,710	34,629	15,556	45,679
	(22,485)	(23,293)	(21,469)	(9,635)	(17,974)
Dropout	0.31	0.36	0.24	0.22	0.23
Individuals	126,395	84,849	41,546	12,070	37,768

NOTE.— Standard deviations in (parentheses). Each column pertains to a specific sample, from left to right; the full analysis sample, subsamples of student from a low and a high socioeconomic background, and subsamples of students with average student income equivalent to only having grants and equivalent to having both grants and the maximum allowed level of student loan in their two first years of enrollment, respectively. Earnings and student aid incomes pertain to incomes from the two first two years of study and are measured in SEK (1 SEK corresponded to approximately €0.1 in 2001). See Table 3.1 for a list and description of the included variables.

4 Results

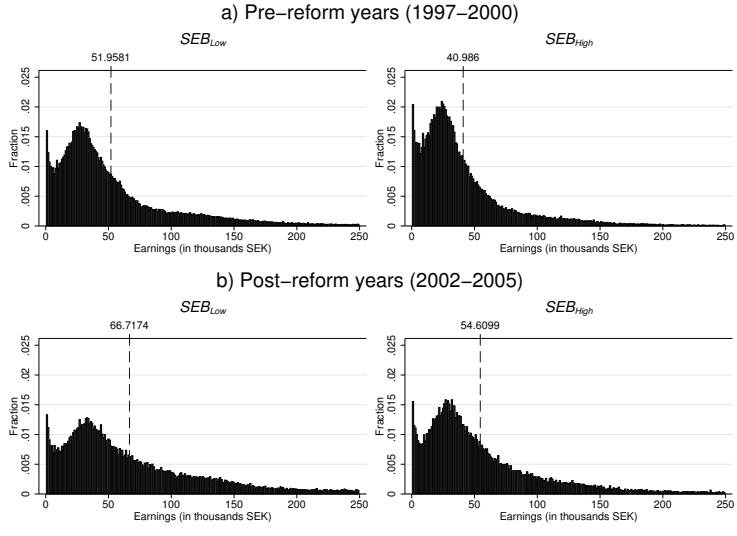
This section reports results from the empirical analysis outlined in the last section. We first present some descriptive evidence before subsequently turning to formal regression analysis. Finally, we discuss the results from a number of sensitivity analyses in order to corroborate the main findings of the paper.

4.1 Descriptive results

Figure 4.1 illustrates the sample distribution of student annual earnings for years before (upper panel) and after (lower panel) the intervention took place for students from a low (left sub-panels) and high (right sub-panels) socioeconomic background. The average earnings for each marginal distribution are also indicated in the plots. The figure shows *i)* that students from a low socioeconomic background earn substantially more than students from a high socioe-

conomic background on average before the reform, and *ii)* that both student groups earn more after, relative to before, the intervention took place. This earnings pattern hence provides some indication that students from a lower socioeconomic background were on average more constrained by the exempt amount than students from a higher socioeconomic background.

Figure 4.1. Distribution of annual student earnings for college students across time, by student socioeconomic background



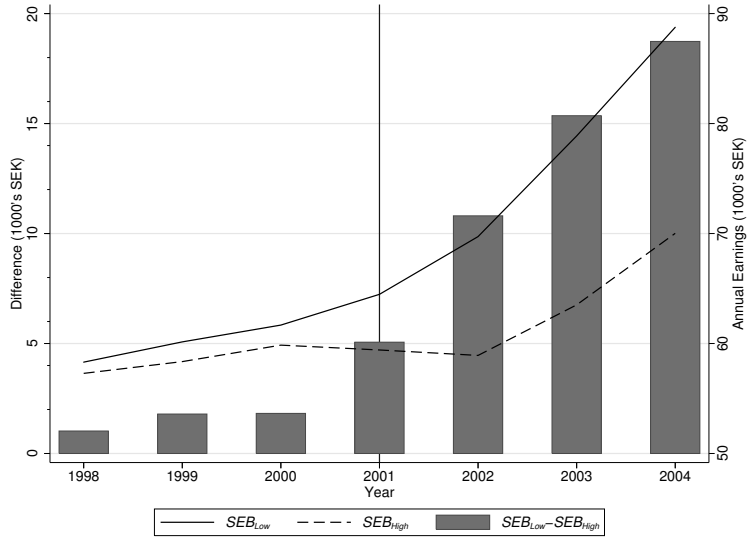
NOTE.—The figure illustrates the sample distributions of student earnings by student socioeconomic background and for years before/after the student aid reform. Panel *a)* pertains to earnings in the pre-reform years and panel *b)* pertains to the post-reform years. SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. The vertical line in each panel indicates the mean of each respective distribution in 1000s of SEK (1 SEK corresponded to approximately € 0.1 in 2001).

To further investigate whether the earnings exempt amount constrained students to work less than they optimally would have wanted, we study the subsample of students who were close to the earnings threshold prior to the intervention. More specifically, as the pre-reform annual earnings exempt amount was about 55,000 SEK, we study the earnings of students having average annual earnings between 50,000-60,000 SEK in their first two years of study, *before* the intervention took place.

Figure 4.2 presents the average earnings of these students over time and by student background. The vertical line in the figure indicates the year of the intervention and the grey bars illustrate the average earnings difference between the student socioeconomic groups for each year. Interestingly, the figure shows a trend break between the student groups at the time the reform was introduced; students from a low socioeconomic background increased their earnings relatively more than students from a high socioeconomic back-

ground. Hence, without making causal claims, this pattern further strengthens the conjecture that students from a lower socioeconomic background were more constrained by — and therefore reacted more strongly to the changed rules in — the public student financial aid system.

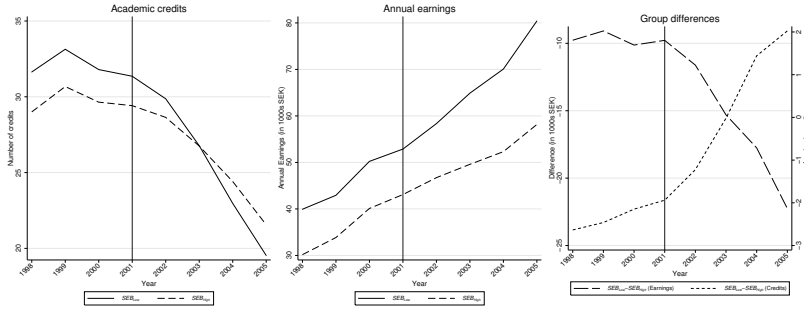
Figure 4.2. Annual earnings for college students close to the pre-reform earnings exempt amount across time, by student socioeconomic background



NOTE.—The figure illustrates the average earnings trends of students from a high (solid line) and a low (dashed line) socioeconomic background with average pre-reform earnings between 50,000-60,000 SEK in their two first years of study. The gray bins in the figure indicate the relative student group earnings difference for each year. The earnings exempt amount was approximately 55,000 SEK before the intervention took place. SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. The vertical line indicates the year of the policy change. Incomes are measured in SEK (1 SEK corresponded to approximately €0.1 in 2001).

We now turn to analyzing the full estimation sample. The left and middle panels of Figure 4.3 show the trends in average study pace and earnings by student socioeconomic background for all students in our sample, respectively. Furthermore, the right panel shows the average difference in these variables across student groups over time. The figure clearly shows that the year of the intervention coincided with both a relative increase in earnings, but also a relative *decline* in study pace, for students from a low socioeconomic background. This result is consistent with the empirical deduction of the time reallocation hypothesis, in which students from a lower socioeconomic background responded relatively stronger to the contents of the student aid reform by increasing their labor supply which, in turn, adversely affected their relative study pace.

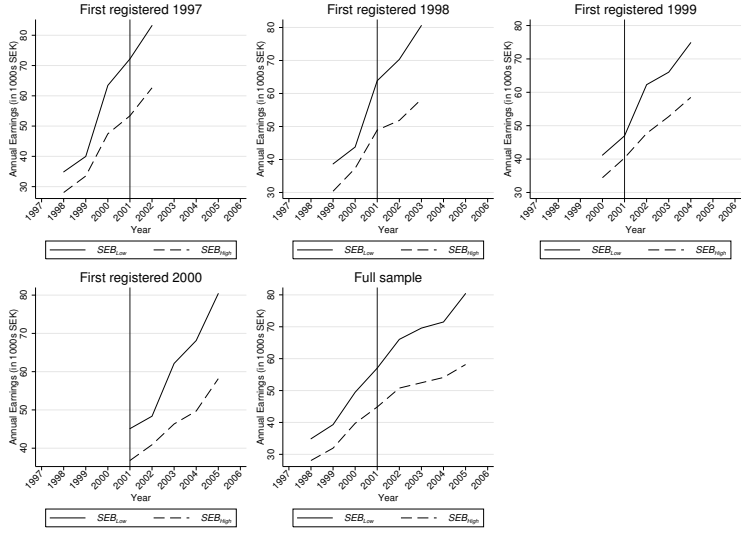
Figure 4.3. Average earnings and study pace for college students across time, by student socioeconomic background



NOTE.— The left-most (middle) panel plots the average number of completed academic credits (earnings) by student socioeconomic background over time. The right-most panel plots the corresponding student socioeconomic group difference for the same variables. SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. The vertical line indicates the year of the policy change. Incomes are measured in SEK (1 SEK corresponded to approximately € 0.1 in 2001).

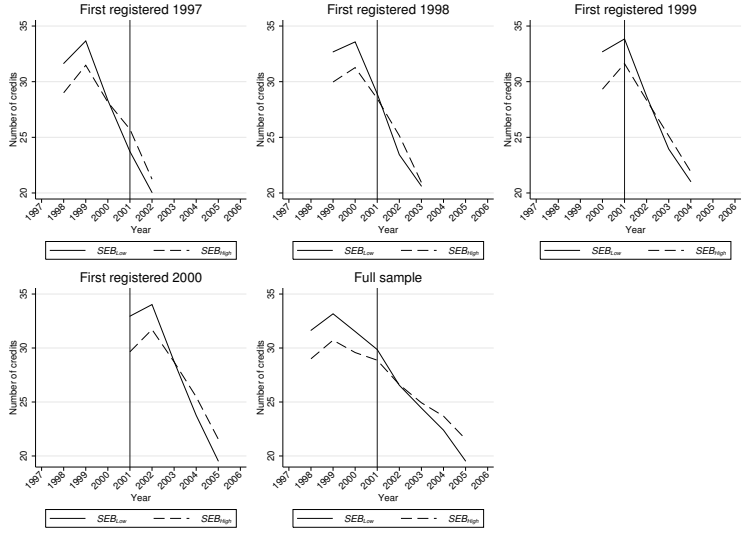
The descriptive evidence provided thus far, while indicative, should not be interpreted as causal effects from the changed rules in the student financial aid system. In particular, the observed outcome trends might simply reflect heterogeneity across the student socioeconomic groups over the course of their studies. To evaluate this concern, Figures 4.4 and 4.5 illustrate the trends in annual earnings and study pace of the two student socioeconomic groups for each enrollment cohort separately. These figures indicate that, while there is little sign of heterogeneity in study trends across cohorts, there is substantial variation across the student groups within cohorts. In order to control for these confounding study spell trends, we next turn to formal regression analysis.

Figure 4.4. Average earnings for college students across time, by enrollment cohort and student socioeconomic background



NOTE.— Each panel pertains to a specific enrollment cohort. SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. The vertical line indicates the year of the intervention. Incomes are measured in SEK (1 SEK corresponded to approximately € 0.1 in 2001).

Figure 4.5. Average number of completed academic credits for college students across time, by enrollment cohort and student socioeconomic background



NOTE.— Each panel pertains to a specific enrollment cohort. SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. The vertical line indicates the year of the intervention. Academic credits are measured on a calendar year basis.

4.2 Regression results

Table 4.1 reports coefficients from estimation of models (3.1) and (3.2) in the empirical section, with annual earnings (first two columns) and annual number of completed academic credits (last two columns) as outcomes, respectively. For each outcome, the first column reports the estimated β_3 from (3.1) and the second column reports the set of estimated γ_y coefficients from (3.2), with 1997 set as the base year. All models include the full set of controls, fixed effects for cohort and study year and a non-parametric study progression trend interacted with the student socioeconomic background indicator.

Starting with the estimated effect on student earnings, column (1) reports that students from a high socioeconomic background earn on average about 2,400 SEK less than students from a low socioeconomic background per year after, relative to before, the student aid reform took place. The estimate is highly significant and precisely estimated. As a comparison, a student from a low socioeconomic background earned on average around 10,000 SEK more per year than a student from a high socioeconomic background prior to the intervention, implying that the reform was responsible for a 25 percent increase in relative earnings between the student groups. Furthermore, from evaluating the year-by-year effects reported in column (2) we find a striking pattern; the relative trend in earnings is constant for all pre-reform years but shows a significant discontinuous jump in the year of the intervention and remaining so for all subsequent years. This estimated effect pattern thus provides some further empirical evidence that, in particular, students from a low socioeconomic background increased their labor supply as a consequence of the changed rules in the student financial aid system.

Next, the results for the study pace is reported in the last two columns of Table 4.1. We find a remarkably similar, but reversed, effect pattern on the relative number of annually completed academic credits. Specifically, the reported coefficient in column (3) is interpreted as that students from a high socioeconomic background complete a statistically significant 0.3 *more* academic credits on average each year after the reform took place, relative to students from a low socioeconomic background. Since students from a high socioeconomic background on average completed approximately 2.5 more academic credits annually prior to the reform, this corresponds to an increase in this difference of about ten percent. Finally, the estimated year-by-year effect pattern reported in column (4) brings further support that this change was in fact attributable to the changes in the student financial aid system.[20]

[20] At first sight it seems strange that we find no estimated effect in 2001 on completed academic credits. However, as completed academic credits are measured on the calendar year (rather than the academic year) level and as students typically complete most of their academic credits in the spring semester, any theoretical effects in 2001 would be captured by the estimated coefficient for 2002.

Table 4.1. *Estimated relative changes in annual earnings and study pace*

	Annual Earnings (in 1000s SEK)		Completed Number of Academic Credits	
	(1)	(2)	(3)	(4)
$1998 \times SEB_{High}$		-1.109		0.239
		(0.724)		(0.209)
$1999 \times SEB_{High}$		-1.114		0.356**
		(0.708)		(0.203)
$2000 \times SEB_{High}$		-1.122		0.233
		(0.709)		(0.200)
$2001 \times SEB_{High}$		-3.272***		0.218
		(0.753)		(0.233)
$2002 \times SEB_{High}$		-4.740***		0.495**
		(0.911)		(0.253)
$2003 \times SEB_{High}$		-4.685***		0.582**
		(0.885)		(0.284)
$2004 \times SEB_{High}$		-5.286***		0.804**
		(1.159)		(0.373)
$2005 \times SEB_{High}$		-4.725**		1.473**
		(1.895)		(0.577)
$R \times SEB_{High}$	-2.392***		0.287**	
	(0.361)		(0.128)	
Controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓
Study trends	✓	✓	✓	✓
Observations	426,546	426,546	426,546	426,546

NOTE.—The table reports the estimated β_3 and γ_y coefficients from estimation of regression models (3.1) and (3.2) in the empirical section. Columns (1) and (3) of the table report the results from including a binary indicator variable equal to one after the year of the policy change interacted with the student socioeconomic background indicator. Columns (2) and (4) instead include a full set of calendar year indicator variables, both in levels and interacted with the student socioeconomic background indicator with 1997 set as the base year. Controls includes age of first registration, gender and immigrant dummies, field of education, local unemployment rates, high school GPA, fixed cohort, study year and college effects and study year trends interacted with the student socioeconomic background indicator. SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. Incomes are measured in SEK (1 SEK corresponded to approximately €0.1 in 2001). Academic credits are measured on the calendar year and one credit corresponds to one week of full-time studies. Robust standard errors in parentheses. *, ** and *** denote significance at the 10, 5 and 1 percent levels. For variable definitions, see Table 3.1

The combined results on study pace and earnings lend substantial support for the time reallocation hypothesis. Students who were incentivized to increase their earnings from the changed rules in the student financial aid system seem to also have been exposed to an unintended spillover effect on their study pace, resulting in an widened academic achievement gap between students from a low and a high socioeconomic background. The estimated effects might seem small when compared to the total earnings and completed credits of an average college student in our sample. However, we stress the importance of keeping in mind that we estimate only relative and not average effects of the intervention. The changed rules might very well have substantially affected students from a higher socioeconomic background and the variation we

have at our disposal does not allow us to estimate the average effect of the intervention. On theoretical grounds, there is no reason to believe that the latter effect could not have been substantially greater than the relative effects we estimate here.

4.3 Robustness checks

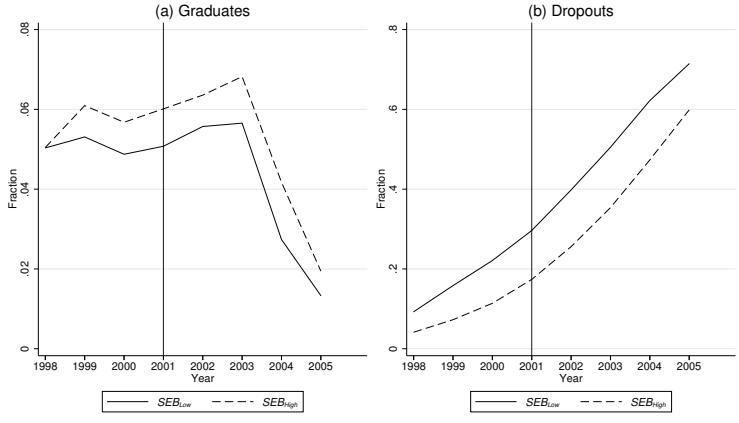
In this section we briefly discuss the results from a number of robustness analyzes we have carried out to examine the sensitivity of our main findings. Specifically, we investigate the potential endogenous censoring arising from changes in drop-out and graduation behavior, the ambiguous impact on the total uptake of student aid from grants and loans, respectively, and the sensitivity of the results with respect to model specification and the definition of student socioeconomic background.

First, if the contents of the student aid reform had an asymmetric impact on drop-out and graduation behavior of students from different socioeconomic backgrounds, then the censoring of these students might have introduced bias due to changes in the student composition over time. Moreover, if e.g. students from a higher socioeconomic background were *a priori* relatively more informed about the changes in the student aid system, they may have acted in a manner opposite to the time reallocation hypothesis — i.e. working less and studying more intensively — to avoid being subject to the new rules.

In order to investigate these concerns, panels *a)* and *b)* of Figure 4.6 illustrate the trends over time in the share of drop-outs and graduates among active students by student background group, respectively. From the trend patterns in the figure we see no indication that the two groups responded differently to the intervention in terms of these factors.[21] In addition, we have also re-estimated our empirical model including dropouts, resulting in slightly larger estimated effects (not reported). Intuitively, as students from a lower socioeconomic background were likely to have responded more to the changed rules also with respect to the likelihood of dropping out from college, we interpret this finding as that the reported estimates in Table 4.1 provide conservative inference of the parameter(s) of interest.

[21] We have also estimated regression models of type (3.1) using the two censoring mechanisms as outcomes. The results (not reported here) show no indication that the relative probabilities to drop out or graduate were significantly related to the intervention.

Figure 4.6. Share of college graduates and drop-outs across time, by student socioeconomic background

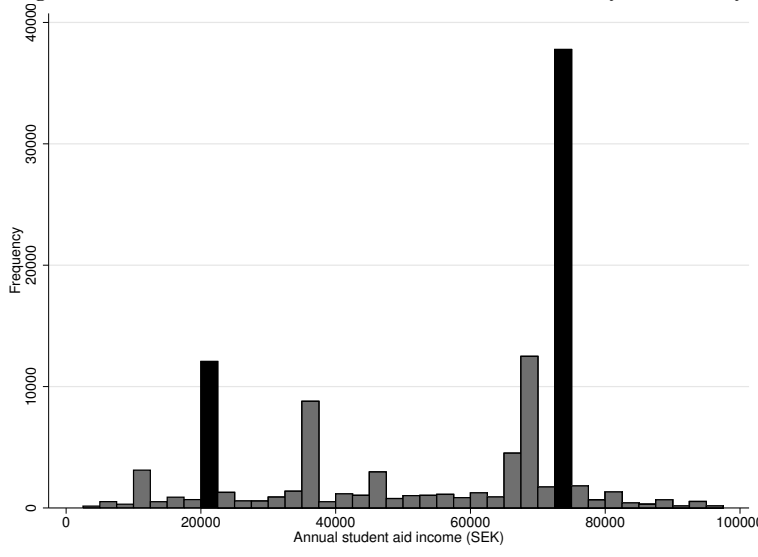


NOTE.—The figure illustrates the fraction of college *a)* graduates and *b)* dropouts for students from a high (dashed line) and a low (solid line) socioeconomic background over time. A college graduate is defined as a student with a total of at least 120 completed academic credits who had less than 120 completed academic credits the previous calendar year. A dropout is defined as a previously enrolled student with three consecutive semesters without any registered nor completed academic credits.

As we cannot distinguish between student loan and grant incomes in the data and the reform also entailed an increase in the grant share of the aid, we are restricted in assessing the impact of the student aid reform on the uptake of student loans. However, depending on the relative student aid densities across student socioeconomic background, the estimated effect may pick up compositional differences in the respective marginal student aid distributions rather than individual student behavior. In particular, if students from a high socioeconomic background include a relatively larger share of students with only grants than students from a low socioeconomic background, the effect could be, at least partially, driven by these compositional effects.

Even if we do not have information on the exact loan and grant shares of for each student, we could still analyze potential composition effects by comparing students with different levels of student aid income prior to the intervention. Specifically, we may "condition" on the effect from the increased grant share by comparing students who have pre-reform student incomes equivalent to only receiving the grant share and students receiving both loans and grants, respectively. Figure 4.7 illustrates the distribution of pre-reform student aid income for the full sample of students average over their first two years in college. The two black bars in the figure indicate the number of students whose average student aid income corresponded to only receiving grants (the left bin) and receiving grants *in addition* to the maximum allowed level of student loans (the right bin). Descriptive statistics for these student groups, denoted AID_{Low} and AID_{High}, can be found in the last two columns of Table 3.2.

Figure 4.7. Aid distribution of students in their second year of study



NOTE.—The figure illustrates the distribution of student aid income for students in their second year of study. The black bins indicate students with only grants (left bin) corresponding to a student aid income of between 20,000-22,500 SEK p.a. and students with maximum loans and grants (right bin) corresponding to a student aid income of between 70,000-72,500 SEK p.a. Incomes are measured in SEK (1 SEK corresponded to approximately €0.1 in 2001).

We estimate the same models for student earnings and study pace as before but replacing the student socioeconomic background indicator for the student aid income group indicator. Again, the estimated β_3 from equation (3.1) is reported separately by student socioeconomic background in columns (1) and (2) of Table 4.2. Moreover, in column (3) we use the full sample of students and estimate a "triple difference" model by including a full set of level and first and second level interactions of the two student group categories and the reform dummy.[22]

The results from this exercise show that students with both loans and grant, irrespective of socioeconomic background, earn significantly more and decrease their study pace more than students who only had the grant before the intervention took place. Moreover, these effects are also larger in magnitude than the estimated effects from comparing students from different socioeconomic backgrounds. This was expected since we now compare a group which were incentivized to decrease earnings (AID_{Low}) with a group which were incentivized to increase earnings (AID_{High}).

[22]Column (3) is related to the difference of columns (1) and (2) in that they would be equivalent if we, in the former specification, also included an interaction variable between the student socioeconomic background indicator and each included control variable.

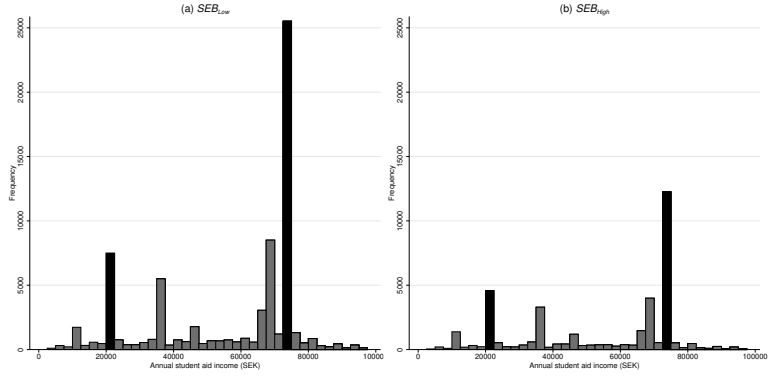
Table 4.2. *Estimated relative changes in annual earnings and study pace for students with different pre-reform student aid incomes, by student socioeconomic background*

	SEB_{low} (1)	SEB_{high} (2)	$SEB_{high} - SEB_{low}$ (3)
Panel A. Earnings			
$R \times AID_{high}$	5.525***	5.127***	-1.624
	(1.037)	(1.102)	(1.430)
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Study trends	✓	✓	✓
Panel B. Study pace			
$R \times AID_{high}$	-1.893***	-1.220***	0.943***
	(0.229)	(0.284)	(0.349)
Controls	✓	✓	✓
Fixed effects	✓	✓	✓
Study trends	✓	✓	✓
Observations	102,933	60,380	163,313

NOTE.—The table reports the estimated β_3 coefficients from estimation of the regression model (3.1) in the empirical section separately by student socioeconomic background. The AID_{Low} group is defined as having student income corresponding to only receiving student grants (left black bin in Figure 4.7) and the AID_{High} group is defined as receiving student income corresponding to both grants and the maximum allowable level of loans (right black bin in Figure 4.7). Columns (1) and (2) pertains to estimating the model separately for students with a high and a low socioeconomic background, respectively, while column (3) includes the full estimation sample in a "triple-difference" design. Controls includes age of first registration, gender and immigrant dummies, field of study, local unemployment rates, high school GPA, fixed cohort, study year and college effects and study year trends interacted with the student socioeconomic background indicator variable. SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. Incomes are measured in SEK (1 SEK corresponded to approximately € 0.1 in 2001). Academic credits are measured on the calendar year and one credit corresponds to one week of full-time studies. Robust standard errors in parentheses. *, ** and *** denote significance at the 10, 5 and 1 percent levels. For variable definitions, see Table 3.1

If the share of students from a low socioeconomic background who had both grants and loans were relatively greater than the corresponding share of students from a high socioeconomic background, we would suspect that some of the effect reported in Table 4.1 could stem from compositional effects. To investigate this hypothesis, Table 4.2 is complemented by Figure 4.8 and Table 4.3, which shows the relative shares of students with a high and a low student aid income by student socioeconomic background, respectively. In particular, from Table 4.3 we see that there is a slightly larger share of students having both grants and student loans in the low socioeconomic student group. Hence, from this analysis we cannot reject that part of the effect of the student aid reform is due to relative differences in the student aid composition of students from different socioeconomic backgrounds.

Figure 4.8. Student aid distribution of students in their second year of study, by student socioeconomic background



NOTE.—The figure illustrates the distribution of student aid income for students in their second year of study separately for students from a high and a low socioeconomic background. The black bins indicate the two student aid income groups having only grants (left bin) corresponding to a student aid income of between 20,000-22,500 SEK p.a. and maximum loans and grants (right bin) corresponding to a student aid income of between 70,000-72,500 SEK p.a. SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. Incomes are measured in SEK (1 SEK corresponded to approximately € 0.1 in 2001).

Table 4.3. *Students with high and low pre-reform student aid incomes, by student socioeconomic background*

	Number of students			
	AID_{high}	AID_{low}	Total	Share AID_{high}
SEB_{Low}	25,514	7,485	32,999	0.77
SEB_{High}	12,254	4,585	16,839	0.72
Total/Difference	37,762	12,070	49,838	0.05

NOTE.— The table reports the number and share of students receiving only grants corresponding to a student aid income of between 20,000-22,500 SEK p.a. and maximum loans and grants corresponding to a student aid income of between 70,000-72,500 SEK p.a., for students from a high and a low socioeconomic background respectively. The AID_{Low} group is defined as having student income corresponding to receiving only student grants (left black bin in Figure 4.7) and The AID_{High} group is defined as having student income corresponding to having both grants and maximum loans (right black bin in Figure 4.7). SEB_{Low} is defined as college students where neither parent has a post-secondary education and SEB_{High} as college students where both parents have a post-secondary education. Incomes are measured in SEK (1 SEK corresponded to approximately € 0.1 in 2001).

Next, we have examined the sensitivity of our results with respect to model specification. Specifically, we re-estimated the model including cohort-specific, rather than socioeconomic group-specific, study progression trends. From exploring the study progression trends in Figure 4.4 and Figure 4.5, we expected that leaving out the student socioeconomic trends would lead to an upward bias of the estimated parameters. We find, in line with this expectation, that the magnitude of the estimated effect is increased when we substitute the socioeconomic group-specific for the cohort-specific study trends (not reported).

Moreover, we have also estimated the model by including individual, instead of cohort, fixed effects with qualitatively unchanged results (not reported).

Finally, we investigated the robustness of the results with respect to the classification of the student socioeconomic background groups. In particular, we included a third student group in the estimations, consisting of college students who had one parent with post-secondary education. Interestingly, the estimated results for this student group on earnings and study pace generally falls in-between the low and high socioeconomic background groups. This was expected if parental education is a good proxy for dependency on the public student aid system. Furthermore, we also redefined the student background groups using parental taxable income rather than education level. As income and level of education are highly correlated in the data, it was no surprise that this definition gave qualitatively similar results and is, for this reason, not reported here.

5 Summary and concluding remarks

This paper exploits a comprehensive Swedish reform in the public student financial aid system to assess the role of the design of the student financial aid system for student academic achievement. While keeping the total amount of student aid unchanged, the reform incorporated three major changes: *i)* the grant component of the total aid increased at the expense of the loan component, *ii)* the exempt amount on earnings while studying was significantly increased, and *iii)* the rules for the repayment of the student loans were significantly tightened. These changes gave college students an incentive to increase their earnings and to decrease their student loans by reallocating time from studying to working, potentially affecting their study pace. Given recent reports of dramatically declining hours of studying among college students, it is an interesting and relevant question for education policy to investigate whether an increased student labor supply could crowd out time for studies.

To empirically test the time reallocation hypothesis, we estimate the relative change in earnings and study pace attributed to the policy change for students from a high and a low socioeconomic background, respectively, defined by the students' parental educational attainment. We argue, on both theoretical and empirical grounds, that students from a lower socioeconomic background are more dependent on — and hence more sensitive to changes in — the public student aid system, implying that any effects from the intervention should be relatively greater in magnitude for these students than for students from a higher socioeconomic background.

Using longitudinal administrative data on on earnings and academic achievements, we sample and follow program students enrolling prior to the intervention over their course of study. Our results show that relative earnings increased with approximately 25 percent for students from a low socioeco-

nomic background as a consequence of the student aid reform. Furthermore, we find that this change coincided with a ten percent reduction in the relative study pace for the same students. In other words, the changes in the financial student aid system increased relative times-to-graduation for students from a low socioeconomic background. This result is well in line with the empirical expectations of the time-reallocation hypothesis. We also perform a number of robustness checks to evaluate the stability of the inferences we make. The results from these exercises generally strengthen the interpretation of our main findings.

In conclusion, long study durations are likely to create considerable social and individual costs. One of the primary objectives of the Swedish student financial aid system is to reduce the cost of higher education for individuals from a lower socioeconomic background. However, relative to individuals from a higher socioeconomic background, the student aid reform of 2001 rather appears to have increased the cost of higher education for these students. Even though this increased cost may have been later counteracted by an improved labor market attachment among these students, the results obtained in this paper indicate that both the latter, intended, and the former, spillover, effect should be taken into account when performing cost-benefit analyses of student financial aid policies.

References

BABCOCK, P. and MARKS, M. (2010). *Leisure College, USA: The Decline in Student Study Time*. Education Outlook 7, American Enterprise Institute for Public Policy Research, www.aei.org.

— and — (2011). The Falling Time Cost of College: Evidence From Half a Century of Time Use Data. *The Review of Economics and Statistics*, **93** (2), 468–478.

BAUMGARTNER, H. J. and STEINER, V. (2006). *Does More Generous Student Aid Increase Enrolment Rates into Higher Education? Evaluating the German Student Aid Reform of 2001*. Discussion Paper 2034, IZA Bonn.

BECKER, G. (1993). *Human Capital – A Theoretical and Empirical Analysis with Special Reference to Education*, vol. 3. Chicago: The University of Chicago Press.

BETTINGER, E. (2004). How Financial Aid Affects Persistence. In C. M. Hoxby (ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It*, The University of Chicago Press, pp. 207 – 38, National Bureau of Economic Research.

BILL NO.1999/2000:10 (1999). *Ett Reformerat Studiestödssystem*. Government bill, http://www.regeringen.se/sb/d/108/a/2751.

BOSTON GLOBE (2010). What Happened to Studying? Keith O'Brien, Available: http://www.boston.com/bostonglobe/ideas/articles/2010/07/04/what_happened_to_studying [Last accessed: 24 October 2013].

BOUND, J., LOVENHEIM, M. and TURNER, S. (2012). Increasing Time to Baccalaureate Degree in the United States. *Education Finance and Policy*, **7** (4), 375–424, MIT Press.

BRODATY, T. O., GARY-BOBO, R. J. and PRIETO, A. (2008). *Does Speed Signal Ability? The Impact of Grade Repetitions on Employment and Wages*. Discussion paper 6832, CEPR.

BRUNELLO, G., LUCIFORA, C. and WINTER-EBNER, R. (2003). The Wage Expectation of European Business and Economics Students. *Journal of Human Resources*, **39** (4), 1116–42.

CAMERON, S. and TABER, C. (2004). Estimation of Educational Borrowing Constraints Using Returns to Schooling. *Journal of Political Economy*, **39** (4), 132–82.

CARD, D. (1999). Causal effects of Education on Earnings. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics 3A*, Amsterdam: Elsevier Science., pp. 1801–1863.

DYNARSKI, S. (2002). The Behavioral and Distributional Implications of Aid for College. *American Economic Review*, **92** (2), 279–285.

— (2003). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review*, **93** (1), 279–288.

ECKSTEIN, Z. and WOPLIN, K. (1999). Why Youth Drop out of High School: The Impact of Preferences, Opportunities and Abilities. *Econometrica*, **67** (6), 1295–1339.

EHRENBERG, R. and SHERMAN, D. (1987). Employment While in College, Academic Achievment, and Postcollege Outcomes: A Summary of Results. *Journal of Human Resources*, **22** (1), 1–23.
FREDRIKSSON, P. (1997). Economic Incentives and the Demand for Higher Education. *Scandinavian Journal of Economics*, **9** (1), 129–142.
GEEL, R. and BACKES-GELLNER, U. (2012). Earning While Learning: When and How Student Employment is Beneficial. *LABOUR*, **26** (3), 313–40.
HÄKKINEN, I. (2006). Working while Enrolled in University: Does it Pay? *Labor Economics*, **13** (2), 167–189, Elsevier.
HÄKKINEN, I. and UUSITALO, R. (2003). *The Effect of a Student Aid Reform on Graduation: A Duration Analysis*. Working Paper 2003:8, Uppsala University, Department of Economics.
HAMMARSTRÖM, E. (1996). *Varför inte Högskola? En Longitudinell Studie av Olika Faktorers Betydelse för Studiebegåvade Ungdomars Utbildningskarriär*. Göteborg: Acta Universitatis Gothoburgensis, Göteborg Studies in Educational Sciences 107.
HOENACK, S. (1971). The Efficient Allocation of Subsidies to Higher Education. *American Economic Review*, **61** (3), 302–311.
HOLMLUND, B., LIU, Q. and NORDSTROM-SKANS, O. (2008). Mind the Gap? Estimating the Effects of Postponing Higher Education. *Oxford Economic Papers*, **60** (4), 683–710, Oxford University Press.
HOTZ, V. J., XU, L. C., TIENDA, M. and AHITUV, A. (2002). Are There Returns to the Wages of Young Men From Working While in School? *Review of Economics and Statistics*, **84** (2), 343–75.
KANE, T. J. (1994). College Entry by Blacks Since 1970: The Role of College Costs, Family Background, and the Returns to Education. *The Journal of Political Economy*, **102** (5), 878–911.
VAN DER KLAAUW, W. (2002). A Regression-Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment. *International Economic Review*, **43** (4), 1249–1287.
LIGHT, A. (2001). In-school Work Experience and the Return to Schooling. *Journal of Labor Economics*, **19** (1), 65–93.
LINSENMEIER, D., ROSEN, H. and ROUSE, C. E. (2006). Financial Aid Packages and College Enrollment Decisions: An Econometric Case Study. *Review of Economics and Statistics*, **88** (1), 126–145.
MANSKI, C. and WISE, D. (1983). *College Choice in America*. Harvard University Press, Cambridge.
MCPHERSON, M. S. and SHAPIRO, M. O. (1991a). Does Student Aid Affect College Enrollment? New Evidence on a Persistent Controversy. *American Economic Review*, **81** (1), 309–318.
— and — (1991b). *Keeping College Affordable: Governance and Educational Opportunity*. Washington, DC: The Brookings Institution.
MOORE, R. L., STUDENMUND, A. H. and SLOBKO, T. (1991). The Effect of the Financial Aid Package on the Choice of a Selective College. *Economics of Education Review*, **10** (4), 311–321.
NATIONAL AGENCY FOR HIGHER EDUCATION (HSV) (2001). *Swedish Universities and University Colleges*. Annual report, http://www.hsv.se.

NATIONAL AGENCY FOR HIGHER EDUCATION (HSV) (2004). *Swedish Higher Education. A Survey 1977-2000*. Tech. rep., http://www.hsv.se.
NATIONAL AGENCY FOR HIGHER EDUCATION (HSV) (2006). *OECD Thematic Review of Teritary Education, Country Background Report for Sweden*. Tech. rep., http://www.hsv.se.
NATIONAL AGENCY FOR HIGHER EDUCATION (HSV) (2007). *The Swedish Higher Education System*. Tech. rep., http://www.hsv.se.
NATIONAL AGENCY FOR HIGHER EDUCATION (HSV) (2010). *Swedish Universities and University Colleges*. Annual report, http://www.hsv.se.
NATIONAL BOARD FOR STUDENT AID (CSN) (2007). *The Swedish Higher Education System*. Tech. rep., http://www.csn.se.
NIELSEN ARENDT, J. (2013). The Effect of Public Financial Aid on Dropout from and Completion of University Education: Evidence from a Student Grant Reform. *Empirical Economics*.
OECD (2005). *Education at a Glance 2005: OECD Indicators*. Tech. rep., OECD Publishing, http://dx.doi.org/10.1787/eag-2005-eg.
— (2008). *Education at a Glance 2008: OECD Indicators*. Tech. rep., OECD Publishing, http://dx.doi.org/10.1787/eag-2008-eg.
REUTERBERG, S.-E. and SVENSSON, A. (1994). Financial Aid and Recruitment to Higher Education in Sweden: Changes between 1970 and 1990. *Studies in Higher Education*, **19**, 33–45.
ROUSE, C. E. (1994). What to Do After High School: The Two-Year Versus Four-Year College Enrollment Decision. In R. G. Ehrenberg (ed.), *Choices and Consequences: Contemporary Policy Issues in Education*, Ithaca, NY: ILR Press.
SCHRØTER JOENSEN, J. (2010). *Timing and Incentives: Impacts of Student Aid on Academic Achievement*. Meeting Papers 823, Society for Economic Dynamics.
SKYT NIELSEN, H., SØRENSEN, T. and TABER, C. (2010). Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform. *American Economic Journal: Economic Policy*, **2** (2), 185–215.
STINEBRICKNER, R. and STINEBRICKNER, T. R. (2003). Working During School and Academic Performance. *Journal of Labor Economics*, **21** (2), 473–91.