Essays in Empirical Health Economics

Inauguraldissertation

zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2017

vorgelegt von

Sebastian Wichert

Referent:	Prof. Dr. Joachim Winter
Korreferent:	Prof. Dr. Oliver Falck
Promotionsabschlussberatung:	31. Januar 2018

Datum der mündlichen Prüfung: 15. Januar 2018

Namen der Berichterstatter: Joachim Winter, Oliver Falck, Amelie Wuppermann

To my parents &
in memory of my grandparents

Acknowledgements

First and foremost I would like to thank my supervisor, Joachim Winter. He sparked my interest in empirical economics and always had an open door for me at his chair. He showed me what it means to be an academic teacher and an applied researcher. I am very grateful for his constant support and for giving me the opportunity to spend several months at the University of Essex.
I am also deeply indebted to Oliver Falck for agreeing to serve as second supervisor on my thesis committee and his constant advice. I want to thank both of them for their faith which enables me to continue my academic journey. I also thank Amelie Wuppermann who kindly agreed to serve as the third reviewer on my dissertation committee.

In addition, I am indebted to many people for their help with writing this dissertation. In particular, I would like to thank my co-authors, Amelie Wuppermann, Nadine Geiger and Simon Reif for their collaboration that has been and continues to be a source of inspiration and motivation. During the course of writing this dissertation, I benefited greatly from inspiring discussions with my colleagues and friends. I particularly would like to thank all my former colleagues of the *Chair of Empirical Economic Research at LMU* and my new colleagues at the *ifo Center for Industrial Organisation and New Technologies* for their professional help, insightful feedback and moral support. I would further like to thank the department of economics at LMU for financial support and the university archives staff for their help with the historical medical files.

My greatest gratitude goes to my parents and my grandparents for their constant guidance and their support throughout my life. Without them this dissertation would not have been possible. This work is dedicated to them.

Contents

List of Figures

List of Tables

Preface

When it comes to the connection of health and wealth, there is a heated, often uninformed debate in the media, politics and the public in general. Ample research in various fields of science tries to inform this discussion and provide guidance. Economists and health economists in particular contribute their share in several ways. First, they have highlighted "the dual relation between health and economic status" (Smith 1999, p.145; Deaton 2013). More specifically, they have formalized this association by separating and analysing causes for and consequences of an individual's health theoretically as well as empirically (Grossman 1972; Currie 2011). Second, economists are able to evaluate potential causal pathways empirically by using microeconometric methods based on quasi-experimental settings, even in absence of randomized controlled trials. For example, building on an earlier medical literature and several such settings, it has been documented that (shocks to) the prenatal environment in the womb and the perinatal environment at birth can have short- and long-term consequences on various life domains for a newborn (see e.g. Barker 1990; Almond and Currie 2011; Almond et al. 2010). Chapter 2 and 3 of this dissertation add to this literature. Another ongoing public discussion is about the potential health effects of modern leisure technology. The international evidence so far is mixed and any well-identified studies on this topic for Germany are virtually non-existent. Chapter 1 tries to fill this gap.

Summing up and giving an outlook on what follows, this dissertation provides insights into how access to modern technology at home, institutional settings in hospitals and external shocks during pregnancy affect various aspects of health. To be precise, the first chapter focuses on the effect of broadband Internet access on the health status of various demographic groups. The following chapter investigates the consequences of reimbursement relevant birth weight thresholds, which may coincide with medical diagnostic thresholds, for treatment decisions and the health outcomes of newborns in German hospitals. The last chapter examines how and why health at birth in hospital was affected by the onset of World War II. The following paragraphs give an introduction to each chapter before a comparison of their similarities and differences ends this preface.

Chapter 1, the essay *"The Internet and its impact on health - Evidence from Germany"*, examines whether broadband Internet access affects an individual's physical health status and mental well-being. Economists have dealt with the Internet from very early on and were aware that Internet connection speed is an important determinant of actual Internet dissemination and usage (see e.g. MacKie-Mason and Varian 1994). More recently, well-identified studies were conducted to evaluate the causal effects of broadband Internet on various life domains in diverse institutional and technological settings, highlighting several causal mechanisms and using different identification strategies. Surprisingly, potential health effects of broadband Internet access or actual broadband Internet use have been widely neglected in the economics and the medical literature so far.[1]

There are several potential mechanisms through which Internet use in general may affect different domains of physical health and mental well-being. A first potential mechanism is health information on the Internet. This may either enhance or deteriorate health, depending on the quality of the health information acquired and on whether individuals use it as a complement or a substitute for qualified health advice by trained physicians. Second, there is a communication and peer group channel. The Internet facilitates communication through emails, instant messaging or video chats, which may either be used to foster social interactions or may lead to isolation because real-life relationships are neglected in favour for online acquaintances (see e.g. Putnam 2001).[2] Third, there is an addiction channel which clearly has an adverse effect on physical as well as on mental health. Internet use and computer gaming are classified as potential risk factors to psychological well-being with a high addictive effect, especially for young men, children and adolescents (see RKI 2014; RKI 2015). Furthermore, the World Health Organisation warned already in 2000 that increased computer use, which today is almost synonymous with Internet use, in the developed world promotes physical inactive behaviour, which in turn is a key driver of rising obesity rates around the world (WHO 2000, p.113). Interestingly, these warnings of "adverse Internet health effects" mostly rely on observational studies without a proper econometric identification strategy, thus providing only correlational evidence. The present study tries to

[1] To the best of my knowledge, DiNardi et al. (2017) and the present study are the only exceptions.

[2] Bauernschuster et al. (2014) investigate the effect of broadband Internet access on social capital. However, they do not look at potential health consequences.

bridge this gap by evaluating potential causal effects of *broadband Internet access* on an individual's health status. Using the Socio-Economic Panel (SOEP), the most comprehensive German social science data set, I consider weight variables (Body-Mass-Index, overweight and obesity) and measures of life and health satisfaction in my analysis as health status outcomes. Information on household-level broadband Internet access serves my main explanatory variable. To evaluate a potential effect heterogeneity, I split the sample further into four demographic groups by gender and age. Aguiar et al. (2017) find that only men aged 30 and younger spend a greater share of their leisure time for recreational computer use in recent years. Therefore, I expect negative health effects of broadband Internet to be more pronounced for younger cohorts and males. The association of high-speed Internet access and the health status variables points into this direction. OLS results (controlling for several county- and individual-level characteristics and household fixed effects) show a significant and sizeable weight gain for men and cohorts aged 30 and younger who have broadband Internet access, but there is no effect on satisfaction levels for these groups. Neither for women nor for older cohorts aged 30+ I find any significant effects. Despite using controls, there still might be endogeneity issues. Hence, in absence of a randomized controlled trial, one needs a quasi-experimental identification strategy to generate exogenous variation in broadband Internet access and thus identify a causal effect on health outcomes. Technical broadband Internet availability in an area can be used to generate a strong instrumental variable for actual broadband Internet use in Germany as the previous literature in different contexts has shown (see e.g. Falck et al. 2014a). The first stage of the instrumental variable analysis shows that the instrument has high predictive power. Nevertheless, all second-stage coefficients are insignificant. Hence, I cannot rule out that broadband Internet access has no effect on the health status of an individual in Germany.

In Chapter 2, titled *"Is it good to be too light? Consequences of birth weight thresholds in hospital reimbursement systems"*, my co-authors and I analyse how reimbursement relevant low birth weight thresholds trigger different quantity and quality of care for newborns above and below these thresholds. In per-case hospital reimbursement systems, in which hospitals receive more money for otherwise equal newborns with birth weight just below compared to just above specific birth weight thresholds, there is a strong incentive to under-report birth weight. This phenomenon is called "upcoding" and has been documented in several countries around the world (e.g. Jürges and Köberlein 2015; Shigeoka and Fushimi 2014). Related literature has found that in hospital systems, in which no financial incentives for birth weight manipulation are in place, diagnostic birth weight thresholds in medical guidelines can benefit light newborns. More specifically, despite the fact that low birth weight is typically linked to worse health outcomes (e.g. Hack et al. 2002; Hummer et al. 2014) newborns with weight just below the 1500g diagnostic threshold have been found to have higher survival chances than newborns with weight just above (e.g. Almond et al. 2010; Bharadwaj et al. 2013; Breining et al. 2015). We bring these two strands of the literature together and explore whether newborns benefit from having a reported birth weight just below a birth weight threshold in the German hospital system where reimbursement relevant and medical birth weight thresholds can coincide. In such a setting birth weight manipulation may be beneficial for the newborn, if hospitals provide additional care to these newborns due to the increase in revenue.

However the fact that the diagnostic thresholds are relevant for reimbursement in our setting imposes a challenge to the empirical analysis: As Jürges and Köberlein (2015) show, hospitals primarily tend to upcode newborns for whom staff expects more care. These are relatively fragile newborns that still have non-negligible survival chances and will therefore receive a lot of treatment. Hence newborns on different sides of a birth weight threshold may differ in potentially unobserved ways, which would prevent one from any meaningful comparison. In our empirical analysis which is based on an administrative hospital claims data set, covering the universe of hospital births in Germany in the years 2005–2011, we take this into account in three ways. First, we control for a large set of variables which cannot be manipulated and which capture a newborn's health at birth (e.g. sex, plurality of births). Second, we exclude fragile newborns for whom hospital staff may have expected an early death and for which the hospital staff has no incentive to upcode, namely newborns who die on their first day of their life. Third, we take possible differences in coding practices into account by including hospital fixed effects in our estimations.

We find that for all but the highest two thresholds (2000g and 2500g), our results are insignificant in this most extensive specification. We interpret this as evidence that neither reimbursement differences, nor the diagnostic threshold of 1500g or thresholds in medical guidelines trigger additional care or reduce mortality among newborns in Germany. While our results may indicate that newborns with weight just below 2000g and 2500g benefit from being below these thresholds, we believe that the differences rather reflect systematic upcoding related to factors that we cannot control for perfectly.

In Chapter 3 of this dissertation, *"Birth in Times of War - An investigation of health, mortality and social class using historical clinical records"*, my co-author and I examine how the onset of the World War II (WWII) affects health at birth and infant mortality. Furthermore, we explore the effect heterogeneity with respect to parental social status and we analyse possible causal mechanisms. This project is related to a large and established literature that has identified the time in utero as the most critical time period in life (Almond and Currie 2011; Barker 1990). Moreover these studies show that adverse early-life events have long-run effects on later-life outcomes, such as worse health status, lower educational attainment, occupational status and even human behaviour. To establish a causal link between adverse early-life circumstances and later-life outcomes, several studies use WWII, by far the greatest shock that has affected living cohorts in Western Europe, as a natural experiment (see e.g. Atella et al. 2016; Van den Berg et al. 2016; Jürges 2013; Kesternich et al. 2014; Kesternich et al. 2015). This literature leaves several questions unanswered. As historical individual-level data on birth outcomes is hardly available, it is unclear whether the negative effects of WWII remained latent until later-life or were already present at time of birth. Furthermore, these findings are based on samples of the surviving population, which may bias any long-term estimates if individuals who survive infancy during the war do systematically differ from survivors of other cohorts. Moreover, most studies investigate only very extreme and in our days very rare events related to WWII, namely famines and hunger, bombing and combat, or dispossession.[3]

To bridge this gap and estimate the short-term effects of WWII on infant health outcomes, we collected an entirely new data set of historical clinical birth records from the largest birth hospital in Munich, Germany. Our unique data set contains around 10,000 births and miscarriages which took place in hospital between December 1937 and September 1941. In our empirical strategy we exploit the unexpected onset of WWII as natural experiment. We focus on the first two years of WWII, a period when military operations took place outside of Germany and there was no nutritional shortage. Our main contribution is to document an effect of WWII on perinatal child mortality even in the absence of extreme conditions. The onset of WWII acted as shock to individuals and the public health system, which initially led to a jump in perinatal mortality and then gradually faded out. This interpretation is consistent with historical evidence, showing that the onset of the war caused turmoil in the health system and disrupted daily life. Two mechanisms are potentially driving our results. First,

[3] Lindeboom and Van Ewijk (2015) and Van Ewijk and Lindeboom (2016) are a rare exception.

maternal stress levels are likely to be increased due to drafting of husbands and a high level of uncertainty that comes along with the war. These higher stress levels should lead to a decrease in birth weight (Currie and Rossin-Slater 2013; Torche 2011). Second, the large-scale drafting of doctors can lead to a decrease in medical quality, which could put a newborn's life at risk. Our results indicate that there is no effect of the onset of the war on birth weight, but we find a large increase in mortality rates for live born children, whose health is at least partially under the control of physicians and hence may suffer from their absence. Hence we conclude the decline in medical quality to be the more important mechanism. Our results have important implications for the literature on long-term effects of WWII. We document a disproportional increase in mortality for very low birth weight infants, suggesting that studies using samples of the surviving population provide a lower bound for the true effect.

Each chapter of this dissertation is the result of a separate research project, which cover a broad range of topics, and therefore can be read independently. Furthermore, these research projects differ among many lines. For example, each chapter is based on different data sources that imply different data access methods. For Chapter 3 my coauthor and I saved, with great support from the LMU university archives staff, molded historical medical files from decay, before we were able to digitize any data for our research project.[4] Chapter 2 is based on administrative data, namely the universe of hospital reimbursement claims in Germany from 2005-2011, whereas Chapter 1 relies on georeferenced survey data combined with technical and cartographic information. For both projects it was necessary to work with an external research data center - either in person or with controlled remote access. Furthermore, the observational periods of the chapters are different. While Chapter 3 clearly has a historical focus, the other chapters cover present day topics. Last, the units of observation vary greatly: Newborns in Chapters 2 + 3 and adults in 1; one Munich hospital in 3 vs. all German hospitals in 2; patients and doctors in 2 + 3 and the general public in 1.
However, this is not an arbitrary collection of research papers. They share key characteristics and are the result of a common research approach, already defined in the title of this dissertation. First, all chapters try to find out whether and how a (subjective or objective) measure of health status is affected - either by institutional incentives in hospitals, new technologies or sudden declines in the quality of hospital care. This

[4] This initial effort has a long lasting impact beyond our project. Complete collections of historical medical files covering several hospitals in Munich and more than 80 years can now be accessed at the LMU university archives (see http://www.universitaetsarchiv.uni-muenchen.de/index.html [last accessed: 17 September 2017]).

means that all chapters in this dissertation focus on health as the outcome, and not as an input. Second, all chapters use German data and institutional knowledge to tackle the respective research question for Germany. In some cases, research on similar topics already exists. However, these papers refer to other countries and often lack external validity due to widely differing institutional contexts around the world. This dissertation shall help to overcome this gap. Finally and most importantly, all chapters are written with a clearly defined causal research question in mind. The combination of modern microeconometric methods, a quasi-experimental setting and diverse data sources enable me to tackle these causal questions empirically and evidence-based.

Chapter 1

The Internet and its impact on health - Evidence from Germany

1.1 Introduction

As early as 1995, six years after its "publication" and long before it became the public good it is today, Bill Gates, the co-founder of Microsoft, identified the Internet as a "tidal wave" that "will have fundamental impact on work, learning and play".[1] History (and the web address in the previous footnote) show that he was right and that the Internet has transformed almost all parts of our lives. Economists have dealt with the Internet from very early on and were aware that Internet connection speed is an important determinant of actual Internet dissemination and usage (see MacKie-Mason and Varian 1994). However, only recently well-identified studies were conducted to evaluate the causal effects of broadband Internet on *various life domains* in *diverse institutional and technological settings*, highlighting *several causal mechanisms*, and using *different identification strategies*, a point which will be elaborated in more detail in Section 1.2. Interestingly, none of these published studies has looked at potential direct or indirect health effects of broadband Internet, even though the public, the media, and medical and psychological professions have recognized that the Internet can affect physical and mental health outcomes. For example, "Internet addiction" has recently been recognized as an illness by the American Psychiatric Association and the Robert-Koch-Institute (RKI), the official body for federal health reporting in Germany, identifies high electronic media consumption as a potential health risk (RKI 2015).

[1] https://www.justice.gov/sites/default/files/atr/legacy/2006/03/03/20.pdf [last accessed: 17 September 2017].

This chapter tries to bridge this gap by estimating the effect of broadband Internet access on health status outcomes. The database for this research project is the Socio-Economic Panel (SOEP), the most comprehensive German social science data set, which consistently covers various health status measures. In particular the weight and height variables of the SOEP are relevant for this study, because they allow a calculation of several measures of overweight, a serious risk factor to public health all over the world. Moreover, I include life and health satisfaction variables as measures of mental well-being in my analysis and use information on household-level broadband Internet access as my main explanatory variable. To evaluate the effect heterogeneity, I split the sample further into demographic groups. I evaluate the following five demographic groups: all individuals, men, women, individuals at most 30 years of age and individuals older than 30 years in the main analysis. These age thresholds separate so-called "digital natives", who already were exposed to computers and the Internet in their youth or early adulthood and therefore are more likely to be early adopters of broadband Internet access.

I begin my empirical investigation with the association of high-speed Internet access and the health status variables using OLS regressions and controlling for a rich set of individual- and household-level characteristics. Despite using controls, there still might be endogeneity issues, namely omitted variables bias and/or reverse causality. Sicker individuals could be more likely to have a broadband subscription because it facilitates communication without leaving your home, which would bias any estimated effect of high-speed Internet on health. Therefore, I use a quasi-experimental identification strategy to generate exogenous variation in broadband Internet access in a household which has been used in the economics literature in other contexts before (see Falck et al. 2014a). In Germany, broadband Internet access via Digital Subscriber Line (DSL) technology, almost the only available high-speed Internet option in 2008, relies on the preexisting telephone network. Only if the wire distance between a household and its telephone/Internet node which is in turn connected to the nationwide telephone/Internet network, is less than 4.2 kilometers, a DSL Internet connection is technically feasible. I calculate these distances using a detailed map of the German telephone catchment areas including Internet nodes and a georeferenced SOEP data set to generate a binary (DSL availability) instrumental variable for the actual DSL subscriptions in a household.

The OLS results show a significant and sizeable weight gain for men and cohorts aged 30 and younger with DSL access, but there is no "DSL effect" on satisfaction levels for these groups. Neither for women nor for older cohorts aged 30+ I find any signif-

icant effects. Regard to the IV estimations, the first-stage regressions show that the instrument has high predictive power since living outside the 4.2 km radius around an Internet node is associated with significantly lower DSL subscription rates. Nevertheless, all second-stage IV-coefficients are insignificant for all but one variable. Hence, I cannot rule out that broadband Internet access has no effect on the health status of an individual in Germany.
The rest of this chapter is organized in the following way: In Section 1.2 I give a short review of the literature related to the effects of broadband Internet. The data is described in Section 1.3. Section 1.4 explains the empirical methodology in combination with the underlying institutional and technological background which is used for identification. Results are presented in Section 1.5. Section 1.6 contains concluding remarks.

1.2 Literature review

The empirical economic literature examining the causal effects of broadband Internet with well-defined econometric identification strategies is already large and expanding rapidly. These studies cover a *broad range of outcome variables*, have *diverse institutional and technological backgrounds*, highlights *several causal mechanisms*, use *different identification strategies* based on *multiple data sources*. Bertschek et al. (2015) provide an overview of the literature on the economic effects of broadband Internet on economic growth, employment and regional development, as well as productivity and firm performance.[2] Table 1.1 gives an overview of the recent broadband literature in several other fields of economics which are related in one or another way to my research topic. All studies use a quasi-experimental design which provides exogenous variation in broadband Internet availability/penetration to identify causal effects on the outcome of interest and not just a mere correlation.[3] More specifically, detailed geographical information is the foundation of all identification strategies used, which is either the number of broadband Internet providers in an area (Guldi and Herbst 2017), the distance from a household to an Internet node (i.a. this study) or the geo-

[2] Some more recent publications include Briglauer et al. (2016), Canzian et al. (2015), Dettling (2017), and Falck (2017) and further sources mentioned therein. DellaVigna and Ferrara (2015) summarizes economic and social impacts of the media more broadly.

[3] Broadband Internet availability is used in other contexts as well. For example, researchers in industrial organisation use broadband Internet dissemination as an outcome variable and try to investigate potential drivers, barriers or regulatory questions (see e.g. Lange 2017; Nardotto et al. 2015). Moreover Falck et al. (2016) use broadband Internet availability as an instrument for ICT skills to determine its effects on later earnings in an international comparison as well as within Germany.

Table 1.1: Effects of broadband Internet - Literature review

Autors	Exogenous variation in broadband Internet access induced by	Identification strategy	Outcome	Country
Czernich (2012)	Technical availability	IV	Voting behaviour	Germany
Falck et al. (2014a)				
Campante et al. (forthcoming)	Technical availability	IV	Voting behaviour	Italy
Gavazza et al. (2017)	Technical availability	IV	Voting behaviour, government expenditure	UK
Poy and Schüller (2016)	Roll-out	DID	Voting behaviour	Italy
Lelkes et al. (2017)	Regulatory availability	IV	Political polarisation	US
Hjort and Poulsen (2017)	Roll-out	DID	Employment	Africa
Gebhardt (2010)	Technical availability	IV	Retail competition	Germany
Ahlfeldt et al. (2017)	Technical availability	IV	Housing market	UK
Mang (2016)	Technical availability	IV	Housing prices, Local online service	Germany
Bauernschuster et al. (2014)	Technical availability	IV	Social capital	Germany
Bhuller et al. (2017)	Roll-out	IV	Print media	Norway
Bhuller et al. (2013)	Roll-out	IV	Sex crime	Germany
Nolte (2016)	Technical availability	IV	Crime	Germany
Billari et al. (2017b)	Technical availability	IV	Fertility	Germany
Guldi and Herbst (2017)	Number of providers	DID	Fertility	US
Bellou (2015)	Roll-out	IV	Marriage market	US
Billari et al. (2017a)	Technical availability	IV	Sleep	Germany
DiNardi et al. (2017)	Number of providers	DID	Health	US
Amaral-Garcia et al. (2017)	NA	NA	Health care choice	UK

graphical roll-out of Internet infrastructure (Hjort and Poulsen 2017). This shows that this literature seeks to identify the effects of broadband Internet *access*, not Internet *use*, even though they are highly correlated. From a normative perspective the results of these studies show that broadband Internet access has positive and negative effects on individuals, firms or regions. For example, Bauernschuster et al. (2014) show that broadband Internet access increases social capital (in Germany), whereas Bhuller et al. (2013) find for Norway that broadband Internet access leads to more sex crimes.

Most closely related to my research question in this chapter is the analysis of DiNardi et al. (2017), which is to the best of my knowledge the only other study evaluating health effects of broadband Internet.[4] They use the roll-out of broadband Internet providers in the US to identify causal effects of broadband coverage on physical health measures, namely body weight and health behaviours. They find that greater broadband Internet coverage leads to a significant weight gain, and an increase in overweight and obesity rates. These effects apply only for white women, neither for men nor other ethnic groups. This is surprising, because Aguiar et al. (2017) show for the US as well that only men aged 30 and younger reduced their working hours in the last 15 years and spend a greater fraction of their increased leisure time on recreational computer use due to innovations in leisure technology. Neither women nor older cohorts show similar patterns. Hence, one could expect that any weight gain is more pronounced for this subgroup, acknowledging that the effects of recreational Internet and computer *use* may differ from the impacts of broadband Internet *access*.

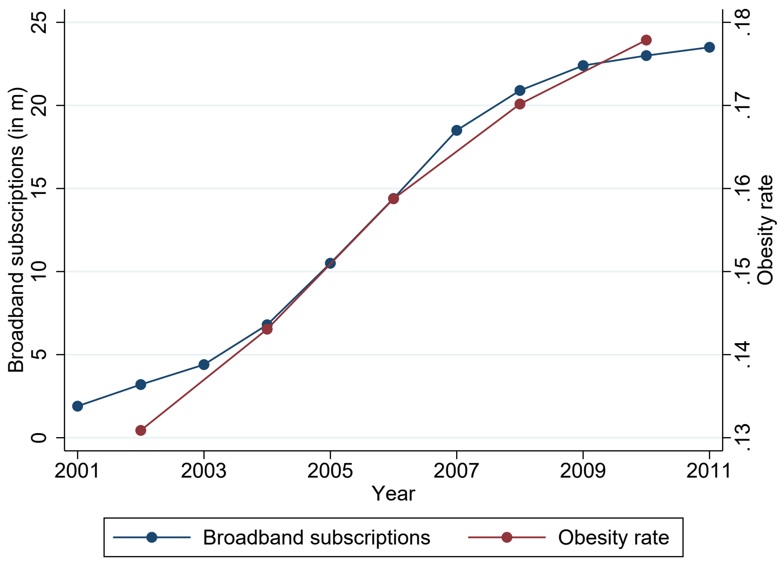
[4] The study of Amaral-Garcia et al. (2017) is still unpublished and therefore results are not available.

To evaluate the direction of potential health effects of broadband Internet, it makes sense to think about causal channels through which it may operate. First, the Internet provides access to health information. This may enhance or decrease health depending on the quality of the health information acquired and on whether individuals use it as a complement or a substitute for qualified health advice by trained physicians. Second, there is a communication and peer group channel. Broadband Internet facilitates communication through emails, instant messaging or video chats, which may be used to foster social interactions as shown by Falck et al. (2014a) or it may lead to isolation because real-life relationships are neglected in favour for online acquaintances. Third, there is an addiction channel which clearly has an adverse effect on physical as well as on mental health. The RKI (2014, pp.65) classifies Internet use and computer gaming as a potential risk factor to psychological well-being with a high addictive effect, especially for young men. Furthermore the World Health Organisation warned already in 2000 that increased computer use, which today is almost synonymous with Internet use in the developed world, promotes physical inactive behaviour which in turn is a key driver of rising obesity rates around the world (WHO 2000, p.113). A naïve look at Figure 1.1, which shows increasing trends over time for broadband subscriptions and obesity rates in Germany, seems to support this claim. However, in absence of a randomized control trial one has to exploit quasi-experimental settings to evaluate whether there really is a causal relationship between broadband Internet access and health status outcomes, not just a mere (spurious) correlation. The present study tries to do exactly that by instrumenting actual broadband Internet access with exogenous, geographically defined technical broadband Internet availability.

1.3 Data

The primary data source used for the empirical analysis is the German Socio-Economic Panel (SOEP), a representative, annually conducted household survey that is comparable to the Panel Study of Income Dynamics (PSID) in the US and the British Household Panel (BHP). All household members aged 16 and older are interviewed in person and answer questions covering virtually all parts of their life. Additional information about the household as a whole (e.g. household amenities etc.) and about any children aged 16 or younger living in the household are provided by the household head. The survey asks for "objective" facts (e.g. biography, education/qualification, time use) and "sub-

Figure 1.1: Obesity and Internet access



Notes: Share of overweight individuals and broadband Internet subscriptions 2001-2011.

Sources: Obesity rates calculated based on SOEP v32.1. Broadband Internet subscription rates taken from Bundesnetzagentur (2010) and Bundesnetzagentur (2017).

jective" assessments (attitudes, values, fears, satisfaction, opinions, perceptions).[5] Especially the following properties of the SOEP data help to tackle the research question of this chapter, namely if broadband Internet access and individual health are (causally) related: First, the SOEP entails detailed health information from all its participants biannually starting from 2002. In the main results in Section 1.4 I focus on the self-reported health status outcomes *Body-Mass-Index (BMI), overweight, obesity, health satisfaction and overall life satisfaction*.[6] BMI (for adults) is defined as weight (in kilograms) divided by height (in meters) squared. All (adult) individuals with a BMI of 25-30 are classified as overweight and as obese if they have a BMI larger than 30.[7] Health and life satisfaction are measured on an 11-point scale (0 [completely dissatisfied] to 10 [completely satisfied]) and are z-standardized for the empirical analysis. Second, data on slow Internet access (= Modem or Integrated Services Digital Network (ISDN) with a maximum speed of 128 kbit/s) in the household is available since 2000

[5] For an introduction to the German Socio-Economic Panel see Wagner et al. (2007) and for a comprehensive description of all variables see www.paneldata.org [last accessed: 17 September 2017].

[6] The evaluation of the effects of broadband Internet access on health behaviours, health care utilisation and health insurance choice is left for future research.

[7] The main analysis is conducted with the original weight variables/responses by the survey participants, not the imputed weight data since the imputation/correction method is not clearly documented. Nevertheless results with imputed data remain very similar.

in the SOEP, but on broadband Internet access only since 2008. The survey specifically asks for DSL connections (Digital Subscriber Line with a maximum speed of 6 Mbit/s), the most common high-speed Internet connection in Germany at that time (see Bundesnetzagentur 2010) which was well before fast, inexpensive and convenient mobile Internet access became available (see Falck et al. (2014a, p.67) for technical details). This setting allows me to study the effects of broadband Internet access.[8]

Third, the SOEP provides a vast amount of background characteristics which I use as control variables. More specifically, I control for marital status, number of children (age < 18) in the household, education, occupational status, migration background, flat/home ownership, log household net income, and survey month.

In 2008, about 20,000 individuals in 12,000 households were interviewed. Since the identification strategy presented in Section 1.4 will only work for West Germany, I restrict the estimation sample to all individuals living in West Germany in 2008 with (full) information on the aforementioned health variables, DSL Internet access in the household and the control variables. This leaves me with 12,873 observations (age 17 and older) in 2008. Table 1.2 presents summary statistics on these variables. 61% of all individuals already have broadband access at home. Younger cohorts and men are even more likely to have fast Internet. Almost 64% (17%) of all men and 44% (15%) of all women are considered as overweight (obese), which reflects the official statistics of the federal statistical office based on the German Microcensus quite well.

[8] Unfortunately Internet time use is only measured crudely in the SOEP, but Falck et al. (2014a) show that DSL access is positively correlated with Internet time use and that slow ISDN connection is not compensated with more Internet time use.

Table 1.2: Descriptive statistics

	Overall	Men	Women	Age $\leq$ 30	Age > 30
Outcome variables					
BMI	26.0	26.8	25.3	23.6	26.4
Overweight	0.54	0.64	0.44	0.29	0.58
Obese	0.17	0.19	0.15	0.078	0.19
Health satisfaction (std.)	-0.021	0.0024	-0.042	0.45	-0.11
Life satisfaction (std.)	0.057	0.062	0.053	0.17	0.036
Broadband Internet access in hh	0.61	0.64	0.58	0.74	0.58
Control variables					
Male	0.48	1	0	0.47	0.48
Marital status					
Married	0.64	0.67	0.62	0.16	0.73
Single	0.29	0.30	0.28	0.84	0.19
Widowed	0.067	0.032	0.100	0	0.080
Owner of Home/Flat	0.57	0.59	0.56	0.41	0.60
Occupational status					
Not employed	0.076	0.014	0.13	0.067	0.077
Apprentice	0.061	0.063	0.059	0.37	0.0028
Unemployed	0.037	0.037	0.037	0.052	0.034
Retired	0.26	0.26	0.26	0.00099	0.30
Blue collar worker	0.15	0.21	0.096	0.16	0.15
White collar worker	0.36	0.33	0.37	0.33	0.36
Entrepreneur	0.063	0.083	0.045	0.020	0.071
Schooling					
Lower secondary	0.37	0.37	0.36	0.19	0.40
Medium secondary	0.24	0.20	0.28	0.29	0.23
Higher secondary	0.11	0.10	0.12	0.28	0.081
Other	0.059	0.059	0.060	0.023	0.066
University of applied sciences	0.072	0.092	0.053	0.035	0.079
University	0.13	0.16	0.11	0.067	0.15
Pupil	0.019	0.020	0.018	0.12	0
Migration background					
None	0.79	0.79	0.78	0.73	0.80
Direct	0.13	0.13	0.13	0.10	0.13
Indirect	0.084	0.084	0.084	0.17	0.069
Age	49.8	49.8	49.7	24.1	54.5
# Children in hh	0.49	0.49	0.50	0.42	0.51
Net hh income p.m.	3040.9	3151.4	2939.3	2905.9	3066.1
Observations	12873	6166	6707	2027	10846

Notes: Samples: All individuals in 2008 in West-Germany.
Source: Own calculations based on the SOEP v32.1.

1.4 Empirical strategy

As mentioned before, I want to investigate whether broadband Internet access has an influence on the health status of an individual. A simple regression of health outcomes on broadband access in the household is prone to severe endogeneity problems, namely omitted variable bias. As noted previously by Falck et al. (2014a, p.2244) broadband Internet providers primarily install (expensive) DSL infrastructure and offer broadband contracts in areas with individuals who have a high willingness to pay. Since these individuals typically have a higher income which in turn is associated with better health, the estimated coefficient would be severely overestimated. Therefore, the following model is the base for all further empirical analyses:

$$Y_i = \alpha + \beta DSL_i + \gamma \boldsymbol{X_i} + \delta_s + \epsilon_i \tag{1.1}$$

where the index denotes individuals i. The outcome variable Y_i is either one of the aforementioned weight related variables (BMI, overweight status, obesity) or satisfaction variables (health or overall life satisfaction). The main explanatory variable DSL_i is defined as 1 if the individual has a DSL Internet connection at home and zero otherwise. Therefore β is the coefficient of interest and denotes whether high-speed Internet access affects health status. Equation 1.1 further contains a vector $\boldsymbol{X_i}$ of individual-level and county-level control variables. More specifically individual-level controls include indicator variables for gender (male/female), marital status (married, single, widowed), house/flat ownership (yes/no), occupational status (not employed, apprentice, unemployed, retired, blue collar worker, entrepreneur), schooling (pupil, university, university of applied sciences, other degree, higher secondary, lower secondary), migration background (none, direct, indirect), interview month (January-December) and the continuous variables age, age squared and log net household income. County-level controls include the unemployment rate and GDP per capita. To exclude any distortive unobserved state-specific factors that are related to DSL availability and health outcomes, state fixed effects δ_s are included. ϵ_i is the error term.

Even if one were able to eliminate any omitted variable bias concerns by including control variables, there still remains (at least) one other threat to identification: reverse causality. For example, sicker persons (e.g. individuals who can't leave their home easily) could prefer fast Internet connections because it facilitates communication and social interaction (e.g. via Skype), which would introduce a downward bias to the estimated coefficient. To identify any causal effect of broadband access on health, and in absence of a randomized control trial, one has to look for a natural experiment that generates exogenous variation in DSL access.

The structure of the German fixed-line telephone network offers such a possibility. I will first explain the construction of an instrumental variable which I exploit in the following.[9] As mentioned before, broadband Internet access in (West) Germany in 2008 was primarily provided via DSL technology which relied on the existing fixed-line telephone network infrastructure and therefore was inexpensive to implement for Internet service providers (compared to installing new underground cables). This network was established in the 1960s and (almost) every household in West Germany was connected. More precisely, (telephone) catchment areas (CA) have been defined and within these areas all households have been connected via copper wires to a main distribution frame (MDF). These MDFs are in turn connected through a backbone network to offer a nationwide telephone network. Figure 1.3 in Appendix section 1.7 depicts all MDFs and CAs in West Germany.

For traditional telephone communication the distance within a CA between an MDF and a household was irrelevant, but for the availability of DSL Internet access, it's crucial. If a household is more than 4.2 kilometers away from its assigned MDF, broadband Internet provided via DSL is (technically) not available, because of the physical transmission loss which would reduce the transmission bandwidth below the defined minimum necessary to call it a DSL connection. Most of these households could only subscribe to low-speed Internet, because high-speed alternatives to DSL (e.g. broadband Internet via TV-cable, fibre wires, or satellite) were hardly available at the time in Germany (Bundesnetzagentur 2010).

This technical limit to DSL availability due to a household's distance to its MDF in their CA is used to create a binary instrument. It takes the value 1 for individuals living in households more than 4.2 km away from their assigned MDF in their CA and zero otherwise.[10] At the time of the German reunification in 1990 the telephone network of the former German Democratic Republic in East Germany was completely run down and new telecommunication infrastructure (i.e. MDFs, lines and wires) was rolled out afterwards. Therefore the distribution of the MDFs is potentially not as-good-as-random (Falck et al. 2014a, p.2248) in East Germany. Hence I exclude all

[9] This paragraph draws on Falck et al. (2014a) and Falck et al. (2014b). See their papers and the references therein for more technical details.

[10] The information on the location of the MDFs and the CAs is taken from Falck et al. (2014b). The distances have been calculated using *QGIS* and a georeferenced SOEP data set which can only be accessed at the "Research Data Center SOEP" at the DIW Berlin. See Goebel and Pauer (2014) for more information on access regulations, data protection and scope of this data set.

individuals living in East Germany (and Berlin) from my main analysis and estimate the following first-stage regression using all individuals living in West Germany:

$$DSL_i = \alpha + \beta \mathbb{1}(distance > 4.2km) + \gamma \boldsymbol{X_i} + \delta_s + \kappa_i \tag{1.2}$$

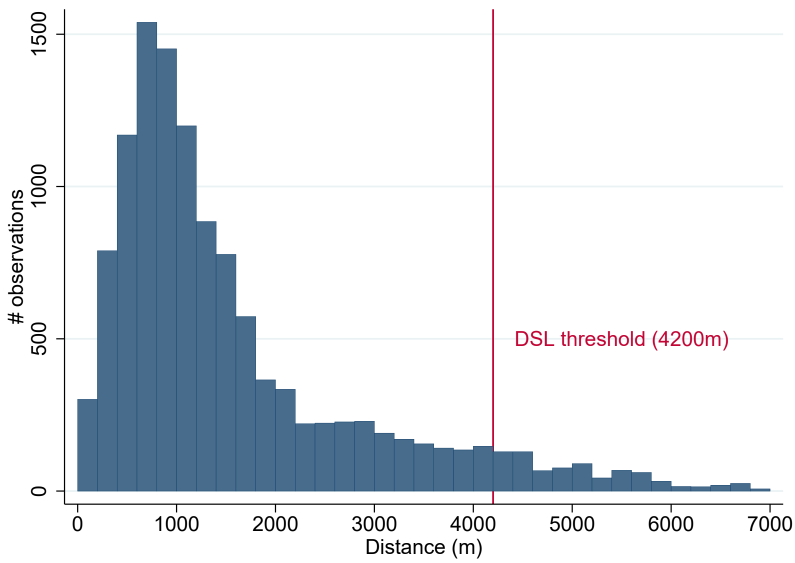
with $\mathbb{1}(distance > 4.2km)$ being a binary variable that takes the value one for individuals living in households more than 4.2 km away from the MDF in their CA and zero otherwise.[11] This identification strategy is slightly different to Falck et al. (2014a) who use a first difference instrumental variable approach. In the SOEP the collection of health variables started in 2002, while DSL access in the household was recorded in 2008 for the first time. In 2002 there were already 3.2 million DSL connections in Germany (Bundesnetzagentur 2010). Hence I cannot observe if some households already had DSL access by 2002. Therefore I rely on the cross sectional instrumental variables approach described above.

To evaluate the effect heterogeneity I split the sample further into demographic groups. As already noted in Section 1.2, Aguiar et al. (2017) show for the US that men aged 30 and younger substituted working hours with leisure time spent on recreational computer activities due to innovations in leisure technology, while neither older cohorts nor females react to these innovations. Since DSL access at home clearly is a household amenity, which can be used for recreational computer activities (e.g. web-gaming, online-dating, or instant messaging), I hypothesize that any health effects of broadband Internet access are more pronounced for men and younger cohorts. Hence I evaluate the following five demographic groups: all individuals, men, women, individuals at most 30 years of age and individuals older than 30 years in the main analysis. The younger cohort represents the so-called "digital natives", who already were exposed to computers and the Internet in their youth or early adulthood and therefore are more likely to be early adopters of broadband Internet access. As technical DSL availability varies on the household level, standard errors are clustered accordingly. Figure 1.2 shows the distribution of distances between households and their MDFs in my sample. One can clearly see that the average distance from a household to an MDF is well below 4.2 km. Nevertheless there are still quite a lot of households above the threshold.

This instrumental variables strategy identifies the local average treatment effect, i.e. the average treatment effect of broadband access on health for individuals who are less likely to have a DSL connection at home just because their home is located more

[11] Other technical and administrative parameters have an impact on DSL availability as well, making the 4.2 km limit a rather fuzzy threshold. In combination with sample size restrictions in the vicinity of the threshold, a Regression Discontinuity Design analysis is unfortunately not possible.

Figure 1.2: Distance to closest telephone node



Notes: Distance to closest node (in meters) for individuals in West Germany in 2008.
Source: Own calculations based on SOEP v32.1.

than 4.2 km away from their MDF, but would otherwise have broadband subscription (compliers). They key identifying assumption (besides the monotonicity, the "as-good-as-random" and the strong instrument assumption) is that technical DSL availability affects an individual's health status only through actual DSL access at home.

1.5 Results

1.5.1 OLS results

First, I present the effects of broadband Internet access in a household on an individual's health status outcomes using simple OLS regressions described above in equation 1.1 in a West German sample. Each cell in Table 1.3 displays the results for one of the five health outcome variables (columns) and one of the five subgroups (panels) defined above. All regressions include individual-level and county-level control variables and state-fixed effects. Standard errors are clustered on the household level.

Table 1.3: Association between broadband Internet access and health

	BMI	Overweight	Obese	Health satisfaction (std)	Life satisfaction (std)
Panel A			**Overall**		
Broadband access in hh	0.19* (0.11)	0.0014 (0.011)	0.0096 (0.0089)	-0.022 (0.023)	0.018 (0.026)
Observations	10852	10852	10852	10852	10852
R^2	0.112	0.129	0.040	0.137	0.089
Panel B			**Men**		
Broadband access in hh	0.48*** (0.15)	0.024 (0.015)	0.029** (0.013)	0.026 (0.031)	0.025 (0.032)
Observations	5176	5176	5176	5176	5176
R^2	0.086	0.110	0.044	0.147	0.116
Panel C			**Women**		
Broadband access in hh	-0.067 (0.15)	-0.018 (0.015)	-0.0052 (0.012)	-0.067** (0.030)	0.015 (0.032)
Observations	5676	5676	5676	5676	5676
R^2	0.117	0.104	0.048	0.136	0.076
Panel D			**Age $\leq$ 30**		
Broadband access in hh	0.88*** (0.25)	0.069** (0.027)	0.033** (0.017)	-0.13** (0.051)	0.078 (0.061)
Observations	1558	1558	1558	1558	1558
R^2	0.124	0.111	0.066	0.073	0.085
Panel E			**Age > 30**		
Broadband access in hh	0.083 (0.12)	-0.0078 (0.012)	0.0065 (0.0100)	-0.0028 (0.026)	0.0049 (0.027)
Observations	9294	9294	9294	9294	9294
R^2	0.071	0.090	0.030	0.107	0.096
Controls	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions; Individual-level controls include gender, marital status, house/flat ownership, occupational status, schooling, migration background, age, age squared and log net household income and interview month; County-level controls include population density, the unemployment rate and GDP per capita; Sample: All individuals in 2008 in West-Germany; Standard errors clustered on household level in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Own calculations based on the SOEP v32.1.

For the entire sample (Panel A of Table 1.3) there doesn't seem to be a connection between broadband Internet access and health variables. The coefficients are very small and only one is weakly significant. The results for the subsample of males (Panel B of Table 1.3) show a different picture: There is a significant and sizeable association between DSL access and the weight variables: Men in households with DSL access are 2.9 percentage points more likely to be obese and have a 0.48 points higher BMI (conditional on covariates). The coefficients for the satisfaction variables are small and insignificant. When one turns to cohorts aged 30 and younger (Panel D of Table 1.3), one can see that the effects on weight are even more pronounced. Young cohorts with DSL connection are 6.9 percentage points more likely to be overweight and 3.3 percentage point more likely to be obese. Younger cohorts with DSL access even exhibit slightly worse health satisfaction. As expected neither for women nor older cohorts any effects of broadband Internet on health are significant - with the exception of a small negative effect of broadband Internet on health satisfaction (see Panel C and E of Table 1.3). For women the effects even point into the opposite direction. These results are in stark contrast to the effects found by DiNardi et al. (2017). They show for the US that an increase in broadband providers in a county leads to a significant weight gain and higher obesity rates only for women.

1.5.2 Instrumental variable results

As mentioned in Section 1.4, caution is warranted when it comes to a causal interpretation of these effects, because there might still be endogeneity issues despite controlling for individual-/county-level covariates and state fixed effects. Hence, I now present the results of the instrumental variables approach that uses exogenous variation in technical DSL availability to instrument actual DSL subscriptions in a household.[12] Table 1.4 has the same basic structure as Table 1.3 (columns = outcomes, panels = subsamples). Additionally the first column shows the results of the first stage regressions in equation 1.2 and the row denoted as "First-stage F-statistic" refers to the Kleibergen-Paap F-statistic.

[12] Reduced-form estimates can be found in Table 1.5 in Appendix section 1.7.

Table 1.4: Broadband Internet access and health - Instrumental variable results

	Broadband access in hh	BMI	Overweight	Obese	Health satisfaction (std)	Life satisfaction (std)
Panel A	**Overall**					
Threshold	-0.11***					
	(0.025)					
Broadband access at home		1.13	-0.0051	0.017	-0.072	0.40
		(1.77)	(0.19)	(0.15)	(0.40)	(0.41)
Observations	10159	10159	10159	10159	10159	10159
First-stage F-statistic	17.9					
Panel B	**Men**					
Threshold	-0.098***					
	(0.029)					
Broadband access at home		-0.59	-0.24	-0.040	-0.22	0.20
		(2.30)	(0.28)	(0.23)	(0.57)	(0.57)
Observations	4845	4845	4845	4845	4845	4845
First-stage F-statistic	11.8					
Panel C	**Women**					
Threshold	-0.12***					
	(0.027)					
Broadband access at home		2.57	0.21	0.072	0.031	0.57
		(2.34)	(0.24)	(0.16)	(0.47)	(0.47)
Observations	5314	5314	5314	5314	5314	5314
First-stage F-statistic	18.7					
Panel D	**Age $\leq$ 30**					
Threshold	-0.18***					
	(0.051)					
Broadband access at home		-1.80	-0.44	-0.24	-0.84*	-0.46
		(2.71)	(0.30)	(0.20)	(0.48)	(0.52)
Observations	1468	1468	1468	1468	1468	1468
First-stage F-statistic	12.3					
Panel E	**Age > 30**					
Threshold	-0.095***					
	(0.026)					
Broadband access at home		1.88	0.13	0.092	0.19	0.68
		(2.07)	(0.22)	(0.18)	(0.48)	(0.51)
Observations	8691	8691	8691	8691	8691	8691
First-stage F-statistic	13.6					
Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: IV regressions; Individual-level controls include gender, marital status, house/flat ownership, occupational status, schooling, migration background, age, age squared and log net household income and interview month; County-level controls include population density, the unemployment rate and GDP per capita; Sample: All individuals in 2008 in West-Germany; Standard errors clustered on household level in parentheses; First-stage F-statistic refers to the Kleibergen-Paap F-statistic; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Own calculations based on the SOEP v32.1.

The first-stage results indicate that technical broadband Internet availability is a strong instrument. The F-Statistics for all subgroups vary between 11.8 (for men) and 18.7 (for women). Individuals living more than 4.2 km away from the MDF in their CA have a 11 percentage point lower probability to have DSL access in their household compared to individuals living closer to an MDF. Except for a weakly significant negative effect of broadband Internet access on health satisfaction, all coefficients for all subgroups are insignificant. Interestingly, all instrumental variable results are larger (in absolute value) than the corresponding OLS results. Furthermore the coefficients for the weight variables switch sign for all but one subgroup (individuals older than 30).[13] Hence I cannot rule out that broadband Internet access has no effect on health status in Germany.

1.6 Conclusion

The Internet has fundamentally transformed almost every part of human life in the 21st century. Interestingly, there is little (causal) research on potential health effects of this revolutionary technology. In this chapter I try to bridge this gap by analysing a specific part of the Internet, namely the influence of broadband Internet access on the health status of five different demographic groups. Using German individual-level data from the SOEP, I measure health status in terms of weight (BMI, overweight status and obesity), and life and health satisfaction. I begin the empirical analysis by estimating OLS regressions with a broad range of control variables. The results show a significant weight gain for men and cohorts aged 30 and younger with DSL access, but there is no "DSL effect" on satisfaction levels for these groups. Neither for women nor for older cohorts aged 30+ I find any significant effects. On the one hand, this finding is in line with the results of Aguiar et al. (2017) who show for the US that technological innovations increased the share of leisure time spent on recreational computer use for young men, but not for women or older cohorts. On the other hand, DiNardi et al. (2017) find that more broadband providers in an area increase weight only for white women in the US.

Since these simple OLS regressions may be plagued by endogeneity issues, a quasi-experimental identification strategy, which has been used in the economics literature in other contexts, is applied in the next step. I employ georeferenced SOEP data in combination with the structure of the preexisting telephone network to calculate a binary instrumental variable which states whether a DSL access is technically possible for a household to instrument actual DSL subscriptions in a household. In comparison

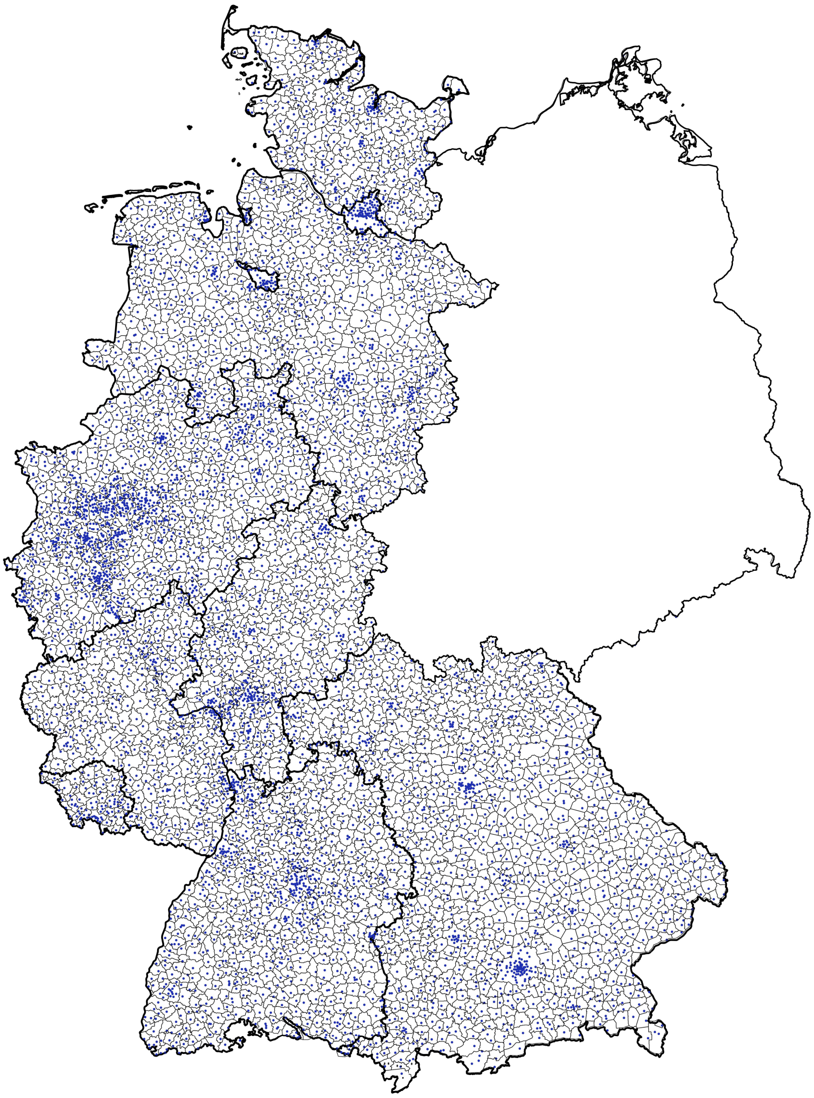
[13] Regressions with alternative sets of control variables, standard errors clustered on CA-level or probit regressions for binary outcomes yield similar results and are available from the author on request.

to the OLS results, the IV coefficients are insignificant for all but one variable. Hence I cannot rule out that broadband Internet access has no effect on the health status of an individual in Germany.

Nevertheless, in future research it would be interesting to assess whether and how individuals with broadband Internet access change their health behaviours or whether they choose health care providers/insurance differently. Furthermore the health care supply side (e.g. general practitioners, hospitals and health insurance firms) and other demographic groups (e.g. children), which I have not been able to observe, might be affected by broadband Internet access as well. Future research on these topics may reveal interesting insights into further - beneficial or adverse - side effects of modern technology adoption in the health domain and possibly imply government action.

1.7 Appendix

Figure 1.3: Structure of the German telephone network



Notes: Map of Germany with all telephone access main distribution frames (blue dots) and catchment areas in West Germany.

Table 1.5: Broadband Internet access and health - Reduced-form results

	BMI	Overweight	Obese	Health satisfaction (std)	Life satisfaction (std)
Panel A			**Overall**		
Threshold	-0.12	-0.000059	-0.0020	0.0071	-0.043
	(0.19)	(0.020)	(0.016)	(0.043)	(0.043)
Observations	10159	10159	10159	10159	10159
R^2	0.111	0.130	0.038	0.133	0.087
Panel B			**Men**		
Threshold	0.059	0.023	0.0038	0.021	-0.020
	(0.23)	(0.026)	(0.023)	(0.056)	(0.056)
Observations	4845	4845	4845	4845	4845
R^2	0.084	0.111	0.043	0.142	0.116
Panel C			**Women**		
Threshold	-0.32	-0.026	-0.0090	-0.0058	-0.067
	(0.26)	(0.027)	(0.019)	(0.054)	(0.052)
Observations	5314	5314	5314	5314	5314
R^2	0.116	0.105	0.046	0.132	0.073
Panel D			**Age $\leq$ 30**		
Threshold	0.32	0.078	0.043	0.15*	0.083
	(0.48)	(0.048)	(0.034)	(0.080)	(0.090)
Observations	1468	1468	1468	1468	1468
R^2	0.115	0.106	0.070	0.071	0.085
Panel E			**Age $>$ 30**		
Threshold	-0.18	-0.013	-0.0089	-0.018	-0.065
	(0.19)	(0.021)	(0.017)	(0.045)	(0.046)
Observations	8691	8691	8691	8691	8691
R^2	0.070	0.090	0.029	0.104	0.094
Controls	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions; Individual-level controls include gender, marital status, house/flat ownership, occupational status, schooling, migration background, age, age squared and log net household income and interview month; County-level controls include population density, the unemployment rate and GDP per capita; Sample: All individuals in 2008 in West-Germany; Standard errors clustered on household level in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Source: Own calculations based on the SOEP v32.1.

Chapter 2

Is it good to be too light? Consequences of birth weight thresholds in hospital reimbursement systems

2.1 Introduction

Small changes in birth weight can have important financial implications for hospitals in many of the widespread prospective payment schemes (PPS) which reimburse hospitals with a fixed rate for the treatment of strictly defined diagnosis related groups (DRGs). More specifically, hospitals receive a higher reimbursement for newborns with birth weight just below certain thresholds than for newborns with weight above, leading to an incentive to under-report birth weight. The evidence is accruing that the introduction of birth weight thresholds has led to large under-reporting – so called upcoding – of birth weight (e.g. Jürges and Köberlein 2015; Shigeoka and Fushimi 2014).

At the same time, birth weight thresholds are used to diagnose newborns as having "extremely low" (weight $\leq$ 1000g), "very low" (weight $\leq$ 1500g), or "low" birth weight (weight $\leq$ 2500g) and appear in medical guidelines. Despite the fact that low birth weight is typically linked to worse health outcomes (see e.g. Hack et al. (2002), Hummer et al. (2014)), newborns with weight just below the 1500g diagnostic threshold have been found to have higher survival chances than newborns with weight just above (Almond et al. 2010; Bharadwaj et al. 2013; Breining et al. 2015). To the extent that reimbursement-relevant birth weight thresholds are identical to the diagnostic thresholds, the practice of upcoding newborns' weight may thus be to the benefit of the newborn. Furthermore, as the hospital receives more money for newborns with weight below the reimbursement relevant thresholds, they may also be able or willing to deliver additional care to these newborns.

In this chapter, we investigate whether newborns benefit from having a reported birth weight below reimbursement relevant weight thresholds in the German DRG system. Based on an administrative hospital claims data set covering the universe of hospital births in Germany in the years 2005–2011, we compare the survival chances and the quantity of care that newborns with weight just below the relevant thresholds receive to those of newborns with weight just above the thresholds. We include all eight reimbursement relevant thresholds in the German DRG system. While some of these thresholds overlap with the diagnostic thresholds and/or are explicitly mentioned in medical guidelines, others are only relevant for reimbursement, allowing us to shed light on the importance of diagnostic thresholds, medical guidelines, and reimbursement for the care of newborns.

Different from the settings in the earlier literature on the effects of diagnostic thresholds (Almond et al. 2010; Bharadwaj et al. 2013; Breining et al. 2015), the fact that the diagnostic thresholds are relevant for reimbursement in our setting imposes a challenge to the empirical analysis: If the decision to upcode a newborn's weight below a threshold depends on the newborn's health status, the crucial assumption that newborns with reported weight above and below the threshold have ex ante similar health is not plausible, an issue that Barreca et al. (2011) already raised in the light of the earlier literature. As Jürges and Köberlein (2015) show, it is likely that birth weight manipulation in German hospitals is not random. To the contrary, hospitals primarily tend to upcode newborns for whom staff expects more care. These are relatively fragile newborns that still have non-negligible survival chances and will therefore receive a lot of treatment. We take three steps in our analysis to take this into account: First, we control for a large set of variables capturing a newborn's health at birth. We specifically choose variables such as sex and plurality of births that are not easy to manipulate and observable to the hospital staff who reports newborns' weight. These are variables that may influence reported birth weight. Second, we restrict our analysis to newborns who survive the first day of their life, thus excluding those very fragile cases for whom hospital staff may expect an early death, making upcoding of birth weight not worthwhile. In a third set of results, we further include hospital fixed effects in our estimations, taking into account possible differences in coding practices and treatment across hospitals.

Our findings show that without controls and not restricting to first-day-survivors newborns with weight below almost all of the eight thresholds stay in the hospital longer, receive more procedures, and have lower mortality (during the hospital stay). However, for all but the highest two thresholds (2000g and 2500g) these results become insignificant or even change sign, when we control for health at birth, exclude newborns who die on the first day of life, and include hospital fixed effects. We interpret this as evidence that neither reimbursement differences, nor the diagnostic threshold of 1500g or thresholds in medical guidelines trigger additional care or reduce mortality among newborns in Germany. While our results may indicate that newborns with weight just below 2000g and 2500g benefit from being below these thresholds, we believe that the differences rather reflect systematic upcoding related to factors that we cannot control perfectly. As further discussed in the final section of this chapter, we therefore conclude that DRG-related upcoding in Germany does not affect the care upcoded newborns receive.

Our research project contributes to and brings together two strands of literature. The first focuses on the question whether hospitals upcode diagnoses or other health measures, such as birth weight, to generate higher payments in DRG reimbursement systems. Concerning the upcoding of diagnoses, Dafny (2005) shows that hospitals reacted to a recalibration of Medicare DRGs in 1988 by disproportionally shifting patients to diagnoses codes that became more lucrative. At the same time, she finds no changes in the treatment that patients receive nor in patient mortality. Silverman and Skinner (2004) focus on Medicare patients with respiratory disease and show that the share of patients coded to the highest paying DRG increases significantly over time, particularly so in the group of for-profit hospitals. Furthermore, there is evidence that also hospitals in the Italian region of Emilia-Romagna (Verzulli et al. 2016), Portugal (Barros and Braun 2016) and Norway (Januleviciute et al. 2016) upcode patients to the highest paying DRGs. Concerning upcoding of patient characteristics, Shigeoka and Fushimi (2014) show that hospitals in Japan have manipulated birth weight as a response to the introduction of a partial PPS in a way that increased hospital payments. Similarly, Jürges and Köberlein (2015) find that German hospitals reacted to the introduction of the DRG payment system in 2003 by systematically under-reporting newborns' weight. The second strand of literature focuses on the effect of birth weight thresholds on the quantity and quality of care that newborns receive. Based on the census of U.S. births, Almond et al. (2010) and Almond et al. (2011) find that newborns with weight just below the very low birth weight threshold at 1500g have higher survival chances than newborns with weight above. Based on hospital discharge data for five states they further find that birth weight below 1500g triggers additional care that also results in higher hospital charges. Their results are particularly concentrated among low quality hospitals, i.e., those hospitals that offer no or only low levels of neonatal intensive care. Similar effects of the very low birth weight thresholds have been found for newborns in Chile where medical guidelines explicitly recommend different treatment depending on a very low birth weight diagnosis (Bharadwaj et al. 2013). Breining et al. (2015) focus on Denmark where the treatment recommendations in medical guidelines only vary across the very low birth weight thresholds for newborns with at least 32 weeks of gestation. They find that indeed only for newborns born at 32 weeks of gestation or more, treatment depends on birth weight. For newborns born earlier in the pregnancy they find no treatment differences with birth weight, indicating that medical guidelines have an impact on the care that newborns receive.

This chapter brings these two strands of the literature together in investigating whether newborns benefit from upcoding of birth weight below thresholds that may themselves affect the care that newborns receive as they are diagnostic thresholds. Our analysis contributes to the first strand of the literature by investigating effects of upcoding on the care the patients receive. We add to the second strand by focusing not only on the very low birth weight threshold but also on other thresholds that impact diagnoses and appear in medical guidelines.

The rest of the chapter is structured as follows: In Section 2.2, we give an overview on the institutional background in Germany and discuss which different birth weight thresholds may be related to the quantity and quality of care that newborns receive. We introduce our data in Section 2.3 and the empirical strategy in Section 2.4. In Section 2.5, we present our results, the sensitivity of which we explore in Section 2.6. The chapter closes with a discussion and conclusion in Section 2.7.

2.2 Institutional Background

In this section we give a brief overview on the German DRG (G-DRG) system and then describe reasons why the treatment that newborns receive in German hospitals may vary around birth weight thresholds.

2.2.1 The G-DRG reimbursement system

Until the year 2003, German acute care hospitals were reimbursed through a multiple-source-system consisting of a hospital specific patient/day base-rate, a ward specific rate and case-based lump-sums. To increase efficiency and, in particular, reduce length of stay, the G-DRG system was introduced. Under the old and the new system, reimbursement was and is generally the same irrespective of individuals' insurance status (public or private). Based on the Australian DRG system, 664 DRGs defined by combinations of diagnoses, performed procedures, hours of ventilation, age and (for perinatal DRGs) birth weight were set up. To each DRG a case-weight was assigned which determined the final reimbursement. This case-weight multiplied by a base-rate gives the amount of money a hospital receives for treating the respective patient.

In a transition period from 2003 to 2010 hospital specific base-rates were used. These rates advanced towards state specific rates over the years.[1] Since 2010 most hospitals are reimbursed according to this state base-rate. From 2010 to 2014 these state specific base rates in turn narrowed down to a federal base rate interval (Salm and Wübker 2015).

Case-weights and group definitions vary from year to year. The German Institute for Hospital Reimbursement[2] decides on changes of case-weights and DRG definitions every year using information on actual costs of different treatments reported voluntarily by a group of hospitals. Since the introduction of the G-DRG system, the number of DRGs almost doubled to 1200 in 2015.[3]

The change over time in the number of DRGs in neonatal care, however, was not as drastic. When DRGs were first put into use in Germany in 2003, 38 different groups for the treatment of newborns were defined. Until 2015 this number has increased only slightly to 42 groups. The majority (37/42) of these groups depend on birth weight. The remaining 5 DRGs are for newborns that die within the first four days of their life and for specific cardiovascular conditions. The 37 birth-weight related DRGs depend on eight birth weight thresholds. Within each birth weight interval multiple DRGs cover different degrees of treatment complexity and complications. For our main analysis, we investigate differences across these thresholds.

To compare only similar cases left and right of the thresholds we also conduct analyses within "severity groups" across thresholds. To this end we group the DRGs into six severity groups that have the same level of complications and complexity based on the DRG catalogue. Within each severity group birth weight should be the only differ-

[1] The reason for starting with hospital specific rates was that hospital cost structures were drastically different when the G-DRG system was introduced. Starting off directly with identical reimbursement in each state would have led to major financial struggles for some hospitals.

[2] Institut für das Entgeldsystem im Krankenhaus (InEK), financed by private and social health insurers as well as the German Hospital Association.

[3] For an analysis of these changes see Schreyögg et al. (2014).

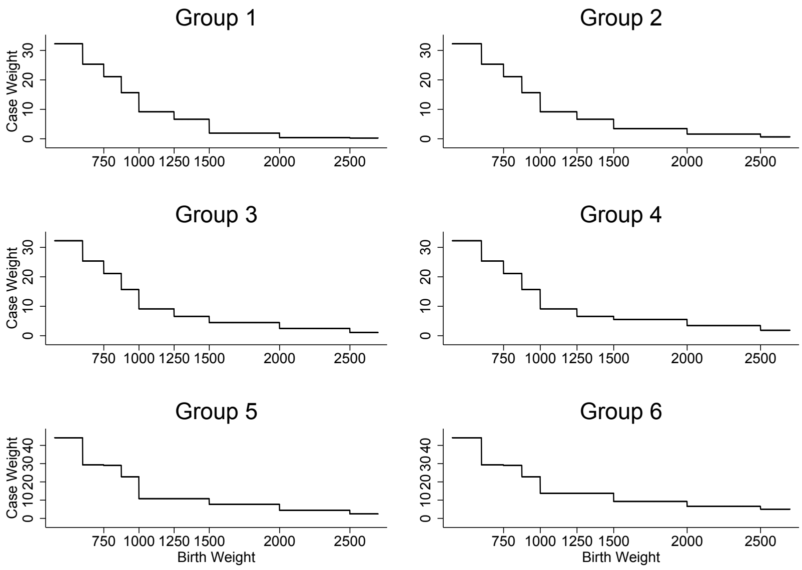
ence across thresholds. Table 2.6 (in the Appendix) shows which DRGs we assign to the different severity groups in the different birth weight intervals. Group 1 in Table 2.6 comprises the least severe cases, group 6 the most severe ones. At very low birth weights there are only two different groups of severity. The first contains newborns with lower levels of complications and less severe health conditions. This group splits up into four groups for higher birth weight intervals (group 1-4). The second group contains more severe cases and splits up into two groups for higher birth weights (group 5/6). Starting from a birth weight of 1000g, three severity groups apply as the joint group 5 and 6 is split into two severity groups. At birth weights of 1500g and higher all 6 groups are present.

2.2.2 Treatment and birth weight

Broadly speaking, there are two reasons for differences in treatment of newborns across birth weight intervals. As explained in more detail below, the first is that DRGs are defined along birth weight intervals, leading to sharp increases in reimbursement when birth weight crosses from above to below thresholds and hence more funds to spend on costly procedures. The second reason is that birth weight thresholds are used in the diagnosis of newborns, define the level of specialisation of the hospital that the newborn should be treated in, and appear in medical guidelines.

Thresholds and reimbursement
Figure 2.1 provides a systematic overview on changes in case-weights with birth weight within the six DRG-severity groups defined above, i.e. holding everything else except for birth weight constant. Generally, the case-weight decreases at 600g, 750g, 875g, 1000g, 1250g, 1500g, 2000g, and 2500g. Reimbursement for newborns with weight just below the thresholds is thus generally higher than for newborns with weight on or above the thresholds. The sizes of the jumps in case-weights (and thus in reimbursement), however, vary across groups and within groups across thresholds. The 2500g threshold for example is not relevant for reimbursement for newborns in the least severe group (group 1), while the 750g threshold only hardly matters for the cases with most severe conditions (group 5/6). As an example for how small changes in birth weight can affect reimbursement let's consider a notional hospital in the German state of Hesse in the year 2010 which had a base-rate equal to the state base-rate of €2,968.56. A generic newborn in this hospital may have had a birth weight of 1004g and was assigned

Figure 2.1: Hospital reimbursement by birth weight



Notes: Development of case-weights for the six DRG-severity groups defined in Tables 2.6 for the year 2008. **Source:** DRG catalogue 2008.

DRG P63Z[4] which in this year had a case-weight of 8.776. Total reimbursement to the hospital would have been €26,052.08. Had the weight instead been 999g, DRG P62D with a case-weight of 15.51 would have been assigned. This means the hospital would have received €46,042.37, a plus of almost €20,000 for a 5g change in birth weight. As the financial incentives apply to the hospital, while treatment is typically decided on by the medical personnel, one might argue that these reimbursement differences should not play a large role for medical decision making. However, Jürges and Köberlein (2015) have shown that reimbursement differences by birth weight have led to manipulation of reported weight. As birth weight is reported by medical personnel, the individuals who decide on treatment are aware of the financial importance of reimbursement differences. If they are also informed about the costs of different procedures, they may also take these differences in reimbursement into account when deciding on which treatment newborns receive.

[4] For example, because it was a girl, ventilation was in use for 48 hours, respiratory distress syndrome (ICD-10-GM: P28.5), systematic inflammatory response-syndrome (ICD-10-GM: R65.0), and some other infectious disease (ICD-10-GM: P37.9) were diagnosed and the girl was treated by monitoring cardiovascular levels (OPS: 8-930), was given a VR-infusion (OPS: 8-811.0), CAPA ventilation was performed (OPS: 8-711.00) and after this did not work out, endotracheal intubation (OPS: 8-701) came into use.

Thresholds and medical guidelines

In addition to reimbursement differences, birth weight determines the level of specialisation of the hospital in which newborns should be treated. In Germany, four degrees of specialisation are defined by the Gemeinsamer Bundesausschuss (2013)[5]: Level 4 indicates regular maternity clinics, Level 3 means that basic neonatal care can be provided, Level 2 clinics have some specialisation in neonatal care, and Level 1 clinics maintain neonatal intensive care units (NICUs). Preterm births with weight above 1500g should be treated in clinics of at least Level 3, birth weight between 1250 and 1500g induces Level 2 care, while newborns below 1250g should be treated in Level 1 clinics.[6] [7] To the extent that better care is provided in more specialized hospitals, birth weight just below the thresholds that require care in more specialized hospitals, in particular below 1250g could be beneficial for newborns.

As results in the earlier literature suggest (Almond et al. 2010; Bharadwaj et al. 2013), further differences in treatment around birth weight thresholds could result as a consequence of thresholds in medical guidelines or as thresholds determine diagnoses, such as the international standard classifications of extremely low (below 1000g), very low (below 1500g), and low birth weight (below 2500g). The German Association of the Scientific Medical Societies (AMWF) publishes a multitude of medical guidelines for neonatal care. They range from guidelines concerning the care of very frail newborns to care for healthy newborns. Additionally, there are guidelines on after-hospital care for groups of newborns with specific conditions. Except for the 875g and 2000g thresholds, all reimbursement relevant thresholds are mentioned in at least one of the guidelines concerning neonatal care. The threshold that appears most often is 1500g, the diagnostic threshold for very low birth weight. Only few guidelines, however, give specific recommendations on care depending on birth weight. One example that does is the guideline on parenteral nutrition. Among other things, it recommends specific nutritional solutions for newborns with weight below 1500g (AWMF 2014). Most recommendations seem to be linked to gestational age rather than weight.

[5] The highest self administration unit in the German health sector which consists of 13 members representing health insurers, providers as well as the general public.

[6] Interestingly, the *German Association for Perinatal Medicine* uses different Level labels (1, 2a, 2b and 3) and slightly different thresholds which would admit less newborns to NICUs (Bauer et al. 2006). Nevertheless, the Gemeinsamer Bundesausschuss (2013) regulations are binding.

[7] Theoretically this structure could offer an incentive for hospitals to increase birth weight of newborns that would be too light for this type of clinic and would be transferred otherwise. Since case-weights are calculated for average newborns within a birth weight bracket, reimbursement would most likely not cover all the costs for newborns at the lower end of a birth weight DRG, making over-reporting of weight to keep newborns implausible (Jürges and Köberlein 2015).

Although they do not contain specific treatment recommendations, several guidelines report risks of specific diseases or survival chances for birth weight intervals. For example, the risks of complications like necrotizing enterocolitis (11% for 401-750g, 9% for 751-1000g, 6% for 1001-1250g, and 4% for 1251-1500g) and patent ductus arteriosus (30% for preterm birth with weight below 1500g, and 50-70% for newborns below 1000g) are reported for birth weight intervals. Furthermore, survival chances for esophageal atresia are reported separately for newborns with weight above or below 1500g, with higher chances for heavier newborns (AWMF 2012). Although risk classifications are no direct recommendations on care, they may result in treatment differences, to the extent that they influence physicians' awareness of risks for newborns in the different weight groups.

As some thresholds are only relevant for reimbursement (875g, 2000g), while others affect reimbursement as well as the type of hospital that a newborn should be treated in (1250g), or overlap with diagnostic thresholds and are mentioned in medical guidelines (600g, 750g, 1000g, 1500g, 2500g), we can shed light on the importance of the drivers of differences in medical care around the thresholds by comparing effects across thresholds.

2.3 Data

Our analyses are based on the universe of German hospital claims from the years 2005–2011. All German hospitals have to submit their DRG claims to the Institute for Hospital Reimbursement (InEK). InEK forwards parts of the data to the German Federal Statistical Office, which makes the data available to researchers.[8]

For each of the roughly 20 million hospital stays per year in Germany, the data contain basic demographic information on the patients, as well as detailed information on the hospital stay. In addition to the exact DRG used for reimbursement, the data contain information on the reason for the hospital stay, the length of the stay, procedures performed during the stay (the German version of ICPM codes, called OPS codes), diagnoses (ICD-10-GM) codes, and the reason of discharge (including regular discharge, death or referral to a different hospitals among others). For newborns, the data further contain information on birth weight.

In our analyses, we restrict the data to all births (cases with a perinatal DRG (those beginning with "P") and "birth" as cause of admission) with valid information on new-

[8] For further information on the data see www.forschungsdatenzentrum.de/en/database/drg [last accessed: 17 September 2017].

borns' sex and a birth weight between 450 and 3000g.[9] After applying these selection criteria our sample amounts to a total of 985,885 cases.

Table 2.1 displays basic descriptive statistics for our data, averaged across all cases with birth weight 450-3000g and separately for the different birth weight intervals that are relevant for reimbursement (< 600g, 600-749g, 750-874g, 875-999g, 1000-1249g, 1250-1499g, 1500-1999g, 2000-2499g, 2500-3000g). It contains information on birth weight, on quality and quantity of care (length of the hospital stay, number of procedures, mortality), on the newborn's health at birth, as well as hospital related control variables.

On average, newborns have a birth weight of 2601g, have 2.6 reported procedures, and stay in the hospital for 8.5 days. Only 0.6% of newborns die during the hospital stay.[10] Birth weight increases mechanically across the birth weight intervals. At the same time quantity of care and mortality generally decrease. The 600g threshold constitutes an exception: Newborns with weight below 600g receive less care than newborns in the next interval. This may be explained by the fact that in the lighter group of newborns a larger share dies very early on (34% die on their first day of life, compared to 13% in the group with weight 600-749g), likely leading to shorter hospital stays and fewer procedures for the group with weight below 600g.

To measure health at birth we use variables that are visible at birth and known health risk factors for newborns. These are important variables for our analysis as they could influence the reporting of birth weight as well as the care that newborns receive in the hospital. In addition to sex, prematurity, and plurality of birth (which are mentioned explicitly as risk factors for newborns' survival in medical guidelines (AWMF 2007), we focus on two conditions that usually become apparent during or right after delivery and are leading causes of death among preterm births: asphyxia and infant respiratory distress syndrome (IRDS). Furthermore, we include several indicators for maternal health-behaviour during pregnancy, namely whether the mother smoked, drank alcohol, or took illegal drugs.[11]

[9] By selecting only cases with "birth" as cause of admission, we exclude cases that are re-admitted after a first discharge. Since patients cannot be linked across hospital stays, this is necessary to avoid double-counting newborns that leave the hospital and are re-admitted later.

[10] We derive the information on whether a newborn dies from the reason of discharge.

[11] Unfortunately, linking the data on newborns to their mothers – who constitute a separate hospital case – is impossible. Most procedures concerning delivery - except for those specifically undertaken for the newborn – are billed as the mother's case, not the newborn's case. Thus, the data contain almost no information on what happened during pregnancy or delivery. To the extent that this information determines the newborn's reported weight and later treatment, this is valuable information that we unfortunately do not have in the data. To account for this as best as we can, we proxy for mothers' behaviour during pregnancy by relying on diagnoses that the newborn receives.

In our data, 44% of all newborns are boys,[12] 19% of births are premature, i.e., born with at least 28, but less than 37 completed weeks of gestation (ICD-10-GM: P07.3). Moreover 1.2% are extremely premature, i.e., have a gestational age of less than 28 weeks (ICD-10-GM: P07.2). 11% of newborns are twins. 0.4% are higher order births. Asphyxia is a form of oxygen deprivation that occurs during delivery. The ICD-10-GM coding system contains three different diagnosis codes for asphyxia, one for newborns with severe asphyxia and a 1-minute APGAR score[13] of 0-3 (ICD-10-GM: P.21.0), one for newborns with mild asphyxia and a 1-minute APGAR score of 4-7 (ICD-10-GM: P.21.1), as well as asphyxia without additional classification of severity (ICD-10-GM: P.21.2). Among all newborns in our data 2.1% had general asphyxia, 1.6% had mild asphyxia, and 0.5% severe asphyxia. IRDS (ICD-10-GM: P22.-) is a lung malfunction which is common among preterm births and begins shortly after birth. About 10.6% of all newborns in our data suffer from IRDS.

Only 1.2% of newborns in our data are (suspected to be) affected by maternal use of tobacco (ICD-10-GM: P04.2). This number is much lower than prevalences of smoking during pregnancy as reported in other sources. Kuntz and Lampert (2016), for example, report a prevalence of 12.1% for children born in Germany in the years 2003–2012 based on survey data. As only severe cases of smoking during pregnancy may be apparent to the hospital staff at the time of delivery, it is likely that only severe cases of smoking are flagged in our data. To the extent that newborns diagnosed with suspected damage due to maternal smoking are those for whom the medical personnel expects higher needs of care during the hospital stay and thus may also manipulate the weight, the indicator available to us flags the relevant cases. On average, there are less than 0.1% of newborns for whom alcohol use (ICD-10-GM: P04.3) and 0.2% for whom drug use during pregnancy (ICD-10-GM: P04.4) are reported. The shares are higher (0.1% for alcohol use and 0.3 to 0.5% for drug use) for lower birth weight intervals. Similar to maternal smoking, we believe that conditions related to maternal alcohol or drug use are coded for severe cases of use and thus flag relevant cases.

[12] This unusually low number may result from the high share of male stillbirths among low birth weight newborns. Stillbirths are not included in our data.

[13] The APGAR score was developed in the 1950s to summarize a newborn's health. It measures **A**ppearance (skin colour), **P**ulse (heart rate), **G**rimace (reflex irritability), **A**ctivity (muscle tone), and **R**espiration.

Table 2.1: Descriptive statistics by birth weight bracket, 2005-2011

	< 600	600-749	750-874	875-999	1000-1249	1250-1499	1500-1999	2000-2499	2500-3000	All births
Birth weight (grams)	527.092	680.801	816.950	948.530	1149.752	1404.486	1808.097	2306.875	2808.359	2601.265
	(43.828)	(45.427)	(34.144)	(37.177)	(69.413)	(72.487)	(134.898)	(136.260)	(140.054)	(441.940)
Outcomes										
Length of stay (days)	53.706	69.951	67.946	61.468	50.428	39.012	23.776	9.821	4.662	8.452
	(62.826)	(55.507)	(43.185)	(34.718)	(27.785)	(21.412)	(16.439)	(9.564)	(4.196)	(14.196)
Procedures (#)	11.729	14.354	13.638	12.192	10.203	8.296	5.942	3.208	1.724	2.559
	(11.393)	(10.688)	(9.407)	(7.569)	(6.662)	(5.129)	(4.154)	(2.988)	(1.601)	(3.275)
Mortality	.487	.236	.126	.069	.042	.025	.012	.003	.001	.006
	(.500)	(.425)	(.332)	(.253)	(.201)	(.155)	(.111)	(.057)	(.028)	(.077)
First-day-mortality	.339	.125	.056	.031	.024	.014	.008	.002	.000	.003
	(.474)	(.331)	(.231)	(.174)	(.154)	(.118)	(.087)	(.044)	(.021)	(.059)
Health-related controls										
Male births	.474	.502	.521	.529	.518	.501	.484	.453	.432	.442
	(.499)	(.500)	(.500)	(.499)	(.500)	(.500)	(.500)	(.498)	(.495)	(.497)
Extreme prematurity	.590	.581	.495	.369	.146	.037	.010	.003	.001	.012
	(.492)	(.493)	(.500)	(.483)	(.353)	(.188)	(.102)	(.055)	(.025)	(.107)
Prematurity	.158	.198	.261	.364	.527	.611	.592	.350	.108	.188
	(.364)	(.399)	(.439)	(.481)	(.499)	(.488)	(.491)	(.477)	(.311)	(.391)
Twin birth	.172	.173	.186	.211	.229	.260	.303	.229	.058	.107
	(.378)	(.378)	(.389)	(.408)	(.420)	(.439)	(.459)	(.420)	(.235)	(.309)
Multiple birth	.022	.026	.026	.031	.043	.046	.026	.004	.000	.004
	(.145)	(.158)	(.158)	(.172)	(.203)	(.208)	(.159)	(.059)	(.012)	(.060)

continued on next page

Table 2.1: Descriptive statistics by birth weight bracket, 2005-2011 (continued)

	< 600	600-749	750-874	875-999	1000-1249	1250-1499	1500-1999	2000-2499	2500-3000	All births
Health-related controls (continued)										
Asphyxia	.127	.130	.125	.103	.096	.080	.056	.029	.013	.021
	(.333)	(.336)	(.331)	(.304)	(.295)	(.272)	(.229)	(.169)	(.112)	(.145)
Severe asphyxia	.065	.055	.051	.036	.029	.022	.013	.006	.003	.005
	(.247)	(.228)	(.220)	(.187)	(.168)	(.145)	(.115)	(.079)	(.050)	(.071)
Moderate asphyxia	.059	.074	.072	.065	.065	.057	.041	.022	.010	.016
	(.235)	(.262)	(.258)	(.246)	(.246)	(.233)	(.198)	(.147)	(.098)	(.124)
IRDS	.622	.784	.799	.790	.717	.611	.399	.160	.040	.106
	(.485)	(.412)	(.401)	(.408)	(.451)	(.487)	(.490)	(.367)	(.196)	(.308)
Maternal smoking	.013	.019	.023	.029	.027	.030	.037	.027	.006	.012
	(.112)	(.136)	(.150)	(.167)	(.161)	(.170)	(.189)	(.163)	(.076)	(.108)
Maternal alcohol use	XXX	.001	.001	.001	.001	.001	.001	.001	.000	.000
	XXX	(.036)	(.038)	(.034)	(.032)	(.037)	(.038)	(.028)	(.013)	(.019)
Maternal drug use	XXX	.002	.003	.003	.003	.005	.005	.004	.001	.002
	XXX	(.039)	(.056)	(.052)	(.054)	(.069)	(.074)	(.064)	(.038)	(.047)
Hospital-related controls										
In-patient ward	.007	.005	.006	.005	.004	.007	.015	.063	.113	.095
	(.085)	(.067)	(.079)	(.071)	(.065)	(.082)	(.121)	(.244)	(.316)	(.293)
In-patient midwife	.016	.014	.014	.011	.014	.014	.022	.051	.082	.071
	(.124)	(.118)	(.117)	(.103)	(.116)	(.118)	(.145)	(.219)	(.274)	(.256)
In-patient doctor	.007	.006	.007	.005	.005	.007	.015	.063	.113	.095
	(.081)	(.076)	(.083)	(.071)	(.072)	(.083)	(.122)	(.243)	(.316)	(.293)
Number of births	3059	4589	3461	5135	9698	15044	49708	167624	727567	985885

Notes: Standard deviations (in parenthesis) below means. Variable definitions: Procedures: Number of stated OPS codes (= German modification of ICPM); Mortality: Infant death during hospital stay; First-day-mortality: Infant death on first-day of live in hospital; Extreme prematurity: Birth with gestational age < 28 weeks (ICD-10-GM: P07.2); Prematurity: Birth with gestational age > 28 and < 37 weeks (ICD-10-GM: P07.3); Twin birth: (ICD-10-GM: Z38.3); Multiple birth: (ICD-10-GM: Z38.6); Severe asphyxia: Asphyxia with APGAR score below 4 (ICD-10-GM: P21.0); Moderate asphyxia: Asphyxia with APGAR score above 3 and below 8 (ICD-10-GM: P21.1); Asphyxia: Any asphyxia (ICD-10-GM: P21.x); IRDS: Infant respiratory distress syndrome (ICD-10-GM: P22.0-P22.9); Maternal smoking (ICD-10-GM: P04.2), alcohol use (ICD-10-GM: P04.3), or drug use (ICD-10-GM: P04.4): Newborn has been harmed by respective maternal behaviour; In-patient ward: Birth place is in-patient ward; In-patient midwife: Birth with external midwife; In-patient doctor: Birth with external surgeon. **Source:** Own calculations based on the DRG-Statistic.

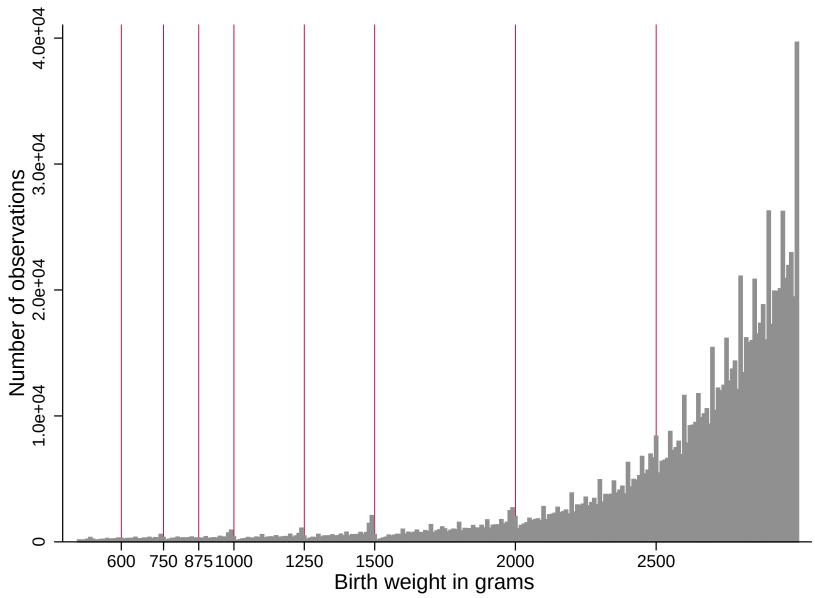
The third part of Table 2.1 shows three measures that capture information on the hospital as well as the medical personnel that delivered the treatment. The first variable is an indicator for whether a newborn was treated in an inpatient ward, defined as a department of the hospital where the staff (doctors, nurses, and midwives) are self-employed rather than employed by the hospital. Even if newborns are not treated in an inpatient ward, care can be partly delivered by self-employed doctors or midwives. Whether a newborn was treated by a self-employed (inpatient) physician or self-employed (inpatient) midwife is captured by the second and third variables. Information on the place of treatment and the doctors and midwives that delivered the treatment is important for our analysis as the incentives for birth-weight manipulation are slightly smaller in inpatient wards and as it is possible that treatment differs when delivered by inpatient doctors and midwives who generally also oversee outpatient treatment of the patient. Overall, 9.5% of newborns are treated exclusively in inpatient wards and by inpatient doctors; 7.1% received care from inpatient midwives.[14] Although inpatient wards and inpatient doctors and midwives are allowed to treat all newborns irrespective of their birth weight, only few inpatient wards are specialized in treating low birth weight newborns. As a consequence, treatment in inpatient wards and by inpatient doctors is rare for newborns with birth weight below 1500g (less than 1% of cases). Even among newborns with weight between 1500g and 1999g only 1.5% are exclusively treated in inpatient wards, compared to 6.3% for those with weight between 2000g and 2499g, and 11.3% in the highest birth weight interval in our data (2500g-3000g). Figure 2.2 provides additional information on the distribution of birth weight in our data. It contains birth weight frequencies pooled across the years 2005–2011. The eight birth weight thresholds that determine DRGs are indicated by the vertical lines. Figure 2.2 shows statistically implausibly large increases in frequencies slightly below the thresholds, especially at 1250g, 1500g, 2000g, and 2500g. These increases are in line with the findings by Jürges and Köberlein (2015) and indicate that hospitals under-report birth weight.

2.4 Empirical strategy

We aim at investigating whether newborns benefit from having a (reported) birth weight below birth weight thresholds defined by the DRG system, medical guidelines

[14] As midwives are mainly present in the hospital during delivery (and delivery is billed on the mother not the newborn), fewer newborns are recorded as having been treated by inpatient midwives than are treated solely in inpatient wards or receive treatment by inpatient doctors.

Figure 2.2: Low birth weight frequencies, 2005-2011



Notes: Distribution of birth weights below 3000g in the years 2005-2011. Red lines mark birth weight thresholds in the G-DRG system. All birth weight bins with less than three observations were omitted due to confidentiality. This does not change the look of the graph. **Source:** Own calculations based on the DRG-Statistic.

or a combination thereof. We therefore document differences in quantity and quality of care comparing newborns with weight just below and just above the eight reimbursement relevant thresholds in Germany. To this end we start by estimating mean differences in 25g weight intervals above and below the thresholds. Quantity of care is measured by the length of the hospital stay and the number of procedures that newborns receive. Quality of care is measured by mortality (during the hospital stay). To exclude the possibility that differences across thresholds can be expected as birth weight is related to health and newborns with lower weight may thus simply need more care even if the difference in weight is not large, we also report results for four placebo thresholds that do not affect reimbursement and do not appear in medical guidelines.

Earlier studies on differences in birth weight thresholds have relied on the assumption that newborns left and right of the analysed thresholds are similar in terms of health (e.g. Almond et al. 2010; Bharadwaj et al. 2013; Breining et al. 2015). The thresholds considered in these studies, however, were not relevant for reimbursement and thus did not give hospitals a financial incentive to manipulate birth weight. As Figure 2.2 and Jürges and Köberlein (2015) (with another data source) show, birth weight manipulation is common in the German DRG system. The assumption that newborns left and right of the thresholds are similar in terms of required care and expected outcomes could therefore only hold if the manipulation of birth weight happened randomly, or at least independent of the newborns' health. Jürges and Köberlein (2015), to the contrary, suggest that manipulation is systematic: there seems to be more under-reporting of birth weight for newborns who are expected to need more care (those who are fragile but not likely to die very early on).

Raw differences in quantity and quality of care around the birth weight thresholds therefore do not constitute causal effects of having a weight below the threshold. To investigate whether at least part of observed difference is likely causal, we proceed by conducting parametric regressions, in which we subsequently add more control variables that account for differences in the likelihood of birth weight manipulation and may explain treatment differences. We implement the following parametric regression equation:

$$Y_i = \alpha + \beta \mathbb{1}(bw_i < T) + f(bw_i - T + 1) + \boldsymbol{\gamma X_i} + u_i \ , \qquad (2.1)$$

where Y_i is an outcome measure for the quality or quantity of care for newborn i, bw_i is the (possibly manipulated) birth weight, and $\mathbb{1}(bw_i < T)$ is an indicator for newborn i having a birth weight below the threshold T. The parameter of interest is β, which measures the difference in outcomes around the threshold. In our baseline regressions we include a (centered) second-order birth weight polynomial $f(bw_i - T + 1)$, fitted separately on both sides of the threshold and use all observations in a 100g bandwidth of the birth weight threshold. In robustness analyses, we vary the bandwidth as well as the order of the included polynomial. In an additional robustness analysis, we exclude newborns in 10g windows left and right of the thresholds – so called "donut regressions", as Barreca et al. (2011) have shown that the exclusion of newborns in small windows around the thresholds may have an influence on the results.

In our baseline analysis, we estimate four different specifications for each outcome. In the first, we estimate equation (2.1) using all observations with weight in the 100g bandwidth and without control variables $\boldsymbol{X_i}$. In the second, we include a vector of control variables, $\boldsymbol{X_i}$ to capture observable differences in the newborns' health at birth and hospital-level controls. If we observed all factors that trigger birth weight manipulation, $\widehat{\beta}$ estimated based on equation (2.1) including our control variables would measure the causal effect of having a weight below the different thresholds on the quantity and quality provided. As, however, it is unlikely that we can account for all factors that hospital staff observes at the time of birth, we conduct a third analysis in which we exclude all newborns who die on their first day of their life. All newborns who die during their first four days of life receive the same DRG irrespective of birth weight. There is thus no incentive to manipulate the weight for newborns for whom death can be expected at the time when birth weight is reported. Newborns who are expected to die are therefore likely over-represented just above the weight thresholds. This may explain lower mortality as well as higher treatment intensity below the thresholds. By excluding newborns who die during their first day of life, we aim at excluding this type of selective upcoding as a reason for differences in care around the thresholds.

In addition to selective upcoding of newborns, differences across hospitals are a possible driver of observed differences around thresholds: For example, it could be the case that well-managed hospitals understand the financial incentives stemming from the DRG system and thus manipulate birth weight, while at the same time they deliver good care. Newborns with (reported) weight below the thresholds would then over-proportionally be treated in hospitals with better care, which would result in average differences in outcomes across thresholds. To exclude this explanation for differences, we include hospital fixed effects in our fourth set of analyses. With the inclusion of hospital fixed effects, we make sure that treatment differences arise within hospitals, not across hospitals.

We further investigate drivers of observed differences in care across thresholds by conducting subgroup analyses splitting the sample into the different severity groups. To the extent that the severity groups succeed in holding health constant across thresholds, differences in received treatment and mortality across thresholds are likely driven by the difference in reimbursement or medical guidelines across thresholds.

To separate reimbursement effects and effects of diagnostic thresholds or medical guidelines, we compare effects across the different thresholds. There are two thresholds, 875g and 2000g that are neither diagnostic thresholds nor appear in any of the relevant medical guidelines. Differences across these thresholds are thus likely driven by reimbursement rather than by medical guidelines.

2.5 Results

Table 2.2 reports differences in means between groups of 25g intervals across the eight reimbursement-relevant birth weight thresholds (600g, 750g, 875g, 1000g, 1250g, 1500g, 2000g, 2500g), as well as for four thresholds that play no role in reimbursement or medical guidelines (700g, 1300g, 2200g, 2700g). In general, newborns with weight just below the relevant thresholds stay longer in the hospital and receive more procedures than their neighbours with weight on the other side of the thresholds. With the exception of an increase in mortality below 875g, the share of newborns who dies during the hospital stay and on their first day of life is smaller below than above the thresholds. At the 600g and 875g threshold hardly any of the differences are significantly different from zero. At the other thresholds most differences are significant.

At least part of the raw difference in length of stay and number of procedures, however, is likely explained by the differences in weight around the thresholds: At all placebo thresholds newborns below the threshold stay significantly longer in the hospital and receive more procedures. Controlling for birth weight as in equation (2.1) is thus crucial when analysing differences in care around the thresholds.

Although differences in the quantity of care exist around placebo thresholds, there are no differences in mortality.[15] This contrasts with the significant decreases in mortality across the reimbursement relevant thresholds (except at 875g). As first day mortality varies even more strongly around the thresholds than overall mortality, a large part of the difference may be driven by selective upcoding.

The mean differences in health-related controls across the thresholds present additional evidence for selective upcoding. In particular at the highest threshold of 2500g, all variables except for extreme prematurity and maternal alcohol use indicate that newborns just below the threshold are in worse health than newborns above. The newborns just below are 3.1 percentage points more likely to be male, 3.3 percentage points more likely to be born prematurely, 3.1 percentage points more like to be twins and 0.1 percentage points more likely to be higher order births. They are more likely to have any type of asphyxia, IRDS or a suspected damage due to maternal smoking or drug use. Only few of these variables show significant differences around the placebo thresholds of 2200 or 2700g, and if the differences are significant they are much smaller in size. This suggests that hospital staff manipulates the weight of at risk newborns – possibly to cover the expected higher costs of treatment. At the other thresholds, the differences in health-related controls are not as drastic. However, they also point into the direction of selective upcoding in favour of the more fragile newborns, highlighting the possibility that selective upcoding may explain differences in treatment and mortality across thresholds.

[15] To the extent that lighter newborns have higher mortality risks ceteris paribus, the additional care that the lighter newborns receive may be effective in decreasing these additional risks.

Table 2.2: Differences above and below threshold, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Outcomes												
Length of stay (days)	3.187	11.364***	2.338	8.629***	2.099	3.135***	4.024***	1.602***	5.589*	2.582**	.758***	.165***
	(3.378)	(2.666)	(1.983)	(1.952)	(1.289)	(1.213)	(.282)	(.078)	(3.077)	(1.064)	(.184)	(.039)
Procedures (#)	.606	1.142**	.656	1.75***	.877***	.618**	1.042***	.737***	.982*	.058	.208***	.049***
	(.611)	(.579)	(.428)	(.444)	(.263)	(.3)	(.079)	(.028)	(.567)	(.269)	(.059)	(.015)
Mortality	-.022	-.088***	.03**	-.033**	-.015*	-.041***	-.007***	0	.008	-.009	0	0
	(.027)	(.022)	(.015)	(.016)	(.009)	(.01)	(.002)	(0)	(.022)	(.008)	(.001)	(0)
First-day-mortality	-.041*	-.065***	.011	-.05***	-.015**	-.043***	-.004**	-.001*	.014	-.001	0	0
	(.023)	(.016)	(.01)	(.015)	(.007)	(.01)	(.002)	(0)	(.016)	(.007)	(.001)	(0)
Health-related controls												
Male births	.004	-.039	-.01	.021	-.003	.029	.021**	.031***	.041	-.035	.007	.002
	(.028)	(.028)	(.026)	(.029)	(.022)	(.023)	(.01)	(.005)	(.026)	(.023)	(.009)	(.004)
Extreme Prematurity	.014	.032	.041	.085***	.01	-.006	0	0	.004	.012	0	0
	(.028)	(.028)	(.026)	(.025)	(.011)	(.007)	(.002)	(.001)	(.026)	(.012)	(.001)	(0)
Prematurity	.001	.021	-.003	.012	-.006	.024	.099***	.033***	-.006	-.043*	.013	.019***
	(.022)	(.023)	(.024)	(.029)	(.022)	(.023)	(.01)	(.005)	(.021)	(.023)	(.009)	(.003)
Twin births	-.004	-.015	-.047**	-.024	-.007	.003	.001	.031***	.019	.015	.006	.015***
	(.021)	(.022)	(.022)	(.024)	(.019)	(.02)	(.01)	(.004)	(.02)	(.02)	(.008)	(.002)
Multiple births	-.005	.008	-.011	-.022*	-.004	-.003	.002	.001***	-.011	.003	.003*	0
	(.008)	(.009)	(.008)	(.013)	(.01)	(.009)	(.002)	(0)	(.008)	(.01)	(.001)	(0)

continued on next page

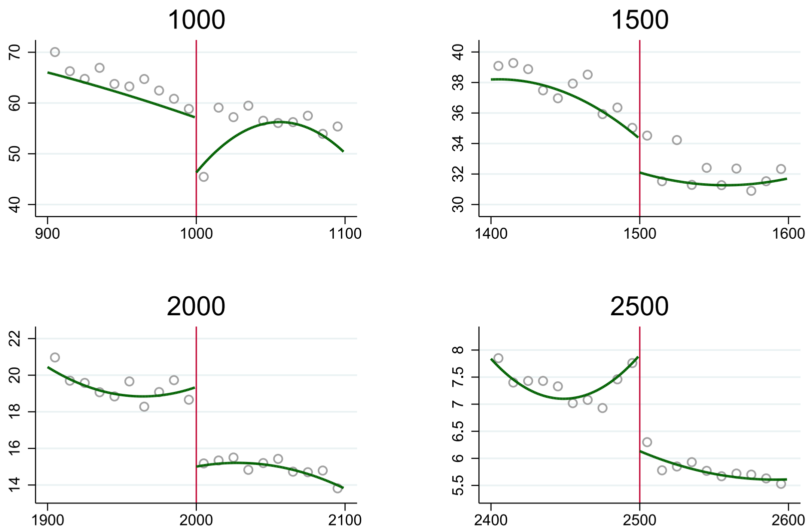
Table 2.2: Differences above and below threshold, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls (continued)												
Asphyxia	.033*	.008	-.006	-.004	.001	.018*	.002	.011***	.031*	-.03**	0	0
	(.02)	(.018)	(.016)	(.017)	(.013)	(.011)	(.005)	(.002)	(.018)	(.012)	(.003)	(.001)
Severe asphyxia	.019	-.013	.005	.003	0	.003	-.002	.002*	.01	-.011	.001	0
	(.014)	(.013)	(.011)	(.01)	(.007)	(.006)	(.002)	(.001)	(.012)	(.007)	(.001)	(0)
Moderate asphyxia	.011	.024*	-.012	-.003	.004	.018**	.003	.009***	.023	-.023**	-.001	0
	(.015)	(.014)	(.013)	(.014)	(.011)	(.009)	(.004)	(.001)	(.014)	(.01)	(.003)	(.001)
IRDS	.046*	.061***	.006	.097***	.024	.119***	.111***	.081***	-.013	.009	.01	.005**
	(.025)	(.023)	(.021)	(.027)	(.021)	(.023)	(.009)	(.003)	(.021)	(.022)	(.007)	(.002)
Maternal smoking	-.006	.005	-.001	.009	.003	.018***	-.001	.007***	-.005	-.002	.001	.002***
	(.008)	(.007)	(.009)	(.009)	(.007)	(.006)	(.004)	(.001)	(.008)	(.007)	(.003)	(.001)
Maternal alcohol use	0	-.004	.001	.002**	-.002	0	0	0	-.001	.002	-.001	0
	(0)	(.003)	(.002)	(.001)	(.002)	(.002)	(.001)	(0)	(.001)	(.002)	(0)	(0)
Maternal drug use	-.002	-.001	.004	0	-.004	0	.002	.003***	0	-.002	.002	0
	(.002)	(.002)	(.003)	(.003)	(.003)	(.003)	(.001)	(.001)	(.002)	(.003)	(.001)	(0)
Hospital-related controls												
In-patient ward	-.001	-.01*	-.006	-.003	-.006*	-.013**	-.024***	-.029***	.001	0	-.008**	-.007**
	(.005)	(.006)	(.004)	(.005)	(.004)	(.006)	(.004)	(.003)	(.003)	(.003)	(.004)	(.003)
In-patient midwife	-.005	-.009	-.009	-.007	-.006	-.001	-.014***	-.019***	.004	-.004	0	-.002
	(.006)	(.007)	(.007)	(.007)	(.006)	(.006)	(.004)	(.003)	(.006)	(.004)	(.003)	(.002)
In-patient doctor	.002	-.002	-.008	.002	.002	-.016**	-.022***	-.03***	.003	-.002	-.008**	-.005*
	(.004)	(.004)	(.005)	(.003)	(.003)	(.006)	(.004)	(.003)	(.004)	(.003)	(.004)	(.003)
N_L	587	1105	880	1881	1957	3881	5665	14432	635	731	5045	20825
N_R	666	446	629	347	693	538	3806	19384	844	1348	8777	37094

Notes: Difference between means below and above birth weight threshold T within a ±25 gram bandwidth ($\bar{X}_{below}$-$\bar{X}_{above}$). (Robust) Standard errors in parentheses below mean difference. N_L: Number of observations left/below threshold within 25 gram bandwidth. N_R: Number of observations right/above threshold within 25 gram bandwidth. *** $p < 0.01$; ** $p < 0.05$ and * $p < 0.1$. For variable definitions see Table 2.1. **Source:** Own calculations based on the DRG-Statistic.

Regression results to facilitate the interpretation of the raw differences in treatment and mortality around thresholds are displayed in Table 2.3. The first row for each outcome displays the estimated coefficient $\widehat{\beta}$ in equation (2.1) when only a second order polynomial in birth weight is included as a control. For the higher birth weight thresholds, this pattern can also be clearly seen by plotting average length of stay and performed procedures around the thresholds (Figures 2.3 and 2.4). The second row in

Figure 2.3: Length of stay (days) at highest thresholds, 2005-2011

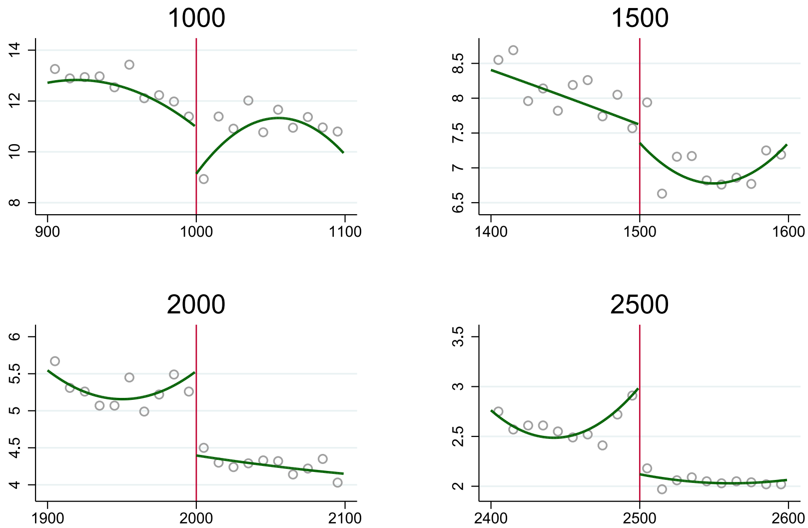


Notes: Circles show mean values in 10 gram intervals. Green lines are fitted values using all observations. Excludes newborns who died on their first day of life. **Source:** Own calculations based on the DRG-Statistic.

Table 2.3 adds the controls for health at birth and hospital-level variables; the third row restricts the data to those newborns who survive their first day of life. The fourth row additionally adds hospital fixed effects to restrict the analysis to within hospital differences.

In line with the comparison of mean differences across reimbursement-relevant and placebo thresholds in Table 2.2, adding controls for birth weight already reduces the number of significant differences in quantity of care. While in Table 2.2, 5 of 8 reimbursement-relevant thresholds showed significant differences in length of stay, and 6 of 8 in the number of procedures, only 4 of the 8 thresholds show significant differences when birth weight is controlled for in Table 2.3. Similarly, the differences at the placebo thresholds are rendered insignificant almost entirely when birth weight is controlled for. The inclusion of health-related and hospital-related controls in row (2) reduces the coefficients for differences in length of stay and number of procedures across

Figure 2.4: Number of procedures at highest thresholds, 2005-2011



Notes: Circles show mean values in 10 gram intervals. Green lines are fitted values using all observations. Excludes newborns who died on their first day of life. **Source:** Own calculations based on the DRG-Statistic.

the reimbursement-relevant thresholds by almost half on average. The coefficients for the placebo thresholds remain almost unchanged. When additionally restricting the analysis to first-day-survivors (row (3)), the coefficients are further reduced towards zero and only 3 out of 8 remain significant for length of stay and 4 of 8 for number of procedures. At least half of the difference in treatment across thresholds thus seems to reflect selective upcoding. When we focus at within hospital differences in row (4), only the differences across the highest weight thresholds (2000g and 2500g) remain positive and significantly different form zero. The differences around the other thresholds are not significant or sometimes even turn negative. Within hospitals, there thus seems to be very little difference in treatment of at risk newborns due to weight thresholds when observable health differences are controlled for. Only for heavier newborns (for whom treatment may not be as decisive as the risk of mortality is much lower) differences persist.

Table 2.3: Regression estimates, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Length of stay (days)												
(1)	7.963	13.542***	2.055	11.249***	2.604	2.266	4.334***	1.733***	5.353	2.973*	.244	.039
	(5.141)	(3.9)	(2.986)	(2.757)	(1.893)	(1.745)	(.415)	(.126)	(4.441)	(1.574)	(.297)	(.062)
(2)	2.821	8.778**	1.117	7.584***	2.8	.677	2.753***	.911***	5.869	2.83*	.155	-.009
	(4.636)	(3.598)	(2.726)	(2.454)	(1.777)	(1.682)	(.393)	(.117)	(4.245)	(1.48)	(.276)	(.058)
(3)	-2.824	5.149	1.876	5.452**	1.722	-.899	2.707***	.898***	6.218	2.653*	.153	-.009
	(5.507)	(3.684)	(2.703)	(2.49)	(1.779)	(1.736)	(.394)	(.117)	(4.362)	(1.464)	(.276)	(.058)
(4)	-9.439*	4.206	-.078	-.457	-.883	-3.446*	1.451***	.834***	5.877	2.79**	-.053	.014
	(5.389)	(3.567)	(2.465)	(2.156)	(1.746)	(1.774)	(.367)	(.113)	(4.132)	(1.322)	(.258)	(.057)
Procedures (#)												
(1)	1.407	1.427*	.647	1.959***	1.049***	.246	1.122***	.859***	1.003	-.296	.164*	.011
	(.945)	(.78)	(.689)	(.55)	(.382)	(.455)	(.123)	(.045)	(.912)	(.394)	(.091)	(.024)
(2)	.234	.559	.486	1.16**	1.02***	-.225	.61***	.449***	1.01	-.171	.135*	-.014
	(.822)	(.708)	(.64)	(.487)	(.354)	(.436)	(.112)	(.039)	(.851)	(.368)	(.081)	(.021)
(3)	-.544	-.056	.566	.87*	.899**	-.487	.605***	.448***	.843	-.186	.132	-.015
	(1.006)	(.758)	(.656)	(.512)	(.363)	(.456)	(.112)	(.039)	(.92)	(.371)	(.081)	(.021)
(4)	-.596	-.221	.44	-.121	.508	-.569	.354***	.397***	1.145	-.083	.112	-.024
	(.964)	(.707)	(.591)	(.475)	(.339)	(.455)	(.105)	(.036)	(.82)	(.336)	(.075)	(.019)
Mortality												
(1)	-.086**	-.134***	.008	-.042*	-.032**	-.059***	-.011***	-.002**	-.038	-.016	0	0
	(.041)	(.031)	(.022)	(.023)	(.014)	(.015)	(.003)	(.001)	(.033)	(.012)	(.002)	(0)
(2)	-.06	-.108***	.009	-.036	-.03**	-.056***	-.01***	-.002**	-.046	-.016	0	0
	(.038)	(.03)	(.022)	(.023)	(.014)	(.015)	(.003)	(.001)	(.032)	(.012)	(.002)	(0)
(3)	.015	-.044*	-.002	.026**	-.004	0	-.005***	0	-.04	-.01	0	0
	(.038)	(.025)	(.018)	(.012)	(.008)	(.007)	(.002)	(.001)	(.028)	(.008)	(.001)	(0)
(4)	.024	-.031	0	.017	.002	-.001	-.005**	0	-.046	-.012	.001	0
	(.04)	(.026)	(.018)	(.014)	(.008)	(.007)	(.002)	(.001)	(.03)	(.008)	(.001)	(0)
Number of births												
(1)+(2)+(3) - N_L	2002	3326	3015	4506	5264	8728	16919	57475	2639	4717	23288	99190
N_R	2639	2581	3254	2470	3781	3938	15667	71505	2990	4892	30853	131754
(4) - N_L	1333	3017	2871	4366	5166	8642	16833	57397	2198	4626	23222	99129
N_R	2198	2424	3140	2388	3697	3867	15604	71449	2773	4803	30792	131682

Notes: Parametric regressions within ±100 gram intervals around birth weight thresholds. Coefficient of binary variable indicating observations below threshold. (Robust) Standard errors in parentheses. N_L: Number of observations left to/below the threshold . N_R: Number of observations right to/above the threshold. All regressions include a second order birth weight polynomial (fitted separately on each side of the threshold). Control variables are gender, (extreme) prematurity, multiples, asphyxia, IRDS, in-patient ward/midwife/doctor, maternal smoking/alcohol consumption/ drug use and years. [Specification] (1) includes only the binary birth weight threshold indicator variable . (2) additionally includes controls. (3) additionally restricts the sample to first-day-survivors. (4) additionally includes hospital-fixed effects. *** $p < 0.01$; ** $p < 0.05$ and * $p < 0.1$. For variable definitions see Table 2.1. **Source:** Own calculations based on the DRG-Statistic.

The share of newborns who die during the hospital stay is significantly lower below all but the 875g threshold when birth weight is included as control as can be seen in the first row of the mortality results presented in Table 2.3. The inclusion of health-related and hospital control variables shrinks some coefficients slightly towards zero but the differences at most thresholds remain statistically significantly different from zero. As row (3) of the results indicates, however, the observable differences in mortality are largely driven by first day mortality. When newborns who die on their first day of life are excluded from our data, only the differences at 750g and 2000g remain negative and statistically significantly different from zero. To the extent that medical personnel expects the early death of newborns who die during their first day of life and thus do not upcode their birth weight, these results suggest that selective upcoding of newborns drives almost all of the observed differences in mortality. These results remain almost unchanged when hospital fixed effects are included (row 4). Differences across hospitals thus do not seem to play a big role.

Table 2.4 displays the $\widehat{\beta}$s estimated for the different severity groups including polynomials in birth weight as well as all control variables and hospital fixed effects and excluding those newborns who die on their first day of life from the data. The results indicate that the differences in quantity of care observed for the higher thresholds (2000g and 2500g) in Table 2.3 are concentrated in severity groups 1 and 2, i.e., those newborns with the least complications and lowest level of severity. The differences within the higher severity groups are not statistically significant, which may, however, also reflect smaller sample sizes for the higher severity groups. Table 2.4 additionally shows significant differences in quantity of care for the severity groups around the 1000g threshold. At the same time, mortality is higher below the 1000g threshold in these severity groups. The reason for this is most likely that hospitals do get extra reimbursement for long ventilation hours if newborns are heavier than 1000g. Here, hospitals have an incentive *not* to upcode newborns that have a high survival probability with long expected ventilation hours.

Concerning the question whether reimbursement relevant thresholds alone trigger additional care for newborns with weight below the threshold, we look at the results for the 875g and 2000g thresholds that affect reimbursement but are not part of medical guidelines. For 875g, none of the differences are significantly different from zero – not even the raw differences in Table 2.2. For 2000g, some differences are significant and remain significant even in the specifications that take selective upcoding and differences across hospitals as well as (ex post) health severity into account in the subgroup analyses. However, there are no differences for the more severe cases in severity groups 4 to

6. While we thus cannot exclude that newborns with weight below the 2000g threshold receive some additional care due to the increase in reimbursement, this effect seems to be concentrated to newborns with the best health, suggesting that among those newborns who require care it is delivered independent of their weight. Furthermore, the results for the 1500g threshold – the reimbursement relevant threshold that is at the same time a diagnostic threshold and that appears most often in medical guidelines – indicates that adding the financial incentives to the possible medical indication to additionally treat newborns with weight below this threshold has likely no detrimental effect on newborns with weight above the threshold.

Overall, the results presented in this section show that the differences in treatment across birth weight thresholds are mainly driven by selective upcoding as are differences in mortality. Differences in treatment, however, remain across the diagnostic threshold for extremely low birth weight (1000g) as well as for relatively healthy and heavy newborns (severity groups 1 and 2 at 1500g, 2000g, and 2500g). The additional quantity of treatment does not seem to result in higher survival rates. It may - of course - have other health benefits for the newborns later on in life; an outcome that we cannot observe in our data.

Table 2.4: Regression estimates by severity group for first-day-survivors, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Length of stay (days)												
Severity group 1						8.398*	1.592***	.186***			.097	.045*
						(4.679)	(.43)	(.046)			(.16)	(.027)
Severity group 2						4.086***	1.287***	1.162***			-.012	.122
	-4.432	5.056	.323	12.425***	-.522	(1.196)	(.336)	(.174)	-2.688	.396	(.257)	(.141)
Severity group 3	(6.235)	(3.454)	(2.499)	(3.503)	(1.264)	-.713	1.125*	.642	(4.113)	(1.275)	.206	.724**
						(1.362)	(.618)	(.418)			(.611)	(.367)
Severity group 4						-8.507***	1.232	-.495			-.646	1.828
						(2.941)	(1.352)	(1.212)			(1.484)	(1.296)
Severity group 5				12.98**	-7.318**	-26.351	-1.261	.235		8.823***	-4.425	-4.647
	-12.335	3.665	-.964	(5.892)	(3.289)	(18.792)	(3.652)	(4.099)	7.472	(3)	(4.388)	(3.774)
Severity group 6	(9.098)	(7.322)	(6.782)	12.206**	-5.219	-1.316	2.23	3.072	(7.939)	5.08	-3.022	-8.268*
				(4.986)	(3.938)	(5.506)	(4.894)	(7.599)		(3.125)	(6.469)	(4.734)
Procedures (#)												
Severity group 1						3.001**	.514***	.06***			.019	.001
						(1.209)	(.093)	(.012)			(.033)	(.007)
Severity group 2						1.24***	.271***	.276***			.084	.027
	.239	.381	.171	2.077***	.417	(.236)	(.073)	(.049)	-.78	-.21	(.057)	(.048)
Severity group 3	(.748)	(.55)	(.437)	(.598)	(.271)	-.403	.133	.503***	(.571)	(.275)	.139	.093
						(.293)	(.153)	(.117)			(.151)	(.108)
Severity group 4						-1.509***	.229	-.123			.309	.41
						(.468)	(.328)	(.383)			(.335)	(.423)
Severity group 5				7.874***	-.25	-3.24	-.428	-1.672		.433	1.821	-.722
	-2.143	-.976	1.064	(1.265)	(.667)	(3.717)	(1.108)	(1.521)	2.678*	(.69)	(1.275)	(1.577)
Severity group 6	(1.78)	(1.49)	(1.808)	5.153***	.188	.277	-.327	.117	(1.554)	.996	.292	-5.389***
				(1.122)	(.708)	(1.676)	(1.37)	(1.888)		(.82)	(2.576)	(1.809)
Mortality												
Severity group 1						-.001	0	0			0	0
						(.002)	(0)	(0)			(0)	(0)
Severity group 2						0	0	0			0	0
	.034	-.028	-.015	.011	.002	(.001)	(0)	(0)	-.018	-.003	(0)	(0)
Severity group 3	(.049)	(.028)	(.019)	(.022)	(.004)	-.002	0	0	(.035)	(.003)	0	0
						(.003)	(0)	(0)			(0)	(0)
Severity group 4						-.002	.004	-.023**			-.004	0
						(.003)	(.005)	(.01)			(.003)	(.003)
Severity group 5				.197***	.002	.008	0	.018		-.002	-.001	0
	.046	-.074	.009	(.042)	(.009)	(.015)	(0)	(.015)	-.029	(.023)	(.007)	(0)
Severity group 6	(.07)	(.055)	(.054)	.158***	-.007	.039*	-.043	.047	(.057)	-.022	-.063	.011
				(.037)	(.019)	(.021)	(.052)	(.04)		(.019)	(.044)	(.018)
Number of births												
1 - N_L						5400	1337	33587			6115	76359
N_R						34	3211	51886			11585	107440
2 - N_L						5400	6902	13677			9036	12098
N_R	785	2030	2111	3536	2061	1021	6305	10332	1354	2025	10613	13249
3 - N_L	1354	1734	2525	650	1844	5400	3829	4754	1930	2536	3674	5591
N_R						1046	2736	4897			3993	5651
4 - N_L						5400	1590	1560			1326	1342

continued on next page

Table 2.4: Regression estimates by severity group for first-day-survivors, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
N_R						618	1001	1155			1334	1321
5 - N_L				961	799	984	383	253		688	247	167
N_R	627	1154	833	359	525	212	203	142	947	656	220	143
6 - N_L	947	754	738	961	1114	1176	662	298	983	953	226	223
N_R				553	694	298	187	192		875	252	234

Notes: Parametric regressions within ±100 gram intervals around birth weight thresholds by severity group for first-day-survivors. Coefficient of binary variable indicating observations below threshold. (Robust) Standard errors in parentheses. N_L: Number of observations left to/below the threshold . N_R: Number of observations right to/above the threshold. Control variables include a second order birth weight polynomial (fitted separately on each side of the threshold), hospital-fixed effects and dummy variables for: gender, (extreme) prematurity, multiples, asphyxia, IRDS, in-patient ward/midwife/doctor, maternal smoking/alcohol consumption/ drug use and years. *** $p < 0.01$; ** $p < 0.05$ and * $p < 0.1$. For variable definitions see Table 2.1. **Source:**. Own calculations based on the DRG-Statistic.

2.6 Sensitivity Analyses

In this section we present results of three sets of sensitivity analyses. First, we explore the sensitivity of the results to a so-called donut regression and thus to excluding from our data newborns with weight very close to the thresholds. Second, we explore whether the choice of the polynomial in birth weight affects the results, and third whether the choice of bandwidth has an impact. All results presented in Table 2.5 are variations of the most comprehensive specification presented in Table 2.3 including controls for birth weight, health-related as well as hospital-related control variables, hospital fixed effects and excluding newborns who die on their first day of life.

The results of the donut regressions are presented in the first rows of results for each of the three outcomes (length of stay, number of procedures, and mortality). Compared to the main results in Table 2.3, excluding newborns in a 10g interval around each threshold induces only minor changes for the quantity of care but renders the mortality results insignificant and even turns around some of the signs. For length of stay, the previously significantly negative coefficients at the 600g and 1500g thresholds become insignificant when newborns with weight in a 10g interval at each side of the respective thresholds are excluded, while the positive coefficients at 2000g and 2500g even increase in value and stay significantly different from zero. The estimates for the number of procedures are hardly affected at all by the exclusion of the donuts in birth weight. In terms of mortality, the only coefficient that was significant in the main results at 2000g becomes insignificant when the donut is excluded. At the same time, differences at 1000g and 1500g become significantly positive. Overall, excluding the donut around the thresholds thus supports the conclusion that differences in treatment exists if at all only at the highest weight thresholds and that these differences do not affect mortality.

The second and third rows for each outcome in Table 2.5 report the results with different orders of the polynomial in birth weight. While we use a second order polynomial in the main specification displayed in Table 2.3, the results in row (2) of Table 2.5 are based on a linear fit and the results in row (3) on a third order polynomial. While there are slight changes in the estimates of the differences across thresholds, changing the order of the polynomial has in general only limited effects on our results. The results for the highest threshold (2500g) are particularly stable, indicating that – irrespective of the order of the polynomial included – newborns with weight below the threshold receive significantly more care than newborns above, while there are no differences in mortality. The results for the next highest threshold (2000g) are also very stable for mortality suggesting reduced mortality below the threshold. The differences in the

Table 2.5: Regression estimates - Robustness, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Length of stay (days)												
(1)	-13.927	4.157	-2.075	2.877	-1.965	-1.722	2.941***	1.05***	5.905	3.982*	.328	-.044
	(9.022)	(5.9)	(3.987)	(3.458)	(2.247)	(1.873)	(.627)	(.155)	(7.966)	(2.209)	(.402)	(.089)
(2)	-4.953	.591	.26	-1.645	-.796	-.716	1.034***	.648***	5.858**	1.024	.072	.028
	(3.532)	(2.264)	(1.659)	(1.404)	(1.083)	(.993)	(.25)	(.069)	(2.767)	(.931)	(.164)	(.035)
(3)	-8.751	9.253*	2.089	-3.183	-.718	-4.214*	.541	.786***	6.894	.446	-.148	.12
	(7.83)	(4.752)	(3.351)	(2.78)	(2.233)	(2.347)	(.505)	(.167)	(5.26)	(1.953)	(.372)	(.091)
(4)	-5.3	6.527	4.274	-1.636	.697	-5.259*	1.048*	.797***	5.416	.859	-.436	.089
	(8.25)	(5.109)	(3.718)	(3.171)	(2.431)	(2.701)	(.556)	(.182)	(5.67)	(2.265)	(.39)	(.1)
(5)	-6.541	1.279	-.352	-1.488	-.386	-.202	.841***	.509***	6.33	.178	.006	.043
	(4.166)	(3.025)	(2.193)	(1.352)	(.891)	(.448)	(.168)	(.049)	(4.541)	(1.941)	(.164)	(.037)
Procedures (#)												
(1)	-2.305	1.007	-.387	-.339	.151	.409	.91***	.471***	.814	.272	.202*	.007
	(1.614)	(1.234)	(.776)	(.875)	(.543)	(.421)	(.161)	(.051)	(1.32)	(.551)	(.114)	(.03)
(2)	.084	-.146	.25	-.174	.164	.073	.31***	.281***	1.093**	-.16	.035	-.002
	(.653)	(.462)	(.364)	(.338)	(.23)	(.244)	(.069)	(.022)	(.538)	(.224)	(.048)	(.012)
(3)	.429	-.173	.734	-.128	.716	-1.04*	.02	.41***	1.765	-.43	.048	-.037
	(1.347)	(.924)	(.833)	(.532)	(.436)	(.621)	(.149)	(.053)	(1.166)	(.481)	(.109)	(.029)
(4)	.579	-.455	1.108	.12	.901*	-1.107	.218	.449***	.883	-.463	-.04	-.029
	(1.459)	(.967)	(.894)	(.593)	(.479)	(.718)	(.163)	(.058)	(1.296)	(.53)	(.118)	(.031)
(5)	-.364	-.361	.379	-.433	.201	.039	.239***	.208***	1.062	-.16	.045	-.004
	(.735)	(.61)	(.512)	(.322)	(.202)	(.109)	(.045)	(.015)	(.894)	(.46)	(.048)	(.013)
Mortality												
(1)	.045	-.039	-.02	.04*	-.01	.015**	-.001	0	0	.003	0	0
	(.07)	(.044)	(.029)	(.022)	(.016)	(.007)	(.003)	(.001)	(.051)	(.014)	(.002)	(0)
(2)	-.003	-.027	.007	.009	.006	-.002	-.004***	0	-.008	-.014**	-.001	0
	(.027)	(.017)	(.013)	(.009)	(.005)	(.004)	(.001)	(0)	(.019)	(.006)	(.001)	(0)
(3)	.033	-.056	.018	.004	.007	-.008	-.006**	.001	-.064	-.023**	0	0
	(.057)	(.035)	(.025)	(.018)	(.01)	(.01)	(.003)	(.001)	(.04)	(.01)	(.002)	(0)
(4)	.037	-.018	.025	-.005	.016	-.013	-.006*	.001	-.069	-.024**	-.001	0
	(.06)	(.037)	(.027)	(.021)	(.011)	(.011)	(.003)	(.001)	(.043)	(.011)	(.002)	(0)
(5)	.012	-.027	.005	.018**	-.001	0	-.002**	0	-.063*	-.016*	.001	0
	(.031)	(.023)	(.016)	(.009)	(.005)	(.002)	(.001)	(0)	(.033)	(.008)	(.001)	(0)

Notes: Parametric regressions within different intervals around birth weight thresholds always include control variables, hospital-fixed effects and are restricted to first-day-survivors. Coefficient of binary variable indicating observations below threshold. (Robust) Standard errors in parentheses. N_L: Number of observations left to/below the threshold . N_R: Number of observations right to/above the threshold. Control variables include gender, (extreme) prematurity, multiples, asphyxia, IRDS, in-patient ward/midwife/doctor, maternal smoking/alcohol consumption/ drug use and years. [Specification] (1) uses a second order birth weight polynomial (fitted separately on each side of the threshold) and includes observations within a ±100 gram bandwidth around the threshold, but excludes observations within a ±10 gram bandwidth around the threshold. (2) uses a first order birth weight polynomial (fitted separately on each side of the threshold) and includes observations within a ±100 gram bandwidth around the threshold. (3) uses a third order birth weight polynomial (fitted separately on each side of the threshold) and includes observations within a ±100 gram bandwidth around the threshold. (4) uses a second order birth weight polynomial (fitted separately on each side of the threshold) and includes observations within a ±50 gram bandwidth around the threshold. (5) uses a second order birth weight polynomial (fitted separately on each side of the threshold) and includes all observations between two thresholds. *** $p < 0.01$; ** $p < 0.05$ and * $p < 0.1$. For variable definitions see Table 2.1. **Source:**. Own calculations based on the DRG-Statistic.

quantity of care, however, become insignificant with a third order polynomial. For the lower thresholds, there are hardly any significant differences in any of the outcomes irrespective of the order of the polynomial included.

The fourth and fifth rows for each outcome in Table 2.5 report results when changing the bandwidth. The main results are based on a 100g bandwidth. Row (4) in Table 2.5 reports results for a 50g bandwidth, row (5) using the entire birth weight interval, e.g. 450-599g below 600g, and 600-749g above 600.[16] Again, there are only minor changes in the coefficient estimate of interest due to the changes in bandwidth. Similar to the other sensitivity analyses, the results are extremely robust for the highest threshold (2500g). But also at the lower thresholds, only minor changes occur.

Overall, the results presented in Table 2.5 highlight that our results are not driven by newborns that have weights very close to the thresholds, nor by the choice of the order of the polynomial in birth weight or by bandwidth choice. All sensitivity analyses indicate that – despite controlling for an extensive set of variables related to health and the place as well as the people who deliver the care, hospital fixed effects and excluding newborns who die very early on – newborns below the highest weight threshold receive more care than their immediate neighbours with weight at or above the threshold. The results for differences in care around the next highest threshold (2000g) are somewhat more sensitive to the exact specification but also indicate overall that newborns below the threshold receive additional care. For these groups of newborns, mortality differences are also observed in most specifications. For all lower thresholds, hardly any of the differences are significantly different from zero – irrespective of the specification.

2.7 Discussion and Conclusion

In this chapter, we investigate whether birth weight thresholds in hospital reimbursement systems affect the quantity and quality of care delivered to newborns. Using the universe of hospital births in Germany from the years 2005–2011, we document that newborns with weight below all but one reimbursement relevant threshold receive more care and have lower risk of dying during the hospital stay than their neighbours with weight at or above the respective thresholds. For all but the highest weight threshold (2500g) the differences in care around thresholds, however, do not remain significant when controlling for birth weight and differences in health around the thresholds that

[16] For the placebo thresholds, we use all observations in the interval between the reimbursement relevant thresholds in which the placebo threshold lies. E.g. for the threshold of 700g, we use 600–749g, and for 1300g, we use 1250–1499g.

are observable to us as researchers and also to the medical personnel that reports birth weight. The health differences therefore likely result from selective birth weight manipulation as medical staff in hospitals tends to under-report a newborn's weight if higher costs are expected. Overall, our results suggest that – if at all – only the heaviest newborns (in our sample with birth weight just below 2000g or 2500g) and among those the group with the least severe health conditions as measured by the assigned DRG stay longer in the hospital and receive more procedures because their weight is below the threshold.

These results lead to the questions what drives the differences in care around 2000g and 2500g and why there are no (robust) differences around the other thresholds. A likely explanation is that we do not account for selective upcoding well enough for the higher threshold. The differences in care for the lower thresholds mainly disappear when the analyses are restricted to first-day-survivors. As mortality and first day mortality decrease with weight, whether newborns are expected to survive or not is likely not the margin that determines upcoding for higher weight newborns. Instead, other measures such as gestational age may be more important. Jürges and Köberlein (2015) for example find that newborns in the 25g weight interval below the 2500g threshold have on average almost four fewer days of gestational age compared to newborns with weight of 2500-2525g. As in our data we only observe broad categories of gestational age, we cannot fully control for these differences. While for the lower weight groups this may not matter as differences in gestational age translate into early mortality (which we observe and control for), this is not necessarily the case for the high weight thresholds. To us, the most plausible explanation for the remaining differences in quantity of care around the higher weight thresholds is that we cannot fully control for selective upcoding.

Of course, the differences in care around the higher thresholds could also reflect actual differences, e.g. triggered by the additional reimbursement or the fact that newborns with weight below 2500g receive a diagnosis of low birth weight, which may possibly increase the attention of hospital staff. However, it is unclear why similar differences should not be observed for the lower weight thresholds, specifically as the latter also trigger diagnoses (e.g. 1500g and 1000g) and result in much larger differences in reimbursement for the hospital. Overall, we thus judge it to be more likely that neither reimbursement differences nor medical guidelines or diagnostic thresholds induce differences in neonatal care in German hospitals.

In light of the evidence from other countries concerning the benefits of having a birth weight just below 1500g and as reimbursement differences should if at all only add to the existing differences caused by the diagnostic threshold, these result may seem surprising. However, Almond et al. (2010) show that the differences in neonatal mortality and care are not present in all hospitals. Instead the effects are driven by low quality ones. Furthermore, medical guidelines in Germany give only very few recommendations that depend on birth weight, which may explain differences compared to other countries like Chile and Denmark where explicit recommendations with respect to birth weight exist (Bharadwaj et al. 2013; Breining et al. 2015). Last but not least, like in the other studies on upcoding (Dafny 2005; Verzulli et al. 2016), higher reimbursement only depends on the reporting of characteristics or diagnoses – in this case birth weight – not on the treatment that is delivered. Our findings of little changes in quantity and quality of care thus align with the earlier findings that additional resources acquired through upcoding do not profit the specific patients whose records are manipulated.

To conclude, we interpret our results in a way that financial incentives relating to birth weight in Germany do lead to birth weight manipulation but do not directly impact the care that specific newborns receive. This suggests that hospital staff is willing to manipulate records according to their employers' financial incentives but does not take the implications of these incentives (higher funds available for the specific case) into account when making critical medical decisions. This finding is in line with physicians taking treatment decisions in the interest of their patients.

2.8 Appendix

Table 2.6: DRG-severity groups for newborns in Germany, 2005-2011

2005

	< 600	600 - 749	750 - 874	875 - 999	1000 - 1249	1250 - 1499	1500 - 1999	2000 - 2499	> 2499
Severity group 1	P61A	P61B	P62A	P62B	P63Z	P64Z	P65D	P66D	P67D
Severity group 2							P65C	P66C	P67C
Severity group 3							P65B	P66B	P67B
Severity group 4							P65A	P66A	P67A
Severity group 5					P03D		P04C	P05C	P06C
Severity group 6					P03B		P04B	P05B	P06B

2006-2010

	< 600	600 - 749	750 - 874	875 - 999	1000 - 1249	1250 - 1499	1500 - 1999	2000 - 2499	> 2499
Severity group 1	P61B	P61D	P62B	P62D	P63Z	P64Z	P65D	P66D	P67D
Severity group 2							P65C	P66C	P67C
Severity group 3							P65B	P66B	P67B
Severity group 4							P65A	P66A	P67A
Severity group 5	P61A	P61C	P62A	P62C	P03C		P04C	P05C	P06C
Severity group 6					P03B		P04B	P05B	P06B

2011

	< 600	600 - 749	750 - 874	875 - 999	1000 - 1249	1250 - 1499	1500 - 1999	2000 - 2499	> 2499
Severity group 1	P61B	P61D	P62B	P62C	P63Z	P64Z	P65D	P66D	P67D
Severity group 2							P65C	P66C	P67C
Severity group 3							P65B	P66B	P67B
Severity group 4							P65A	P66A	P67A
Severity group 5	P61A	P61C	P62A		P03C		P04C	P05C	P06C
Severity group 6					P03B		P04B	P05B	P06B

Notes: Severity groups and birth weight thresholds in grams. Within each severity group, birth weight is the only grouping criterion. This table is based on simulations, performed using the G-DRG Webgrouper: http://drg.uni-muenster.de/index.php?option=com_webgrouper&view=webgrouper&Itemid=26 [last accessed: 17 September 2017].

Table 2.7: Differences above and below threshold by severity group, 2005-2011

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Outcomes												
Length of stay (days)												
Severity group 1						23.647***	3.071***	.263***			.313***	.067***
						(2.96)	(.324)	(.033)			(.119)	(.019)
Severity group 2						6.991***	1.775***	1.7***			.515***	.285***
	-2.431	5.338**	3.098*	16.178***	1.631*	(.883)	(.247)	(.126)	-1.09	1.534*	(.189)	(.104)
Severity group 3	(3.838)	(2.492)	(1.684)	(2.893)	(.938)	1.595*	1.623***	1.455***	(2.76)	(.886)	.953**	1.12***
						(.933)	(.429)	(.307)			(.41)	(.271)
Severity group 4						-7.452***	2.305***	.52			.122	1.999**
						(2.228)	(.849)	(.781)			(.947)	(.976)
Severity group 5				11.136***	-1.901	-11.935	-3.819*	-2.74		4.492**	-1.825	-5.328***
	-4.618	7.211	5.525	(3.23)	(2.127)	(10.246)	(2.236)	(2.905)	9.657*	(1.945)	(2.215)	(1.821)
Severity group 6	(5.685)	(4.598)	(4.281)	14.056***	-1.721	3.246	1.282	-.784	(5.811)	3.583*	-3.681	-3.893
				(3.229)	(2.432)	(3.023)	(2.83)	(3.499)		(2.168)	(2.888)	(2.535)
Others	3.019	1.758	.817	-38.702***	5.906	-6.832*	.18	-.418	-.681	3.456	.362	-.377
	(2.169)	(2.775)	(.768)	(3.457)	(4.786)	(3.653)	(1.419)	(.923)	(2.1)	(4.189)	(1.137)	(.749)
Procedures (#)												
Severity group 1						5.233***	.671***	.082***			.026	.019***
						(.227)	(.069)	(.01)			(.028)	(.006)
Severity group 2						2.085***	.335***	.459***			.139***	.106***
	-.132	.477	.617*	3.702***	.438**	(.187)	(.057)	(.04)	-.1	-.148	(.046)	(.04)
Severity group 3	(.533)	(.445)	(.338)	(.467)	(.215)	-.25	.209*	.947***	(.447)	(.224)	.089	.209**
						(.222)	(.115)	(.091)			(.115)	(.088)
Severity group 4						-1.461***	.253	.318			-.01	.341
						(.374)	(.236)	(.243)			(.242)	(.287)
Severity group 5				5.917***	.134	-.986	-.712	-.14		-.019	.267	-.969
	-.636	-.546	1.569	(.682)	(.485)	(2.049)	(.872)	(.809)	1.693	(.476)	(.651)	(.787)
Severity group 6	(1.18)	(1.156)	(1.169)	4.159***	.723	.208	-.492	-.883	(1.105)	1.202**	-.545	-1.49**
				(.727)	(.536)	(1.417)	(.79)	(1.011)		(.574)	(1.094)	(.719)
Others	.598	1.413	.588	-5.452***	1.399	-1.772*	-.092	.223	-.356	-.053	.112	-.185
	(.58)	(1.303)	(.618)	(.904)	(.924)	(1.003)	(.4)	(.297)	(.708)	(1.028)	(.379)	(.267)
Mortality												
Severity group 1						.001*	0	0			0	0
						(.001)	(0)	(0)			(0)	(0)
Severity group 2						.001*	0	0			0	0
	.026	-.031	.005	.011	.001	(.001)	(0)	(0)	.015	-.003	(0)	(0)
Severity group 3	(.031)	(.019)	(.012)	(.015)	(.001)	.001*	0	0	(.023)	(.002)	0	0
						(.001)	(0)	(0)			(0)	(0)
Severity group 4						.001*	.002	-.019**			-.005	0
						(.001)	(.002)	(.008)			(.003)	(0)
Severity group 5				.116***	.003	.002	0	-.009		-.001	0	0
	.027	-.082**	.04	(.017)	(.003)	(.002)	(0)	(.027)	.017	(.013)	(0)	(0)
Severity group 6	(.043)	(.037)	(.035)	.116***	-.004	.023***	-.018	.021	(.036)	-.008	-.014	0
				(.017)	(.016)	(.007)	(.023)	(.015)		(.014)	(.029)	(0)
Others	.016	-.08	.115*	.07	-.076**	-.145***	-.036***	-.009	.014	-.022	.003	-.006
	(.041)	(.063)	(.068)	(.042)	(.036)	(.039)	(.013)	(.008)	(.057)	(.038)	(.012)	(.008)

continued on next page

Table 2.7: Differences above and below threshold by severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls												
Male births												
Severity group 1						-.051	-.034	.004			-.015	0
						(.145)	(.029)	(.007)			(.015)	(.005)
Severity group 2						.078*	.008	.021			.017	.004
	.055	-.063*	-.037	.065	0	(.045)	(.017)	(.013)	.014	-.043	(.014)	(.013)
Severity group 3	(.041)	(.036)	(.03)	(.061)	(.033)	-.003	0	.041**	(.034)	(.032)	-.003	.003
						(.044)	(.024)	(.019)			(.023)	(.02)
Severity group 4						-.032	.006	.062*			-.013	-.012
						(.058)	(.04)	(.036)			(.039)	(.041)
Severity group 5				.161**	-.107*	-.028	.117	-.071		.049	-.074	-.067
	.016	-.015	.001	(.073)	(.056)	(.093)	(.082)	(.09)	.019	(.062)	(.107)	(.108)
Severity group 6	(.048)	(.049)	(.053)	.045	.018	.074	.092	-.073	(.043)	.011	.039	-.104
				(.067)	(.052)	(.092)	(.082)	(.086)		(.053)	(.091)	(.098)
Others	-.048	.029	.071	-.021	.004	.036	-.028	-.023	.091	-.099*	.016	.012
	(.05)	(.064)	(.07)	(.051)	(.047)	(.05)	(.028)	(.024)	(.058)	(.053)	(.028)	(.024)
Extreme Prematurity												
Severity group 1						.012***	0	0			0	0
						(.002)	(0)	(0)			(0)	(0)
Severity group 2						-.012	-.002	-.001			.002	.001
	.013	-.006	.053*	.245***	-.006	(.014)	(.002)	(.001)	-.005	-.006	(.002)	(.001)
Severity group 3	(.04)	(.036)	(.03)	(.03)	(.013)	-.01	.001	-.004	(.034)	(.012)	0	-.002*
						(.013)	(.002)	(.002)			(.002)	(.001)
Severity group 4						-.014	-.01	-.016**			.009	.004
						(.018)	(.008)	(.007)			(.007)	(.004)
Severity group 5				.154**	.025	-.005	.008	-.045		0	0	0
	-.026	.036	.044	(.06)	(.025)	(.034)	(.008)	(.032)	.003	(.034)	(0)	(0)
Severity group 6	(.046)	(.048)	(.053)	.146***	-.031	-.003	-.004	0	(.043)	.079**	-.003	0
				(.056)	(.032)	(.033)	(.024)	(0)		(.034)	(.027)	(0)
Others	-.008	.087	.053	-.041	.057**	.011	.008	.002	-.003	.009	-.009	-.002
	(.05)	(.064)	(.064)	(.047)	(.028)	(.018)	(.008)	(.008)	(.06)	(.034)	(.007)	(.006)
Prematurity												
Severity group 1						.394***	.157***	-.019***			-.005	.013***
						(.125)	(.03)	(.005)			(.014)	(.003)
Severity group 2						.031	.056***	-.03**			-.01	.028**
	-.037	.031	-.004	-.037	-.003	(.045)	(.017)	(.013)	-.03	-.018	(.014)	(.013)
Severity group 3	(.032)	(.031)	(.028)	(.061)	(.032)	.066	.032	-.066***	(.028)	(.032)	.006	.057***
						(.044)	(.023)	(.019)			(.023)	(.02)
Severity group 4						-.044	.028	.031			.065*	.017
						(.054)	(.038)	(.036)			(.038)	(.041)
Severity group 5				-.092	-.047	-.083	-.053	-.306***		-.006	.078	-.144
	.038	.006	.017	(.074)	(.056)	(.087)	(.075)	(.082)	.025	(.061)	(.102)	(.105)
Severity group 6	(.037)	(.039)	(.049)	-.119*	-.019	-.138*	.156*	-.061	(.037)	-.081	-.038	-.02
				(.067)	(.051)	(.078)	(.081)	(.086)		(.052)	(.089)	(.097)
Others	-.015	-.022	-.006	.027	.031	-.023	.052*	-.021	-.008	-.107**	.025	.024
	(.036)	(.047)	(.059)	(.047)	(.048)	(.05)	(.028)	(.023)	(.044)	(.052)	(.027)	(.023)

continued on next page

Table 2.7: Differences above and below threshold by severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls (continued)												
Twin births												
Severity group 1						.026	-.005	.021***			.005	.014***
						(.125)	(.029)	(.005)			(.014)	(.003)
Severity group 2						-.022	.009	.146***			.008	.008
	.022	.007	-.033	-.018	-.035	(.042)	(.016)	(.009)	.025	-.038	(.013)	(.006)
Severity group 3	(.03)	(.027)	(.025)	(.051)	(.029)	.039	-.023	-.173***	(.025)	(.028)	.011	.035**
						(.038)	(.022)	(.017)			(.021)	(.017)
Severity group 4						.016	.016	.057**			-.026	.03
						(.051)	(.034)	(.024)			(.031)	(.023)
Severity group 5				-.012	.04	-.045	0	-.078		.172***	.035	-.192***
	.001	-.039	-.062	(.062)	(.049)	(.089)	(.073)	(.073)	.061*	(.059)	(.095)	(.055)
Severity group 6	(.037)	(.038)	(.045)	-.073	.04	-.099	.075	-.089	(.035)	.017	-.027	-.019
				(.061)	(.044)	(.088)	(.058)	(.063)		(.048)	(.063)	(.045)
Others	-.012	-.001	-.053	-.072*	-.021	-.022	.026	.009	-.058	.033	.011	.025**
	(.039)	(.052)	(.051)	(.038)	(.039)	(.041)	(.021)	(.014)	(.042)	(.043)	(.021)	(.012)
Multiple births												
Severity group 1						.047***	.004	.001**			.002	0
						(.004)	(.005)	(0)			(.001)	(0)
Severity group 2						-.002	-.006	0			.002	0
	-.009	.001	-.016	-.023	-.006	(.02)	(.005)	(.001)	-.021**	.007	(.003)	(0)
Severity group 3	(.012)	(.013)	(.01)	(.028)	(.015)	-.005	.006	.001	(.009)	(.015)	.006	0
						(.02)	(.005)	(.001)			(.005)	(.001)
Severity group 4						.021	-.003	.002			-.001	0
						(.019)	(.01)	(.002)			(.007)	(0)
Severity group 5				-.047	.012	-.091	.013	0		.024	-.032	.03
	-.013	.001	-.009	(.038)	(.026)	(.063)	(.025)	(0)	-.004	(.025)	(.022)	(.03)
Severity group 6	(.013)	(.016)	(.016)	-.063*	-.036	-.036	.009	.01	(.016)	.011	0	0
				(.037)	(.026)	(.045)	(.007)	(.01)		(.025)	(0)	(0)
Others	.005	.013	.012	-.007	.018	.006	.003	.002	-.004	-.034***	.003	0
	(.013)	(.009)	(.018)	(.017)	(.017)	(.015)	(.003)	(.002)	(.012)	(.011)	(.003)	(0)
Asphyxia												
Severity group 1						.066***	0	0**			0	0
						(.005)	(0)	(0)			(0)	(0)
Severity group 2						.041***	-.005	-.013**			.001	.004
	.032	-.005	-.013	.039	-.019	(.015)	(.006)	(.006)	.021	-.01	(.006)	(.007)
Severity group 3	(.029)	(.024)	(.019)	(.029)	(.017)	-.001	-.008	.013	(.021)	(.015)	-.008	.003
						(.022)	(.012)	(.01)			(.011)	(.01)
Severity group 4						-.025	-.01	.065***			-.025	.017
						(.033)	(.024)	(.023)			(.023)	(.025)
Severity group 5				.018	.015	.052	-.022	.034		-.063*	-.033	.03
	-.004	.016	-.016	(.04)	(.029)	(.036)	(.047)	(.036)	-.002	(.034)	(.043)	(.03)
Severity group 6	(.034)	(.032)	(.033)	.048	.034	.039	.023	.087*	(.03)	-.035	.004	-.018
				(.03)	(.033)	(.046)	(.052)	(.048)		(.034)	(.044)	(.075)
Others	.057*	.058	.019	-.061*	-.007	.025	-.036***	0	.076*	-.047*	-.001	-.031***
	(.033)	(.041)	(.045)	(.033)	(.032)	(.026)	(.014)	(.012)	(.042)	(.028)	(.013)	(.012)

continued on next page

Table 2.7: Differences above and below threshold by severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls (continued)												
Severe asphyxia												
Severity group 1						.012***	0	0			0	0
						(.002)	(0)	(0)			(0)	(0)
Severity group 2						.012***	.001	-.003			.001	.003
	.012	-.019	-.004	.016	-.002	(.002)	(.002)	(.002)	-.002	0	(.002)	(.002)
Severity group 3	(.018)	(.016)	(.012)	(.015)	(.006)	.012***	-.003	.001	(.012)	(.008)	.001	-.002
						(.002)	(.005)	(.005)			(.005)	(.005)
Severity group 4						-.014	-.012	.008			.002	.011
						(.018)	(.012)	(.014)			(.013)	(.014)
Severity group 5				.032***	-.011	.021***	-.03	.005		-.032**	.014	.03
	.015	.003	-.018	(.009)	(.019)	(.007)	(.028)	(.03)	.009	(.013)	(.034)	(.03)
Severity group 6	(.022)	(.02)	(.023)	.032***	.03**	.007	.024	.001	(.019)	-.008	.064*	.008
				(.009)	(.013)	(.033)	(.026)	(.024)		(.023)	(.036)	(.041)
Others	.054*	.024	.055	-.007	-.019	-.004	-.016*	-.01	.031	-.023	-.009	-.012
	(.03)	(.035)	(.036)	(.023)	(.024)	(.022)	(.009)	(.009)	(.035)	(.018)	(.008)	(.008)
Moderate asphyxia												
Severity group 1						.052***	0	0*			0	0
						(.004)	(0)	(0)			(0)	(0)
Severity group 2						.028*	-.004	-.01*			-.001	.002
	.017	.018	-.01	.021	-.012	(.015)	(.006)	(.005)	.024	-.012	(.005)	(.006)
Severity group 3	(.024)	(.019)	(.015)	(.025)	(.016)	-.007	-.006	.011	(.018)	(.013)	-.01	.002
						(.021)	(.011)	(.008)			(.01)	(.009)
Severity group 4						-.012	-.004	.05***			-.028	.006
						(.028)	(.021)	(.017)			(.019)	(.021)
Severity group 5				-.012	.026	.03	-.001	.028		-.036	-.048*	0
	-.02	.022	.002	(.039)	(.022)	(.035)	(.038)	(.02)	-.007	(.03)	(.027)	(0)
Severity group 6	(.028)	(.025)	(.026)	.019	.002	.062***	-.001	.096**	(.025)	-.035	-.048**	-.027
				(.029)	(.03)	(.011)	(.047)	(.044)		(.026)	(.023)	(.066)
Others	-.002	.034	-.036	-.042*	.017	.026*	-.02*	.006	.045*	-.031	.013	-.02**
	(.016)	(.023)	(.029)	(.024)	(.022)	(.014)	(.01)	(.008)	(.027)	(.021)	(.01)	(.009)
IRDS												
Severity group 1						.521***	.011	.003***			0	.001
						(.01)	(.008)	(.001)			(.003)	(.001)
Severity group 2						.295***	.021	.076***			-.015	.004
	.01	-.02	.02	.274***	-.009	(.039)	(.013)	(.01)	-.037*	.021	(.011)	(.01)
Severity group 3	(.024)	(.023)	(.02)	(.06)	(.032)	-.035	.056**	.169***	(.022)	(.032)	.021	.03
						(.044)	(.024)	(.018)			(.023)	(.019)
Severity group 4						-.076	.044	.11***			-.038	-.017
						(.057)	(.039)	(.036)			(.039)	(.041)
Severity group 5				.118*	-.046	.054	.158**	.071		.016	.198**	-.131
	.006	.056*	.018	(.064)	(.042)	(.08)	(.075)	(.094)	-.001	(.048)	(.088)	(.111)
Severity group 6	(.027)	(.032)	(.036)	-.105***	-.02	.015	.272***	.221***	(.025)	.024	.109	-.112
				(.032)	(.03)	(.062)	(.079)	(.084)		(.032)	(.087)	(.097)
Others	.037	.113*	.025	-.277***	.107**	-.063	.007	.028	-.003	-.016	.007	.023
	(.048)	(.063)	(.066)	(.049)	(.048)	(.046)	(.022)	(.019)	(.059)	(.053)	(.022)	(.019)

continued on next page

Table 2.7: Differences above and below threshold by severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Health-related controls (continued)												
Maternal smoking												
Severity group 1						.043***	.026**	.005***			-.004	.001**
						(.004)	(.012)	(.001)			(.005)	(.001)
Severity group 2						.011	-.01	.003			.006	.005
	.007	-.004	.006	.023	.013	(.016)	(.007)	(.004)	-.008	-.003	(.006)	(.004)
Severity group 3	(.012)	(.011)	(.011)	(.015)	(.012)	.028**	-.003	.002	(.011)	(.011)	-.006	.006
						(.011)	(.01)	(.008)			(.009)	(.007)
Severity group 4						.017	-.001	-.006			-.007	.001
						(.019)	(.018)	(.014)			(.016)	(.015)
Severity group 5				.022***	.01	.026***	-.019	0		-.016*	0	-.019
	-.013	.002	-.016	(.008)	(.013)	(.008)	(.019)	(0)	-.005	(.009)	(0)	(.019)
Severity group 6	(.012)	(.012)	(.021)	-.009	-.022	.014***	.014*	-.037	(.014)	.015	.021	.011
				(.023)	(.018)	(.005)	(.008)	(.036)		(.019)	(.021)	(.028)
Others	-.017**	.02*	.008	-.023	.005	.004	-.008	.004	0	-.008	-.001	.002
	(.008)	(.011)	(.008)	(.015)	(.003)	(.003)	(.006)	(.004)	(0)	(.011)	(.005)	(.004)
Maternal alcohol use												
Severity group 1						.002**	0	0			0	0
						(.001)	(0)	(0)			(0)	(0)
Severity group 2						.002**	-.001	0			0	0
	0	-.002	-.002	.003**	0	(.001)	(.001)	(.001)	-.002	.005	(.001)	(0)
Severity group 3	(0)	(.004)	(.002)	(.001)	(0)	.002**	0	.001	(.002)	(.004)	-.002	.003*
						(.001)	(.002)	(.002)			(.001)	(.002)
Severity group 4						-.011	.001	0			-.005	-.002
						(.013)	(.005)	(0)			(.003)	(.002)
Severity group 5				0	.003	.002	0	-.023		-.005	0	0
	0	.003	.009	(0)	(.003)	(.002)	(0)	(.023)	-.004	(.005)	(0)	(0)
Severity group 6	(0)	(.003)	(.006)	0	-.008	.002	0	0	(.004)	0	0	0
				(0)	(.008)	(.002)	(0)	(0)		(0)	(0)	(0)
Others	0	-.01	0	0	-.007	.002	0	-.002	0	0	-.001	.001
	(0)	(.01)	(0)	(0)	(.007)	(.002)	(0)	(.002)	(0)	(0)	(.001)	(.002)
Maternal drug use												
Severity group 1						.007***	0	0			0	0
						(.002)	(0)	(0)			(0)	(0)
Severity group 2						.007***	-.001	.003			.003	-.002
	-.003	-.004	.006	.003**	-.002	(.002)	(.002)	(.002)	.001	0	(.002)	(.002)
Severity group 3	(.003)	(.004)	(.004)	(.002)	(.003)	.007***	0	.008	(.003)	(.004)	.001	.006
						(.002)	(.004)	(.005)			(.005)	(.004)
Severity group 4						-.032	.018**	.004			.004	.021*
						(.022)	(.008)	(.011)			(.009)	(.012)
Severity group 5				-.014	0	.002	.016	0		0	0	0
	0	0	.005	(.019)	(0)	(.002)	(.011)	(0)	0	(0)	(0)	(0)
Severity group 6	(0)	(0)	(.005)	.005	-.012	.008**	-.022	.01	(0)	-.011*	0	-.027
				(.004)	(.012)	(.004)	(.022)	(.01)		(.007)	(0)	(.019)
Others	0	.007	0	0	-.004	0	0	.003	0	.004	-.001	0
	(0)	(.007)	(0)	(0)	(.007)	(0)	(.003)	(.003)	(0)	(.008)	(.005)	(.003)

continued on next page

Table 2.7: Differences above and below threshold by severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Hospital-related controls												
In-patient ward												
Severity group 1						-.167	-.024	-.012***			-.007	-.005
						(.108)	(.018)	(.004)			(.01)	(.003)
Severity group 2						0	-.005**	-.024***			.002	-.012**
	.003	0	0	0	0	(0)	(.002)	(.005)	0	0	(.003)	(.006)
Severity group 3	(.003)	(0)	(0)	(0)	(0)	0	.001	-.01***	(0)	(0)	-.002	-.013***
						(0)	(.001)	(.004)			(.002)	(.005)
Severity group 4						0	0	0			-.002	.004
						(0)	(0)	(0)			(.002)	(.004)
Severity group 5				0	0	0	0	0		0	0	0
	.005	0	0	(0)	(0)	(0)	(0)	(0)	0	(0)	(0)	(0)
Severity group 6	(.005)	(0)	(0)	0	0	0	0	0	(0)	0	0	0
				(0)	(0)	(0)	(0)	(0)		(0)	(0)	(0)
Others	-.004	-.033	-.044	.022	-.028*	-.016	-.023	-.037**	.007	.003	-.026	-.01
	(.014)	(.027)	(.028)	(.017)	(.017)	(.023)	(.016)	(.015)	(.017)	(.015)	(.017)	(.018)
In-patient midwife												
Severity group 1						-.079	-.031**	-.009**			.005	-.002
						(.08)	(.012)	(.004)			(.008)	(.003)
Severity group 2						.004***	-.01***	-.015***			0	-.003
	0	.001	-.001	.001	-.001	(.001)	(.003)	(.005)	0	-.002	(.004)	(.006)
Severity group 3	(.005)	(.001)	(.003)	(.001)	(.005)	.004***	.006**	-.012**	(0)	(.004)	-.003	-.014**
						(.001)	(.003)	(.005)			(.004)	(.005)
Severity group 4						.004***	-.001	-.007			.009	.012
						(.001)	(.007)	(.006)			(.007)	(.011)
Severity group 5				.003	0	0	0	0		0	0	0
	.005	.003	.005	(.003)	(0)	(0)	(0)	(0)	0	(0)	(0)	(0)
Severity group 6	(.005)	(.003)	(.005)	.003	.005	.004	0	.021	(0)	0	0	0
				(.003)	(.003)	(.003)	(0)	(.015)		(0)	(0)	(0)
Others	-.009	-.019	-.072	.014	-.029	.032	.022	-.029*	.021	-.012	0	.019
	(.018)	(.033)	(.045)	(.019)	(.025)	(.024)	(.018)	(.016)	(.029)	(.021)	(.019)	(.018)
In-patient doctor												
Severity group 1						-.249**	-.034*	-.015***			-.004	-.003
						(.125)	(.018)	(.004)			(.01)	(.003)
Severity group 2						-.007	-.004	-.024***			.001	-.008
	0	0	-.007	.001	.001	(.008)	(.003)	(.005)	.006	-.003	(.004)	(.006)
Severity group 3	(.005)	(0)	(.005)	(.001)	(.001)	.001	-.002	-.013***	(.004)	(.002)	0	-.014***
						(.001)	(.003)	(.004)			(.003)	(.005)
Severity group 4						.001	.006*	-.001			-.007*	.002
						(.001)	(.003)	(.004)			(.004)	(.005)
Severity group 5				.005	.003	0	0	.014		0	0	0
	.005	.005	-.007	(.004)	(.003)	(0)	(0)	(.014)	-.011	(0)	(0)	(0)
Severity group 6	(.005)	(.004)	(.014)	-.01	.01**	.006*	0	.01	(.008)	-.004	-.012	-.014
				(.016)	(.005)	(.004)	(0)	(.01)		(.004)	(.012)	(.014)
Others	.007	-.01	-.028	.029***	-.004	-.024	-.004	-.037**	.025*	.003	-.025	-.001
	(.009)	(.02)	(.022)	(.011)	(.011)	(.023)	(.015)	(.015)	(.014)	(.015)	(.015)	(.017)

continued on next page

Table 2.7: Differences above and below threshold by severity group, 2005-2011 (continued)

	reimbursement relevant thresholds								placebo thresholds			
	T = 600	T = 750	T = 875	T = 1000	T = 1250	T = 1500	T = 2000	T = 2500	T = 700	T = 1300	T = 2200	T = 2700
Number of births												
1 - N_L						2497	404	7988			1509	16474
N_R						12	778	13960			2984	29625
2 - N_L						2497	2384	3575			1929	2354
N_R	294	745	641	1483	813	124	1439	2790	361	365	3130	4061
3 - N_L	304	264	472	70	328	2497	1327	1402	528	636	756	1041
N_R						135	670	1312			1201	1718
4 - N_L						2497	536	463			257	227
N_R						77	225	317			433	406
5 - N_L				370	302	423	125	71		99	33	33
N_R	218	373	221	52	98	30	53	44	250	186	63	52
6 - N_L	225	142	147	370	419	484	216	97	275	132	47	41
N_R				65	118	31	45	52		262	84	74
Others - N_L	170	150	127	243	423	477	673	836	121	135	514	655
N_R	234	101	84	160	149	129	596	909	170	264	882	1158

Notes: Difference between means below and above birth weight threshold T within a ±25 gram bandwidth in each severity group ($\bar{X}_{below}$-$\bar{X}_{above}$). (Robust)Standard errors in parentheses below mean difference . N_L: Number of observations left/below threshold within 25 gram bandwidth. N_R: Number of observations right/above threshold within 25 gram bandwidth. *** $p < 0.01$; ** $p < 0.05$ and * $p < 0.1$. For variable definition see Table 2.1 **Source:** Own calculations based on the DRG-Statistic.

Chapter 3

Birth in Times of War - An investigation of health, mortality and social class using historical clinical records

3.1 Introduction

Early childhood and the time in utero may be one of the most critical time periods in life (Almond and Currie 2011). To establish a causal effect of adverse early-life environment on later life outcomes, a growing literature exploits historical shocks like natural disasters, recessions, famines and wars. The by far greatest shock that has affected living cohorts in Western Europe is World War II (WWII). Individuals exposed to WWII in utero or early-life have been shown to have higher morbidity and mortality rates, worse socio-economic outcomes and even a modified behaviour at older ages (see e.g. Kesternich et al. 2014; Kesternich et al. 2015; Van den Berg et al. 2016; Jürges 2013; Atella et al. 2016). These findings are based on samples of the surviving population. If individuals who survive infancy during the war do systematically differ from survivors of other cohorts, estimates of long-term effects may be biased (see e.g. Lindeboom and Van Ewijk 2015; Van Ewijk and Lindeboom 2016). As historical individual level data on birth outcomes are hardly available,[1] it is unclear whether the negative effects of WWII remained latent until later life or were already present at time of birth.

[1] A rare exception is the "Dutch Famine Birth Cohort Study". See Lumey et al. (2011) for an overview.

The aim of this research project is to estimate the short-term effect of the onset of WWII on perinatal health and mortality of infants. To explore the relation between changes of perinatal infant mortality and individual characteristics related to outcomes later in life, we estimate heterogeneous treatment effects by social group and infant health. We collected an entirely new data set of historical clinical birth records from the largest birth hospital in Munich, Germany. Our unique data contain around 10,000 births and miscarriages which took place in hospital between December 1937 and September 1941. Besides a rich set of demographic variables, our data set contains detailed socio-economic information, namely legal and occupational status which we classified into a validated measure of social class. In our empirical strategy we exploit the unexpected onset of WWII as natural experiment.

Even 60 years after the end of WWII its consequences continue to shape individual live outcomes. Kesternich et al. (2014) use retrospective life data to document that individuals exposed to WWII are more likely to suffer from diabetes and depression at old ages. Atella et al. (2016) investigate the impact of WWII on health in an Italian context. They show a link between early life stress caused by exposure to intense conflicts and depression, while exposure to famine increases the probability of diabetes in later life. A number of research projects exploit WWII to study the long-term effects of hunger in early life. For example Van den Berg et al. (2016) provide causal evidence that hunger leads to a decrease in adult height and Kesternich et al. (2015) show that individual behaviour can serve as a pathway between early life shocks and later life health. Similarly, the small literature drawing on historical birth records to study the short-term impact of World War II on health at birth mainly focuses on the role of nutritional shortage during gestation. Stein et al. (2004) find individuals affected by the Dutch Hunger Winter 1944/1945 during the third trimester to have decreased birth weight and birth size, while no effect is found for individuals exposed during early pregnancy. Using data similar to ours Floris et al. (2016) study how birth weight evolves over the course of World War I in one Swiss hospital. In their setting food rationing during the end of the war leads to a decrease in birth weight for children from medium SES families, whereas high SES families can compensate the price shocks and low SES families benefit from public interventions. Our work is also related to a strand of literature investigating the impact of shocks in utero and maternal stress using modern data. This literature exploits a variety of shocks, for example natural disasters (Currie and Rossin-Slater 2013; Torche 2011), terrorist attacks (Quintana-Domeque and Ródenas-Serrano forthcoming) or mass layoffs (Carlson 2015). Most of this studies can document a small decrease birth weight due to an external shock. An

exception is Currie and Rossin-Slater (2013) who do not find a change in birth weight, but show that stress in utero affects more extreme health outcomes.
While we do not find any sizeable effects of the onset of the war on health measured as birth weight, we can document a strong, robust increase perinatal infant mortality. This mortality effect can mainly be attributed to live born children who die before leaving the hospital. Perinatal mortality increases for all social classes and disproportionally for very low birth weight infants. Previous literature relating WWII to health outcomes often focuses on extreme effects of the war like bombings, hunger, combat and dispossession. Similarly to Lindeboom and Van Ewijk (2015) and Van Ewijk and Lindeboom (2016), we study less extreme war-related events. We focus on the first two years of WWII, a period when military operations took place outside of Germany and there was no nutritional shortage.[2] Our main contribution is to document an effect of WWII on perinatal child mortality even in the absence of extreme conditions. The onset of WWII acted as shock to individuals and the public health system, which initially led to a jump in perinatal mortality and then gradually faded out. This interpretation is consistent with historical evidence, showing that the onset of the war caused turmoil in the health system and disrupted daily life. Two mechanisms are potentially driving our results. Maternal stress levels are likely to be increased due to drafting of husbands and a high level of uncertainty that comes along with the war, and second drafting of doctors can lead to a decrease in medical quality. As we find the mortality effects to be stronger, where medical quality should matter, namely for live born children, we conclude the decline in medical quality to be the more important mechanism. Our results have important implications for the literature on long-term effects of WWII. We document a disproportional increase in mortality for very low birth weight infants, suggesting that studies using samples of the surviving population provide a lower bound for the true effect.
The remainder of the chapter proceeds as follows: Section 3.2 provides more detailed information on the historical background. Section 3.3 describes our data, the way we constructed our variables and presents first descriptive analyses. We explain our empirical strategy in Section 3.4 and present our results in Section 3.5. Section 3.6 concludes.

[2] Schiman et al. (2017) use a similar time frame and variation in infant mortality in the early stages of WWII to estimate the effects of adverse early-life circumstances on adult outcomes in England. In contrast to our setting living conditions in England had already deteriorated significantly in many ways at that time due to large-scale bombing and severe rationing.

3.2 Historical and institutional setting

3.2.1 General historical background

Events leading to World War II

When Hitler and the Nazi Party seized power in 1933, the transformation from a weak democracy to an autocratic dictatorship began immediately. Within months public institutions, local and regional authorities, judicature and even private clubs were brought under the control of the Nazi party. Non Aryan Germans were dismissed from jobs in the civil service and whoever publicly raised criticism became subject to brutal repression (Evans 2004, p. 498-509). In violation of the treaty of Versailles the Nazis also launched the rearmament of the German military. In 1935 a military law made all male Germans between 18 and 45 liable to military duty. Nevertheless, neither the German public nor other European powers were aware of the imminent threat of a war. When Hitler began with the restoration and expansion of Germany he did so using massive political pressure on foreign governments instead of military force. Between 1935 and 1938 three former German territories separated after WWI were reintegrated into Germany (Territory of the Saar Basin by referendum, Rhineland and Memel Territory by occupation, Austria by voluntary annexation). The first actual expansion took place in 1938 when the German military occupied the Sudeten German territories in Czechoslovakia. The essential military powers in Europe - Great Britain, France, Italy - tolerated this aggression to appease Hitler and to avoid a new war in Europe. Even when Hitler violated previous agreements again in 1939 by occupying the rest of Czechoslovakia, they did not intervene in any military way.

After these successes Hitler and the Nazi state were celebrated by the majority of the German population. Hitler had reached his goals without bloodshed and the population perceived Germany to be a world power again. The general public hoped that wars could be avoided in the future as well - either because Hitler had already archived his goals or because his political measures were sufficient to do so (see Frei 2013, pp.150).

World War II

WWII began with the invasion of Poland in September 1939. For the first time the German military experienced resistance by Polish forces, and its guarantor powers - France and Great Britain - declared war on Germany.[3] This had been unexpected by the German public, to whom it was clear quickly that this conflict would be different

[3] The engagement of the German air force during the Spanish civil war in 1936 was held secret until the end of 1939.

from any other conflict since 1918. There was a great feeling of uncertainty and no euphoria among the population, since most people had experienced the negative consequences of the last world war. Prior to 1944, military operations of WWII mainly took place outside of Germany. Therefore the German population was initially not subject to direct effects of the war like hunger and bombings.

Nevertheless, the onset of WWII marked a distinct break in the daily routine. First, drafting affected a great number of men who were subsequently absent from their families and workplaces. In the end of 1939 around 4.2 million men out of a male population of 33.8 million[4] were serving the military, another 3.5 million men were drafted in 1940 (Overmanns 2009, p.217).[5] Men were drafted based on their year of birth and service in the previous war without social class dependent privileges or exceptions (Absolon 1960, p.4, 152-153).[6] Second, to prioritize production for military purposes, the economy was transformed into a wartime economy. Three days before Germany invaded Poland, the regime announced the introduction of ration stamps for food and other commodities like fabric, leather and soap. The local population in Munich responded to the introduction of ration stamps with a rush to the shops and officials were not well prepared to manage the new circumstances (Stadtarchiv München 1939-1940).[7] While there is no evidence suggesting that the population was affected by serious hardship during the first two years of the war, daily life became more complicated. Long queues in front of shops were common especially in the first weeks of the war and commodities like furniture and bedding eventually became objects of speculation. There was no general shortage of food. However, food quality declined and availability of certain categories of food varied. Pregnant women received preferential treatment. Unlike the general population they were allocated whole milk and when coal was in short supply in February 1940, pregnant women were eligible for extra rations. Records of the hospital do not indicate any problems with the catering of patients or shortage of fuel.

Nevertheless the German health system entered the war ill-prepared. No concept existed on how to operate medical services for the civil population. Instead the military was given full priority. The army made frequent use of its authority to dispose all resources of the civil health system. Besides confiscations of local hospitals, large scale

---

[4] German Reich as of 1937.

[5] Poland was already defeated (with minor German military losses) in October 1939 and lots of soldiers returned on furlough. However, the atmosphere in Germany remained tense as there was a constant threat that soldiers, who had just returned, would be sent to war again.

[6] Only certain conscripts were (temporarily) exempt if their specific occupation duty was classified - again on a case-by-case basis - as indispensable for "homeland defence".

[7] Confidential quarterly reports by the Economic Department give a detailed account of the Economic situation in Munich.

drafting of physicians lead to conflicts between the military and the civil sector. Already in fall of 1939 one third of all available physicians were in military service. To mitigate the shortage of physicians, the state granted final year medical students their approbations prematurely. Turmoil in the health system was greatest during the first weeks of the war, while the situation remained tense throughout (Süß 2003, p.181-212).

Fertility and childbirth under Nazis rule

Childbirth was no longer considered a private matter in Nazi Germany. Between 1900 and 1933 the number of yearly births in Germany had fallen by more than 50% (Sensch 2006), an unacceptable state for a regime adhering to a pro-natalist ideology. However, as the Nazis' world view was based on Eugenics, their goal was not to increase everybody's fertility. The regime used brutal repression to prevent reproduction among those considered to deteriorate the gene pool (Fallwell 2013). To boost birthrates among healthy "Aryan" Germans, the Nazis combined family propaganda, a ban of voluntary abortion[8] and material incentives.[9] Indeed the absolute number of births was increasing in the years prior to World War II. Even the location of delivery became infused with political agenda. The Nazi regime was heavily opposed to the increasing trend towards hospital birth. While the concept of women giving birth at home within their family members fitted in perfectly with the Nazi ideology, home births also spared the resources of the health system. Efforts to propagate home births climaxed in the so called "midwife edict" of September 1939 (RMI 1939). This edict requested hospitals to reject pregnant women without medical or social indication for hospital births. The hospital our data come from was a teaching hospital and therefore exempt from this rule. Due to decisive resistance of the association of gynaecologists the "midwife edict" was modified in 1940, granting women a choice over the location of delivery (Zander and Goetz 1986). Official statistics indicate that the proportion of hospital births in Germany was growing during the Nazi era despite all attempts. In 1935 25% of live births took place within a hospital compared to 38% in 1940 (Statistisches Reichsamt 1933-1940).[10]

[8] In the late 1920's Germany was given of the most liberal abortion policy in the developed world (Usborne 2011).

[9] For example, eligible newly wed couples received marriage loans, whose repayment was reduced with each child born.

[10] Before 1935 official statistics only counted the number of births within *birth* hospitals. In urban areas the proportion of hospital births was even higher.

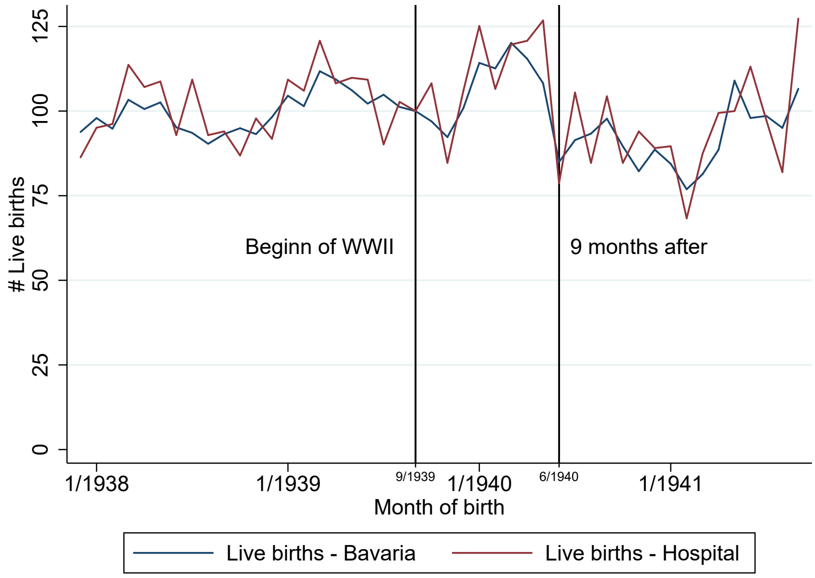
3.2.2 The hospital

The hospital Frauenklinik Maistrasse is the oldest and one of the largest gynaecological hospitals in Munich. It was founded as a state-run university hospital in 1884, succeeding the municipal birth house. In its first years the hospital mainly served lower-class and often single mothers. Women of higher social status traditionally gave birth at home. However, after moving into its current venue in 1916, the Frauenklinik Maistrasse became one of the leading gynaecological hospitals in Germany and attracted patients among all social classes. The hospital was divided into a general and a private ward. Most patients were admitted to the general ward and their treatment was completely covered by public health insurance. The private ward enabled the hospital to extract rents from more affluent or privately insured patients. These patients received special attention by the senior staff.

Deliveries were supervised by both doctors and midwives, but only doctors carried out surgeries and medical procedures. With the onset WWII the drafting of physicians heavily affected the daily routine. The director of the hospital frequently complained in letters to the state administration and applied for exemptions from military service for many of his doctors. For example, in a letter from December 1939 he states that already seven of his doctors were serving the military and four more had received draft calls. Much of the workload was shifted to graduates and unpaid trainees. In the Nazi era the hospital carried out large numbers of forced abortions and sterilisations on women who allegedly suffered from hereditary diseases.[11] Two groups of births are likely to be oversampled in our data: births of mothers with very low socio-economic status and pathological births. Home birth was no option for women living under crowded or unsanitary conditions. Often these women would seek admittance to the hospital weeks before delivery, where they acted as teaching material for medical students and midwives in training. Women in risk of a pathological birth were referred to hospital by midwives and gynaecologists. Still, as hospital births had become quite common especially in big cities by 1937, our sample is broad enough to draw conclusions also for other groups. Around half of our observations equal at least a status of a skilled worker and almost 60% of women entered the hospital without any preexisting risk factors. Between 1938 and 1940 around 17% of all Munich live births took place in the Frauenklinik Maistrasse (see Table 3.20 in Appendix section 3.7.3). Figure 3.1

[11] Most such records state the women suffered from "hereditary feeble-mindedness". Since the 1990's the hospital has endeavoured to shed light on its role during the Nazi era (Stauber 2012).

Figure 3.1: Number of live births in Bavaria and hospital



Notes: Number of live births in Bavaria and our hospital by month of birth, 9/1939=100.
Source: Bayerisches Statistisches Landesamt 1937-1942

shows the monthly trend in the number of live births for our hospital and the whole state of Bavaria, normalized for September 1939. Both trends match quite well and no structural breaks (e.g. at the begin of the war or due to the "midwife edict") point to any differential selection into our hospital.

3.3 Data

3.3.1 Sample selection and variables

Sample selection

We digitized the universe of entries in the hospital's birth records from December 1937 to September 1941.[12] A twin birth results in two observations. The 10,325 observations consist of live and stillbirths, miscarriages and a small number of other conditions. Other conditions comprise women who came to the hospital post birth, women receiving treatment during pregnancy, medically induced interruptions and forced abortions and sterilisations. We exclude these 196 observations from the sample.

Around 1,200 observations are marked as miscarriage in the birth records. These mostly lack information on the child such as weight, length and sex, whereas the gestational

[12] See Appendix section 3.7.2 for a description of the hospital records.

age has been recorded in more than 90% of cases. Miscarriages mostly took place outside the hospital and women sought treatment afterwards. Patterns of selection into the hospital are very likely to vary between women who intend to give birth and women who are treated after a miscarriage. Therefore we exclude miscarriages from our main analysis.[13]

Outcome and control variables

Our primary outcomes are perinatal infant mortality, measured as whether an infant left the hospital alive, and birth weight. Birth weight is an overall measure of health at birth (McIntire et al. 1999), while also being a predictor of future life outcomes, for example educational attainment and adult height (Behrman and Rosenzweig 2004). Currie and Rossin-Slater (2013) find that birth weight is not affected by exposure to stress in utero, while more extreme measures of newborn health are. Therefore we also analyse asphyxia and maturity. Asphyxia is caused by deprivation from oxygen during the process of birth. It often results in the death of the infant and can cause long-term damage to surviving infants. Maturity is an indicator whether the birth takes place at full term. It is assessed by the appearance of the infant.[14]

Our control variables include characteristics of the mother like age, the number of pregnancies and a measure of social status which is derived from the occupational information in the birth records. We categorize this occupational information according to HISCLASS, a validated measure of historical social classes. Each occupation is assigned one out of 12 social classes defined as "a set of individuals with the same life chances" (Van Leeuwen and Maas 2011, p.18). In our empirical analysis we rely on the previous literature and use a compressed 7-class version of HISCLASS (see Abramitzky et al. (2011) and Schumacher and Lorenzetti (2005)).[15][16] For each observation, the birth record contains either the occupation of the father or the occupation of the mother. If the occupation of the mother is given, the entry uses the female version of the occupation. Otherwise the male version is used, mostly with a suffix like -wife, -daughter or -widow. We classify women accordingly as "working", "wife" or

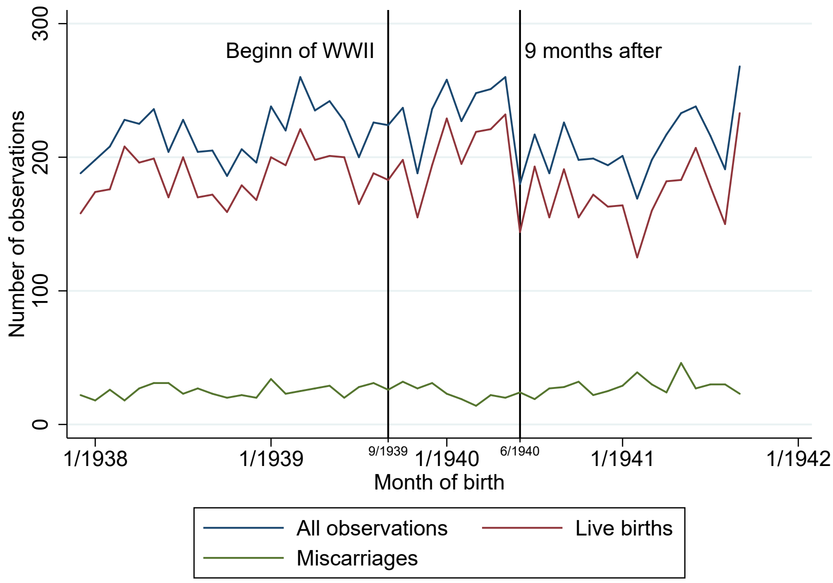
[13] Stillbirths on the other hand almost always include characteristics of the child but do not generally contain a gestational age. Partly the definitions of stillbirth and miscarriage seem to overlap since weight and gestational age of "miscarriages" exceeds 1000 grams and the fifth month in individual cases, while stillbirths" encompass a few infants with a birth weight below 1000 grams. Nevertheless we maintain the categorisation from the administrative records.

[14] To assess maturity, midwives checked colour of skin, body hair, ear conch and the appearance of genitals.

[15] This simplifies the interpretation of regression coefficients, attenuates possible coding errors and increases sample size within classes.

[16] A detailed description of the occupational coding can be found in Appendix section 3.7.1.

Figure 3.2: Timeline of observations in hospital



Notes: Number of all observations, live births and miscarriages by month of birth.

"single". Note that this approach assumes that the categories are mutually exclusive, while in reality a married women may also work. Further control variables include the sex of the infant, multiple births and the fetal position. While fetal malpresentation is one of the most frequent reasons for complications at birth, the fetal position appears to be random in births on term.

Descriptive statistics

Figure 3.2 displays the number of observations by category over our period of observation. The graph shows a distinctive drop in the number of births in June 1940 - nine months after the begin of the war, when many men were drafted for the invasion of Poland. Similarly another drop occurred in February 1941, nine months after the begin of the invasion of France. In mid 1940 many of the German soldiers were granted furlough, leading to an increased number of births towards the end of the observation period. Table 3.1 shows that 96% of the births in our sample are live births. In 93.5% of all births the infant left the hospital alive,[17] implying that in addition to the 4% stillborn children, 2.5% of infants died after birth in hospital. Most births (93%) took place in the general ward. The mothers in our sample are on average 28 years old and experience their second pregnancy, 30% of the women in our sample report an own

[17] The median newborn stayed in hospital for 9 days after birth.

Table 3.1: Descriptive statistics - Births

General characteristics	N	Mean	SD	Min	Max
Birth after 9/1939	8828	0.543	0.498	0	1
General ward	8828	0.931	0.253	0	1
Length of stay	8769	12.704	11.934	0	379
Live birth	8828	0.960	0.195	0	1
Infant leaves hospital alive	8828	0.936	0.246	0	1
Regular fetal position	8688	0.919	0.273	0	1
Mother	N	Mean	SD	Min	Max
Age of mother	8828	27.921	6.211	14	50
Parity	8826	2.208	1.804	1	19
Status is wife	8828	0.651	0.477	0	1
Status is own job	8828	0.310	0.462	0	1
Status is single, divorced or widowed	8828	0.031	0.173	0	1
Social status	N	Mean	SD	Min	Max
Higher managers & professionals	8500	0.069	0.253	0	1
Lower managers & professionals, cleric	8500	0.194	0.396	0	1
Foremen & skilled workers	8500	0.225	0.418	0	1
Farmers	8500	0.072	0.259	0	1
Lower skilled workers	8500	0.133	0.340	0	1
Unskilled workers	8500	0.281	0.450	0	1
Farm workers	8500	0.025	0.157	0	1
Infant	N	Mean	SD	Min	Max
Male	8822	0.527	0.499	0	1
Birth weight	8820	3218.620	601.065	280	5510
Length of infant	8815	49.998	3.108	19	61
No. of infants	8828	1.027	0.164	1	3
Asphyxia	6784	0.023	0.148	0	1

Notes: Descriptive statistics of births in sample (excluding miscarriages).

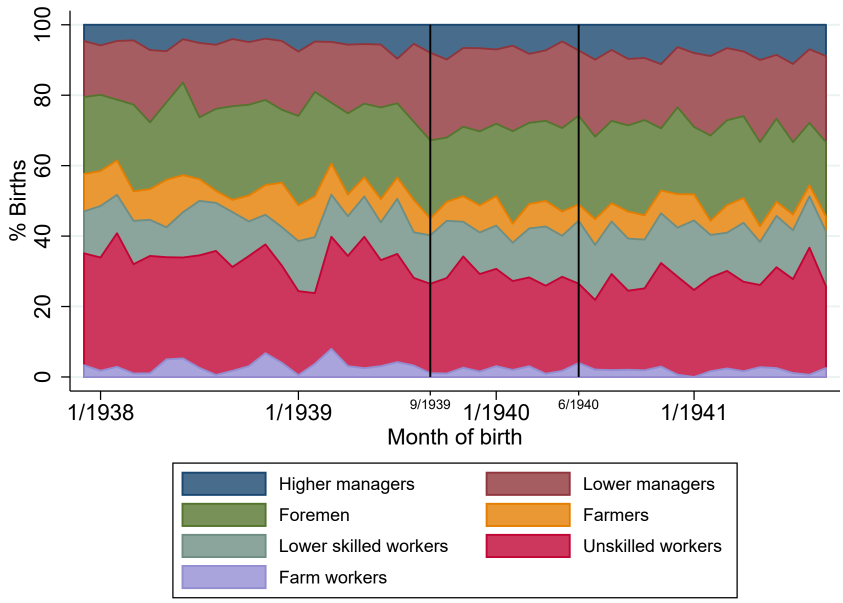
occupation. Unreported analyses show that lower classes are overrepresented among these working women. We compare the unconditional means by prewar and war period in Table 3.2. There is no difference in terms of age and parity of mother, as well as maturity and weight of the infant. The proportion of regular fetal positions does also not change significantly. Since the proportion of regular fetal positions is unlikely to be affected by the war, this suggests that women at risk of a complicated birth were not sent to the hospital more frequently during the war. On the other hand, the perinatal mortality rate is significantly higher during the war and the composition of mothers in terms of social status, labour force participation and marital status does show some changes. This highlights the importance of controlling for socio-economic characteristics. However, no abrupt breaks occur with the begin of the war (see Figures 3.3 and 3.4).

Table 3.2: Mean comparison - Births

General characteristics	Mean before war	Mean after war	Diff	SD	p	N before war	N after war
General ward	0.952	0.914	-0.0374***	0.005	0.000	4035	4793
Length of stay	12.560	12.824	0.2639	0.256	0.303	3979	4790
Live birth	0.966	0.956	-0.0098*	0.004	0.019	4035	4793
Infant leaves hospital alive	0.949	0.924	-0.0247***	0.005	0.000	4035	4793
Regular fetal position	0.920	0.918	-0.0020	0.006	0.728	3954	4734
Mother	Mean before war	Mean after war	Diff	SD	p	N before war	N after war
Age of mother	27.845	27.985	0.1406	0.133	0.289	4035	4793
Parity	2.188	2.224	0.0356	0.039	0.356	4035	4791
Status is wife	0.614	0.682	0.0675***	0.010	0.000	4035	4793
Status is own job	0.339	0.285	-0.0533***	0.010	0.000	4035	4793
Status is single, divorced or widowed	0.037	0.026	-0.0108**	0.004	0.003	4035	4793
Social status	Mean before war	Mean after war	Diff	SD	p	N before war	N after war
Higher managers & professionals	0.055	0.080	0.0253***	0.006	0.000	3878	4622
Lower managers & professionals, cleric	0.174	0.212	0.0375***	0.009	0.000	3878	4622
Foremen & skilled workers	0.226	0.224	-0.0020	0.009	0.826	3878	4622
Farmers	0.084	0.062	-0.0222***	0.006	0.000	3878	4622
Lower skilled workers	0.123	0.141	0.0178*	0.007	0.016	3878	4622
Unskilled workers	0.305	0.261	-0.0434***	0.010	0.000	3878	4622
Farm workers	0.032	0.019	-0.0130***	0.003	0.000	3878	4622
Infant	Mean before war	Mean after war	Diff	SD	p	N before war	N after war
Male	0.525	0.529	0.0042	0.011	0.696	4033	4789
Birth weight	3227.907	3210.802	-17.1054	12.847	0.183	4031	4789
Length of infant	50.198	49.830	-0.3674***	0.066	0.000	4030	4785
No. of infants	1.028	1.026	-0.0019	0.004	0.583	4035	4793
Asphyxia	0.021	0.023	0.0028	0.004	0.483	1991	4793

Notes: T-tests on the equality of means by war (excluding miscarriages). Significance levels: ***$p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

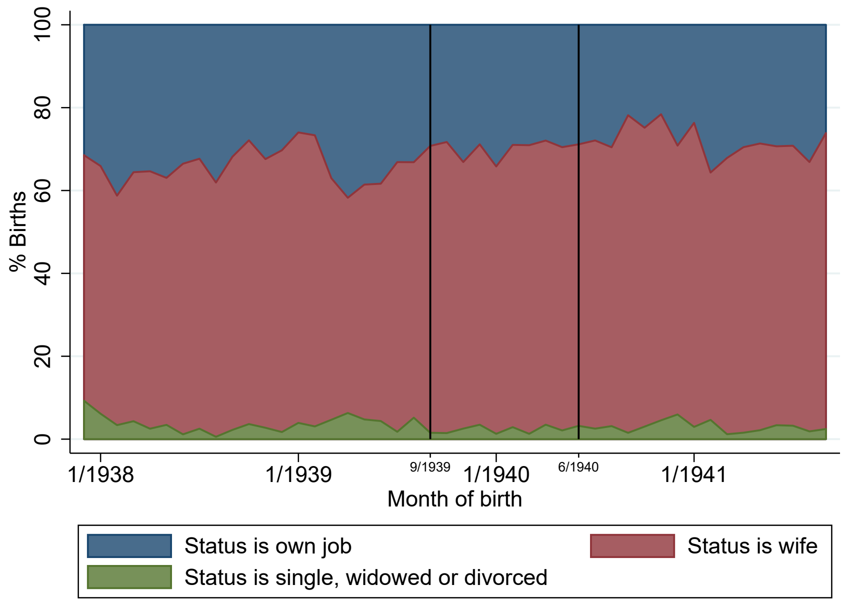
Figure 3.3: Composition in terms of social classes over time



Notes: Proportion of mothers by social class by month of birth.

We also test whether the war had an impact on length of stay in hospital measured in days after birth. One might be concerned that the probability of observing a mortality event increases, when mother and infants remain in the hospital for a longer period. However no significant change occurred. Descriptive statistics and mean comparisons for miscarriages can be found in Table 3.21 and 3.22 in Appendix section 3.7.3. Women who suffer a miscarriage are on average older and have more previous pregnancies than women who give birth.

Figure 3.4: Composition in terms of marital and working status over time



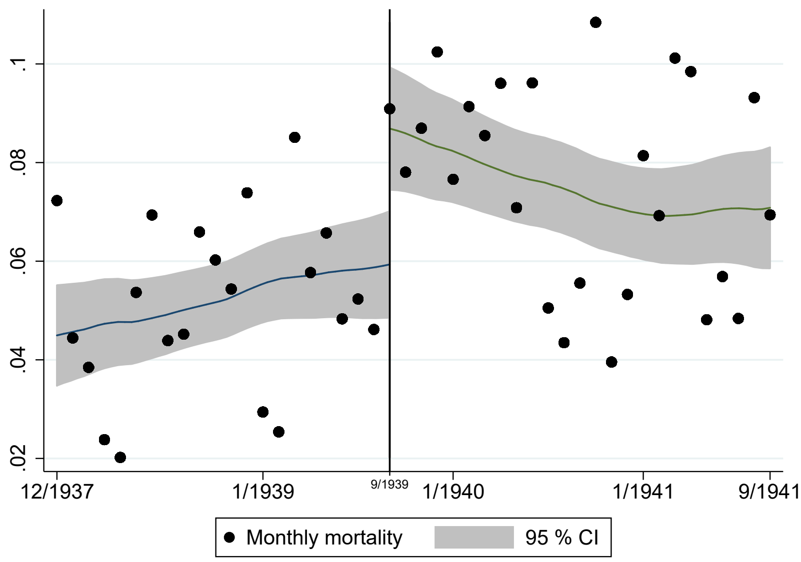
Notes: Proportion of mothers by marital and working status by month of birth.

3.3.2 Descriptive analysis

We begin our analysis by documenting the effect of WWII on perinatal mortality and health graphically. The monthly trend of perinatal infant mortality is presented in Figure 3.5. The dots denote the raw monthly mortality rate. We fit local linear regressions separately for the prewar and the war period. The graph documents a significant jump in perinatal mortality in September 1939. During the following months average perinatal mortality decreases gradually, but remains above prewar levels. In a next step we adjust for observable characteristics.

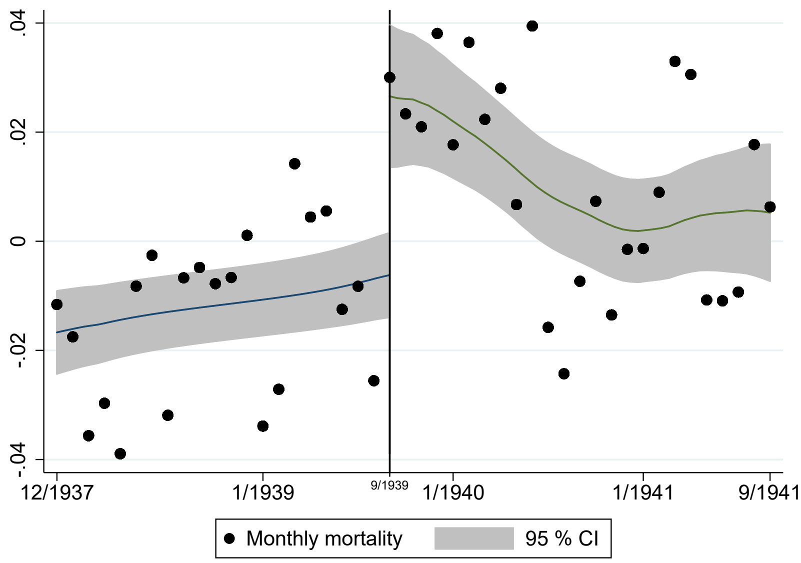
Figure 3.6 displays the monthly averages of residuals obtained from regressions of perinatal mortality on all maternal characteristics given in Table 3.1, infant gender and a dummy for regular fetal position. The jump at the threshold provided by the onset of the war remains significant. The decline in the mortality rate during the war period is slightly more pronounced compared to the plot without adjustment and the mortality rate in 1941 is no longer significantly greater than in the months preceding the war.

Figure 3.5: Raw perinatal mortality by month of birth - All births



Notes: Perinatal death rates (monthly averaged) and local linear regressions with a ROT bandwidth and an Epanechnikov kernel separately for the prewar and the war period.

Figure 3.6: Adjusted perinatal mortality by month of birth - All births



Notes: Regression residuals (monthly averaged) from regressions of perinatal mortality on social status, mother's age, parity, primipara, twinning status, infant's gender, marital status, a dummy for general ward, normal fetal position and working status.

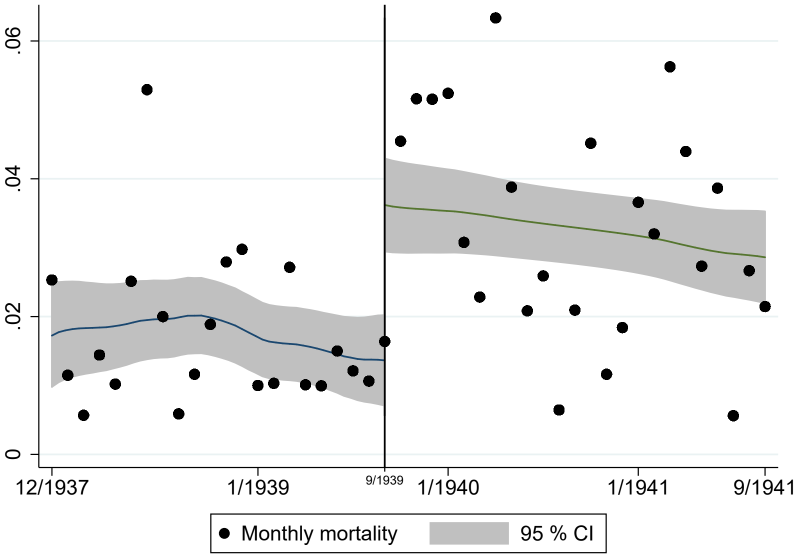
To explore whether the overall increase in perinatal mortality rate is driven by stillborn infants or by live born infants who die in hospital after birth, we repeat the analysis for live births in Figure 3.7 and Figure 3.8. Again we see a significant jump in September 1939 followed by a linear decline in mortality. This suggest that a large part of the overall mortality effect is driven by live born children.

If conditions become worse permanently because of the war, one would expect the effect to stay constant or even accumulate. Our descriptive results point to another interpretation. The onset of WWII might have provided a one time shock, which initially led to a jump in perinatal mortality and then gradually faded out. This explanation is consistent with the evidence presented in Section 3.2.1. The onset of the war was unexpected by the general public and affected the daily routine of individuals. Furthermore a shift of resources towards the military caused turmoil in the unprepared health sector. Yet, prior to 1942 conditions were not as averse as that it would have been impossible for individuals and organisations to adapt.

Given the duration of pregnancy, it is unlikely that the composition of mothers changed abruptly around our threshold. Still, we cannot rule out that the onset of the war had an impact on fertility decisions. To investigate, how the risk profile of mothers differs before and during the war, we regress perinatal infant mortality on our control variables using only observations from the prewar period. We then use the estimated coefficients to predict perinatal infant mortality for the whole sample. Figure 3.9 shows predicted mortality for infants born before the war and the predicted mortality under the assumption that war had no impact on perinatal infant mortality for infants born during the war. No significant change is found around the threshold. Whereas observed mortality decreases during the war, we see an increase in predicted risk after the first quarter of 1940.

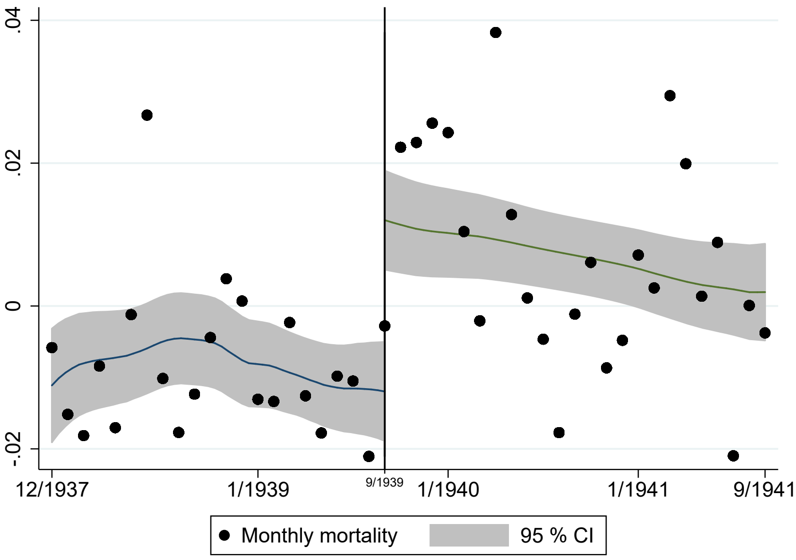
Next we turn to measures of perinatal health. Features of the distribution of birth weight are presented in Figure 3.10. Average birth weight stays almost constant during our whole period of observation. Rather than on the average birth, war might have an impact on more extreme cases. We add lines of the 25th and 75th percentiles of monthly birth weight to our plot to investigate trends for children with highest or lowest birth weight. However, also these cases do not show any trend. Similarly, kernel estimates of the density of birth weight do not indicate that any part of the distribution of birth weight was affected by the war (see Figure 3.11). Graphical results for asphyxia and maturity are given in Figures 3.14 and 3.15 in Appendix section 3.7.3.

Figure 3.7: Raw perinatal mortality by month of birth - Live births



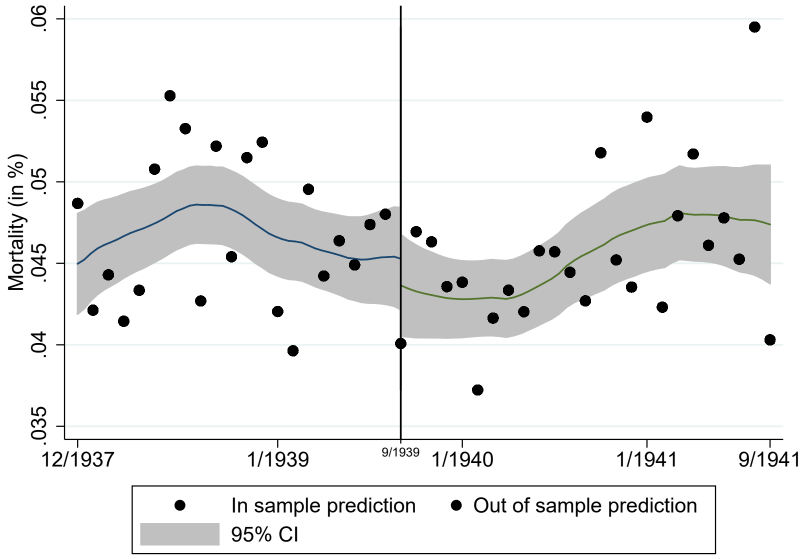
Notes: Perinatal death rates (monthly averaged) and local polynomial regressions with a ROT bandwidth and an Epanechnikov kernel separately for the prewar and the war period for live births.

Figure 3.8: Adjusted perinatal mortality by month of birth - Live births



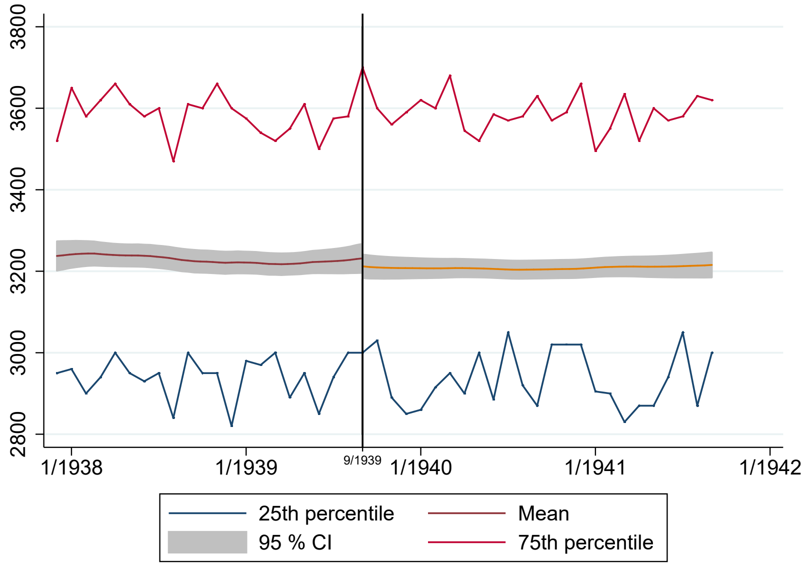
Notes: Regression residuals (monthly averaged) from regressions of perinatal mortality on social status, mother's age, parity, primipara, twinning status, infant's gender, marital status, a dummy for general ward, normal fetal position and working status for live births.

Figure 3.9: Predicted mortality



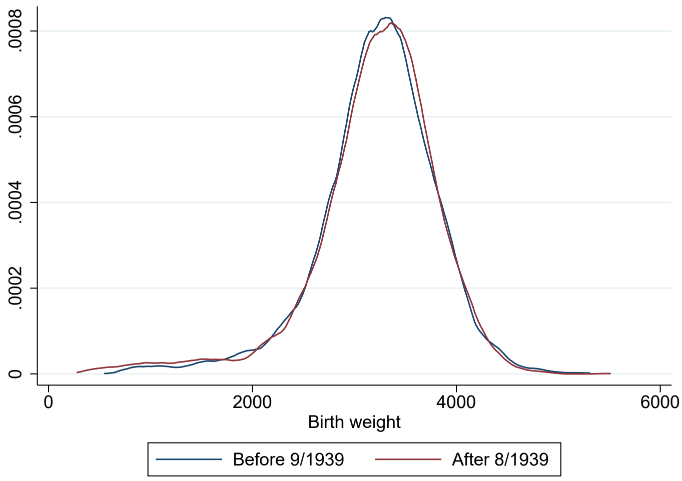
Notes: Predictions (monthly averaged) from regressions of perinatal mortality on social status, mother's age, parity, primipara, twinning status, infant's gender, marital status, a dummy for general ward, normal fetal position and working status.

Figure 3.10: Birth weight by month of birth



Notes: 25th percentile, mean and 25th percentile of birth weight.

Figure 3.11: Birth weight distribution



Notes: Kernel density estimate of birth weight by war status.

3.4 Regression analysis

The aim of this work is to estimate the effect of the onset of WWII on perinatal health and mortality of infants. In our identification strategy we exploit the onset of WWII as a natural experiment. There is no evidence that anticipation of a coming war affected fertility patterns before September 1939 (see Section 3.2.1). Hence we argue that the onset of the war constitutes an unexpected shock for women already pregnant in September 1939. After September 1939 fertility decisions may be affected by the war. Therefore we conduct our analysis using both our whole observation period (1/1938- 9/1941) and a restricted observation period (12/1937- 5/1940). All full term births included in the restricted sample were conceived before the outbreak of the war. However, given that our data do not contain a reliable measure of gestational age, we cannot exclude preterm births conceived during the war period from the restricted sample. Preterm births are associated with a higher risk of perinatal mortality. While preterm birth itself can be a consequence of war, and therefore part of the effect we want to capture with our war dummy, our results will overestimate the true effect on mortality if women with an ex ante high risk of a preterm birth increase their fertility relative to other women during the war. Although we cannot generally rule out such concerns, we argue that an increased share of premature births should be reflected in an on average lower birth weight. Our descriptive analysis of trends in birth weight in Section 3.3.2 do not indicate any change. Additionally we run all our regressions also on a sample restricted to live births, assuming that the share of preterm births is lower among live births.[18] Our baseline results are obtained estimating the following equation:

$$y_i = \alpha + \beta \mathtt{war}_i + \kappa C_i + u_i \tag{3.1}$$

y_i is the outcome (infant mortality, birth weight, maturity, asphyxia), `war` is an indicator if birth took place after the begin of WWII, i.e. in or after 9/1939 and C_i is a set of control variables. In detail, we include maternal age, number of pregnancy, a dummy for first pregnancy, (birth of) multiples, infant's sex, whether the mother is married, or working, a dummy for regular fetal position and a dummy for general ward. β measures the mean difference between the treatment and the control group conditional on observable characteristics.

In Section 3.3.2 we present descriptive evidence that the onset of the war rather than the duration of the war constitutes the actual shock driving our results. Therefore, as a next step, we include a time trend and its interaction with the treatment dummy in our regression equation:

$$y_i = \alpha + \delta \mathtt{war}_i + \lambda_0 \phi(\tilde{t}_i) + \lambda_1 \phi(\tilde{t}_i) * \mathtt{war}_i + \kappa C_i + \pi_i + u_i \tag{3.2}$$

$\tilde{t}_i$ denotes the time trend centered around the onset of the war. In the reported regressions we use a quadratic time trend, such that $\lambda_0 \phi(\tilde{t}_i) = \lambda_{01} \tilde{t}_i + \lambda_{02} \tilde{t}_i^{\,2}$.[19] π_i captures seasonality effects. The coefficient δ measures the jump in mortality at the threshold adjusting for covariates.

As shown in Figures 3.5 to 3.8 the time trend of mortality differs between the prewar and the war period. Furthermore the war might also change the structural relationship between observed characteristics and outcomes. For example the war might have increased the risk disproportionally for certain socio-economic groups. To account for these indirect effects, we additionally estimate an "Average Treatment Effect on the Treated" (ATET) using regression adjustment.[20] This approach is equivalent to estimating equation 3.2 separately for the treatment and the control group and then taking the difference in predicted outcomes under both sets of estimated coefficients for the treatment group. The ATET is constructed as follows:

$$\gamma_{\mathtt{war}}^{\mathtt{ATET}} = (\hat{\theta_0}^{\mathtt{war}} - \hat{\theta_0}^{\mathtt{nowar}}) + \frac{1}{N^{\mathtt{war}}} \sum_{\mathtt{i\ in\ war}} X_i (\hat{\theta}^{\mathtt{war}} - \hat{\theta}^{\mathtt{nowar}}) \tag{3.3}$$

[18] As mentioned earlier in Section 3.3.1 we exclude miscarriages from our main analysis.
[19] We also used a linear time trend and obtained similar results.
[20] For an explanation of regression adjustment see for example Uysal (2015).

X_i denotes all controls and the time trend. $\hat{\theta}^{\texttt{war}}$ and $\hat{\theta}^{\texttt{nowar}}$ are the estimated coefficients from the prewar and the war regression. $\gamma^{\texttt{ATET}}_{\texttt{war}}$ measures the average difference between the predicted effect for the treatment group and the predicted treatment effect for the treatment group if the treatment group had given birth before the war.

3.5 Results

3.5.1 Effect of war on perinatal health

Table 3.3, 3.6 and 3.8 present the effect of war on three measures of perinatal health, *birth weight*, *asphyxia* and *infant maturity*. Panel A shows regression estimates with the full sample (i.e. all births excluding miscarriages), while Panel B restricts the sample further to live births. Results in columns (1)-(4) are based on the entire observation period from 12/1937-9/1941, whereas columns (5)-(8) use only births likely to be conceived before the onset of WWII. We cluster all standard errors at birth level to adjust for twin births. ATETs estimated for the same outcome variables using regression adjustment are reported in separate tables (see Tables 3.4, 3.7 and 3.9 respectively). For neither sample we find any effect of the onset of the war on birth weight. The coefficient of war is small in size and insignificant in all but two specifications. This is in line with the descriptive analysis presented in Section 3.3.2 above.

Table 3.3: Effect of war - Birth weight

Panel A	All observations				Born before 6/1940			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth after 9/1939	-17.1	-21.3*	-19.8	-28.8	-21.5	-27.3*	-8.74	3.34
	(13.3)	(12.3)	(37.5)	(38.7)	(17.4)	(16.1)	(46.5)	(49.4)
Observations	8820	8361	8361	8361	5942	5624	5624	5624
Panel B	Only live births				Only live births born before 6/1940			
Birth after 9/1939	4.98	-8.24	-3.28	-18.0	-7.40	-18.0	18.8	27.5
	(12.2)	(11.5)	(35.5)	(36.2)	(16.2)	(15.2)	(42.5)	(44.9)
Observations	8472	8069	8069	8069	5717	5433	5433	5433
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Trend	No	No	Yes	Yes	No	No	Yes	Yes
Seasonality	No	No	No	Yes	No	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Controls include social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position and general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

Table 3.4: Effect of war - Birth weight - Regression adjustment

	All observations		Observations before 6/1940	
	(1) All births	(2) Live births	(3) All births	(4) Live births
ATET				
Born after 9/1939	-81.5	-43.4	-42.8	-25.9
	(149.4)	(142.4)	(60.5)	(58.1)
Observations	8361	8069	5624	5433

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender, dummy variables for regular fetal position and general ward and a quadratic time trend fitted on each side of the threshold separately.

Table 3.5: Effect of war by time of birth - Birth weight

Panel A	All observations		Live births	
Born 9-11/1339	10.9	-9.68	31.5	7.39
	(28.8)	(27.2)	(25.7)	(25.2)
Born 12/1939-2/1940	-30.2	-42.0*	-26.3	-40.9*
	(26.4)	(24.5)	(24.7)	(23.3)
Born 3-5/1940	-39.6	-32.2	-21.0	-18.9
	(26.4)	(23.0)	(24.5)	(21.8)
Born after 5/1940	-14.2	-16.1	13.2	-1.08
	(15.4)	(14.1)	(13.9)	(13.0)
Observations	8820	8361	8472	8069
Controls	No	Yes	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position and general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

As intrauterine growth takes place during the whole course of pregnancy, the war might manifest itself in lower birth weight only with a delay rather than the day after the war started. Furthermore the war shock may have different effects at different stages of pregnancy. Therefore we split the treatment variable into four categories, depending on whether the shock provided by the onset of the war occurred during late pregnancy (babies born 9-11 1939), during middle pregnancy (babies born 12/1939-2/1940), during early pregnancy (babies born 3-5 1940) or before the pregnancy even started. In the specifications with controls of Table 3.5 we don't see an effect on birth weight at any stage of pregnancy (except a small, weakly significant decrease in birth weight for infants born in the second trimester). Hence we are confident that the shock provided by the onset of the war did not reduce birth weight at any point during pregnancy.

Table 3.6: Effect of war - Asphyxia

Panel A	All observations				Born before 6/1940			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth after 9/1939	0.0028 (0.0039)	0.0034 (0.0040)	0.00064 (0.014)	-0.0061 (0.016)	-0.0018 (0.0044)	-0.00089 (0.0047)	0.012 (0.015)	0.0094 (0.020)
Observations	6784	6440	6440	6440	3906	3703	3703	3703
Panel B	Only live births				Only live births born before 6/1940			
Birth after 9/1939	0.0039 (0.0039)	0.0045 (0.0041)	0.000035 (0.014)	-0.0076 (0.016)	-0.00069 (0.0045)	0.00040 (0.0047)	0.0094 (0.015)	0.0080 (0.019)
Observations	6495	6196	6196	6196	3740	3560	3560	3560
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Trend	No	No	Yes	Yes	No	No	Yes	Yes
Seasonality	No	No	No	Yes	No	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Controls include social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position and general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

Table 3.7: Effect of war - Asphyxia - Regression adjustment

	All observations		Observations before 6/1940	
	(1) All births	(2) Live births	(3) All births	(4) Live births
ATET Born after 9/1939	0.072 (0.16)	0.045 (0.17)	0.015 (0.045)	0.0098 (0.046)
Observations	6440	6196	3703	3560

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender, dummy variables for regular fetal position and general ward and a quadratic time trend fitted on each side of the threshold separately.

Asphyxia was only consistently recorded after November 1938. Therefore we use a smaller sample when estimating the effects for asphyxia presented in Table 3.6. As in the case of birth weight we find a zero effect. Results for infant maturity are mixed (see Table 3.8 and 3.9). There is no significant difference in conditional and unconditional means between the treatment and the control sample. However we find evidence of a drop in the proportion of mature infants at the threshold. The estimates for the ATET in Table 3.9 are larger than the estimated regression coefficients.

Table 3.8: Effect of war - Maturity

Panel A	All observations				Born before 6/1940			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth after 9/1939	0.00044	-0.0047	-0.058***	-0.054**	0.00048	-0.0062	-0.061**	-0.044
	(0.0076)	(0.0075)	(0.021)	(0.022)	(0.0098)	(0.0097)	(0.025)	(0.027)
Observations	8814	8350	8350	8350	5937	5614	5614	5614
Panel B	Only live births				Only live births born before 6/1940			
Birth after 9/1939	0.0075	-0.00017	-0.051**	-0.051**	0.0055	-0.0020	-0.053**	-0.039
	(0.0073)	(0.0073)	(0.021)	(0.021)	(0.0095)	(0.0095)	(0.024)	(0.025)
Observations	8463	8058	8058	8058	5709	5423	5423	5423
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Trend	No	No	Yes	Yes	No	No	Yes	Yes
Seasonality	No	No	No	Yes	No	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Controls include social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position and general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

Table 3.9: Effect of war - Maturity - Regression adjustment

	All observations		Observations before 6/1940	
	(1) All births	(2) Live births	(3) All births	(4) Live births
ATET				
Born after 9/1939	-0.23**	-0.23***	-0.11***	-0.11***
	(0.090)	(0.087)	(0.035)	(0.034)
Observations	8350	8058	5614	5423

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender, dummy variables for regular fetal position and general ward and a quadratic time trend fitted on each side of the threshold separately.

3.5.2 Effect on perinatal mortality

We use the same specifications as in the previous analysis to estimate linear probability models for the effect of the onset of the war on perinatal mortality. The results are displayed in Table 3.10 and Table 3.11.

Table 3.10: Effect of war - Mortality

Panel A	All observations				Born before 6/1940			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth after 9/1939	0.025***	0.025***	0.047***	0.048***	0.035***	0.038***	0.035*	0.031
	(0.0053)	(0.0052)	(0.016)	(0.016)	(0.0075)	(0.0073)	(0.020)	(0.021)
Observations	8828	8363	8363	8363	5950	5626	5626	5626
Panel B	Only live births				Only live births born before 6/1940			
Birth after 9/1939	0.016***	0.016***	0.040***	0.040***	0.024***	0.026***	0.030**	0.026**
	(0.0035)	(0.0036)	(0.0099)	(0.010)	(0.0053)	(0.0055)	(0.012)	(0.012)
Observations	8477	8071	8071	8071	5722	5435	5435	5435
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Trend	No	No	Yes	Yes	No	No	Yes	Yes
Seasonality	No	No	No	Yes	No	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Controls include social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position and general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

Table 3.11: Effect of war - Mortality - Regression adjustment

	All observations		Observations before 6/1940	
	(1) All births	(2) Live births	(3) All births	(4) Live births
ATET				
Born after 9/1939	0.049	0.086***	0.040*	0.053***
	(0.057)	(0.033)	(0.023)	(0.013)
Observations	8363	8071	5626	5435

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender, dummy variables for regular fetal position and general ward and a quadratic time trend fitted on each side of the threshold separately.

Overall perinatal infant mortality increases significantly after the onset of WWII. Deaths in Panel A of Table 3.10 are made up of stillborn children as well live born children who die in hospital after birth. While 5% of births do not result in a living infant leaving the hospital in the prewar sample, this number increases to 7.5% in the war sample. Once we do not compare mean differences but the jump at the threshold in Column (3) and Column (4), the effect becomes even stronger. If we restrict the sample and drop all births which took place after May 1940, we see a larger difference in the means but a smaller jump. The ATET is larger than the regression coefficients in size but only significant in the restricted sample. Altogether these results support our interpretation that the onset of the war provided a shock which faded out gradually.

Effect of war by social class
In Table 3.12 we investigate whether the effect of war on mortality is heterogeneous by social class.[21] Parental social status is highly predictive of future live outcomes of their offspring. If war affects the composition of the population through the channel of selected mortality, this will be reflected in the live outcome of affected cohorts. In our specifications we omit the constant and the overall war dummy. Instead we report baseline levels and treatment effects of all social classes. The onset of the war has a non negative effect on mortality for all social groups. Higher professionals and managers - which constitute our highest social class - do suffer from the war, but also do lower skilled workers. Unskilled workers as well as Foremen & skilled workers are most severely affected. Hence we cannot document a clear socio-economic gradient in infant health/mortality which is commonly reported in the literature (see e.g. Aizer and Currie 2014; Cutler et al. 2006; Deaton 2013).

[21] See Appendix section 3.7.1 for a detailed description of our classification scheme.

Table 3.12: Effect of war by social class - Mortality

Panel A	All observations			Born before 6/1940		
	(1)	(2)	(3)	(4)	(5)	(6)
Higher managers & professionals	0.028**	-0.0079	0.020	0.028**	0.047	0.047
	(0.011)	(0.023)	(0.053)	(0.011)	(0.056)	(0.056)
Lower managers & professionals, cleric	0.044***	-0.010	0.019	0.044***	0.053	0.052
	(0.0079)	(0.020)	(0.053)	(0.0079)	(0.056)	(0.057)
Foremen & skilled workers	0.048***	-0.0070	0.022	0.048***	0.056	0.055
	(0.0072)	(0.020)	(0.053)	(0.0072)	(0.056)	(0.057)
Farmers	0.055***	-0.023	0.0054	0.055***	0.044	0.043
	(0.013)	(0.024)	(0.051)	(0.013)	(0.054)	(0.055)
Lower skilled workers	0.052***	0.00098	0.031	0.052***	0.065	0.065
	(0.010)	(0.021)	(0.054)	(0.010)	(0.057)	(0.058)
Unskilled workers	0.056***	-0.0034	0.026	0.056***	0.061	0.061
	(0.0069)	(0.020)	(0.054)	(0.0069)	(0.056)	(0.058)
Farm workers	0.056***	0	0.027	0.056***	0.065	0.063
	(0.020)	(.)	(0.057)	(0.020)	(0.060)	(0.061)
War * Higher managers & professionals	0.0096	0.019	0.042**	0.048*	0.046**	0.040
	(0.015)	(0.015)	(0.021)	(0.026)	(0.023)	(0.029)
War * Lower managers & professionals, cleric	0.032***	0.031***	0.053***	0.023	0.025*	0.019
	(0.012)	(0.011)	(0.018)	(0.015)	(0.014)	(0.024)
War * Foremen & skilled workers	0.027**	0.028***	0.051***	0.055***	0.056***	0.050**
	(0.011)	(0.010)	(0.018)	(0.017)	(0.015)	(0.024)
War * Farmers	0.060**	0.055**	0.077***	0.058*	0.052*	0.046
	(0.024)	(0.023)	(0.028)	(0.031)	(0.029)	(0.036)
War * Lower skilled workers	0.0060	-0.0048	0.018	0.0053	0.0031	-0.0031
	(0.014)	(0.013)	(0.020)	(0.019)	(0.018)	(0.026)
War * Unskilled workers	0.032***	0.029***	0.051***	0.045***	0.044***	0.038
	(0.011)	(0.010)	(0.019)	(0.016)	(0.015)	(0.025)
War * Farm workers	0.034	0.0075	0.031	0.058	0.051	0.046
	(0.039)	(0.036)	(0.039)	(0.068)	(0.064)	(0.067)
Observations	8500	8363	8363	5729	5626	5626
Panel B	Live births			Live births born before 6/1940		
Higher managers & professionals	0.014*	-0.0094	-0.017	0.014*	0.014	0.019
	(0.0082)	(0.041)	(0.041)	(0.0082)	(0.013)	(0.047)
Lower managers & professionals, cleric	0.018***	-0.018	-0.025	0.018***	0.011	0.014
	(0.0052)	(0.041)	(0.041)	(0.0052)	(0.011)	(0.047)
Foremen & skilled workers	0.018***	-0.019	-0.026	0.018***	0.010	0.014
	(0.0045)	(0.041)	(0.041)	(0.0045)	(0.010)	(0.047)
Farmers	0.013*	-0.027	-0.035	0.013	0.0015	0.0048
	(0.0078)	(0.039)	(0.039)	(0.0078)	(0.013)	(0.044)
Lower skilled workers	0.017***	-0.019	-0.025	0.017***	0.011	0.015
	(0.0061)	(0.042)	(0.042)	(0.0061)	(0.011)	(0.048)
Unskilled workers	0.018***	-0.024	-0.031	0.018***	0.0064	0.0099
	(0.0040)	(0.042)	(0.042)	(0.0040)	(0.0096)	(0.048)
Farm workers	0.0084	-0.030	-0.039	0.0084	0	0.0026
	(0.0084)	(0.043)	(0.043)	(0.0084)	(.)	(0.049)
War * Higher managers & professionals	0.0076	0.010	0.034**	0.018	0.018	0.018
	(0.011)	(0.013)	(0.016)	(0.018)	(0.018)	(0.020)
War * Lower managers & professionals, cleric	0.018**	0.018**	0.041***	0.017	0.018*	0.019
	(0.0082)	(0.0081)	(0.012)	(0.011)	(0.010)	(0.016)
War * Foremen & skilled workers	0.019**	0.022***	0.046***	0.037***	0.040***	0.040***
	(0.0074)	(0.0072)	(0.012)	(0.012)	(0.012)	(0.016)
War * Farmers	0.039**	0.031*	0.055***	0.031	0.018	0.018
	(0.016)	(0.016)	(0.018)	(0.021)	(0.017)	(0.020)
War * Lower skilled workers	0.0065	0.0055	0.029**	0.012	0.013	0.013
	(0.0091)	(0.0090)	(0.013)	(0.014)	(0.013)	(0.017)
War * Unskilled workers	0.014**	0.014**	0.038***	0.029**	0.029***	0.029**
	(0.0069)	(0.0068)	(0.012)	(0.011)	(0.011)	(0.015)
War * Farm workers	0.016	0.0093	0.034	0.052	0.046	0.048
	(0.025)	(0.023)	(0.026)	(0.059)	(0.053)	(0.055)
Observations	8164	8071	8071	5512	5435	5435
Controls	No	Yes	Yes	No	Yes	Yes
Trend	No	No	Yes	No	No	Yes
Seasonality	No	No	Yes	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Controls include mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position and general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

Effect of war by birth weight

Like social class, birth weight is highly correlated with later life outcomes. If low birth weight infants are more likely to die as a consequence of the war, negative effects of war on later live outcomes will be underestimated in studies based on surviving individuals.

Table 3.13: Effect of war by birth weight - Mortality

Panel A	All observations			Born before 6/1940		
	(1)	(2)	(3)	(4)	(5)	(6)
Birth weight below 2000 grams	0.63***	0.67***	0.69***	0.63***	0.69***	0.70***
	(0.044)	(0.061)	(0.062)	(0.044)	(0.064)	(0.065)
Birth weight 2000-2499 grams	0.14***	0.22***	0.24***	0.14***	0.24***	0.25***
	(0.024)	(0.051)	(0.051)	(0.024)	(0.056)	(0.056)
Birth weight 2500-3999 grams	0.028***	0.11***	0.13***	0.028***	0.13***	0.14***
	(0.0060)	(0.043)	(0.044)	(0.0060)	(0.048)	(0.049)
Birth weight 3000 grams and above	0.024***	0.10**	0.12***	0.024***	0.12**	0.13***
	(0.0028)	(0.043)	(0.043)	(0.0028)	(0.048)	(0.048)
War*Birth weight below 2000 grams	0.13**	0.14**	0.15**	0.18***	0.19***	0.18***
	(0.055)	(0.058)	(0.059)	(0.062)	(0.065)	(0.066)
War*Birth weight 2000-2499 grams	0.054	0.059	0.068*	0.049	0.054	0.043
	(0.036)	(0.036)	(0.038)	(0.048)	(0.048)	(0.051)
War*Birth weight 2500-3999 grams	0.039***	0.036***	0.046***	0.045***	0.046***	0.035
	(0.010)	(0.010)	(0.016)	(0.015)	(0.015)	(0.022)
War*Birth weight 3000 grams and above	0.0049	0.0087**	0.019	0.012**	0.017***	0.0065
	(0.0040)	(0.0040)	(0.013)	(0.0057)	(0.0058)	(0.017)
Observations	8820	8361	8361	5942	5624	5624
Panel B	Live births			Live births born before 6/1940		
Birth weight below 2000 grams	0.44***	0.51***	0.51***	0.44***	0.55***	0.54***
	(0.056)	(0.067)	(0.067)	(0.056)	(0.071)	(0.071)
Birth weight 2000-2499 grams	0.072***	0.16***	0.15***	0.072***	0.19***	0.18***
	(0.019)	(0.042)	(0.042)	(0.019)	(0.049)	(0.049)
Birth weight 2500-3999 grams	0.0081**	0.092***	0.088***	0.0081**	0.13***	0.12***
	(0.0033)	(0.034)	(0.034)	(0.0033)	(0.041)	(0.041)
Birth weight 3000 grams and above	0.0031***	0.084**	0.080**	0.0031***	0.12***	0.11***
	(0.0010)	(0.033)	(0.033)	(0.0010)	(0.040)	(0.040)
War*Birth weight below 2000 grams	0.14*	0.15*	0.16**	0.25***	0.24***	0.24***
	(0.075)	(0.077)	(0.077)	(0.086)	(0.088)	(0.088)
War*Birth weight 2000-2499 grams	0.022	0.025	0.041	0.050	0.045	0.048
	(0.028)	(0.028)	(0.029)	(0.041)	(0.041)	(0.040)
War*Birth weight 2500-3999 grams	0.022***	0.020***	0.036***	0.035***	0.034***	0.036***
	(0.0068)	(0.0069)	(0.010)	(0.011)	(0.011)	(0.013)
War*Birth weight 3000 grams and above	0.0083***	0.0088***	0.025***	0.0080***	0.0098***	0.012
	(0.0021)	(0.0023)	(0.0080)	(0.0030)	(0.0033)	(0.0096)
Observations	8472	8069	8069	5717	5433	5433
Controls	No	Yes	Yes	No	Yes	Yes
Trend	No	No	Yes	No	No	Yes
Seasonality	No	No	Yes	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position and general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

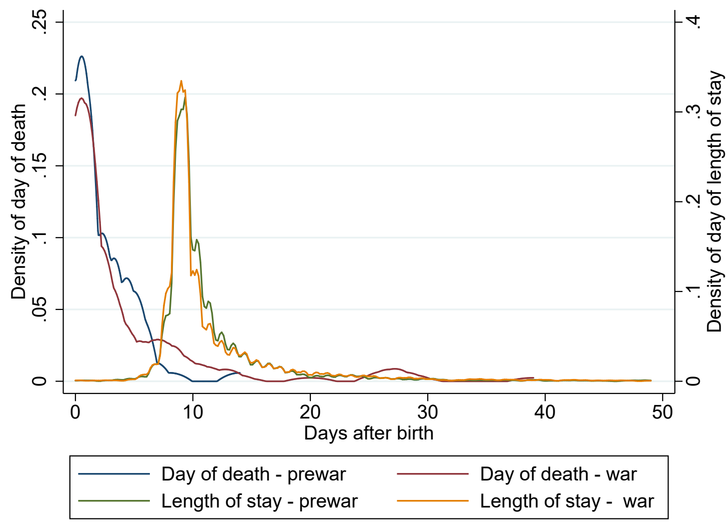
To explore heterogeneity by birth weight, we split our sample at 2000, 2500 and 3000 grams. Table 3.13 displays the estimated treatment effects for all four groups. We find a clear mortality-birth weight gradient in the treatment effect. In any of the specifications the magnitude of the estimated coefficient of the interaction between group effect increases when birth weight decreases. However, due to a small sample size, the effect is not significant for children whose birth weight is below the common low birth weight threshold at 2500 grams but above 2000 grams. For very low birth weight children with less than 2000 grams at birth, the effect is largest. The probability to leave the hospital alive decreases by more than 10 percentage points.[22] Also children born between 2500 and 3000 grams are affected to a larger extend than the group of children above 3000 grams.

3.5.3 Robustness check

Length of stay

While the conventional definition of neonatal mortality includes deaths up to 28 days after birth (WHO 2006), we only observe newborns until they leave the hospital. As long as the day of discharge and our treatment are independent, our definition of infant mortality will not pose a thread to identification.

Figure 3.12: Length of stay and day of death



Notes: Distribution of length of stay in hospital and length of life in days for live born children.

Figure 3.12 shows the distribution of length of stay and length of life in days for live born children separately for the prewar and the war period.[23] First we notice that there is hardly any difference in the distribution of the day of discharge in our treatment and control group. Most observations stay in hospital for around 9-10 days after birth

[22] A surprisingly large number of infants below 2000 grams survives. We checked the most extreme cases in the birth records carefully but found no sign of misreporting. In one case we found a letter stating that a child born at around 1300 grams had left the hospital and was doing well.

[23] To facilitate legibility, we exclude a small number of observations who stayed in hospital for more than 50 days.

and only 1.5% of live born children are discharged before completing the first week of life. Neonatal deaths on the other hand mostly occur within the first four days after birth. Since mothers received postnatal care in hospital the death of an infant does not automatically lead to a discharge of the mother. As a robustness check we estimate our regressions using modified versions of infant mortality. We define an infant to have died if the death occurred either in the first five days (see Table 3.14) or the first seven days (see Table 3.15) after birth. In these specifications we exclude all observations which left the hospital before that specific day. Although the coefficients become smaller in size, we still see a significant effect of the onset of the war on perinatal infant mortality.

Table 3.14: Effect of war - Mortality - Death within five days

Panel A	All observations				Born before 6/1940			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth after 9/1939	0.019***	0.019***	0.036**	0.036**	0.025***	0.026***	0.038*	0.033
	(0.0050)	(0.0050)	(0.015)	(0.015)	(0.0069)	(0.0067)	(0.020)	(0.021)
Observations	8762	8305	8305	8305	5907	5589	5589	5589
Panel B	Only live births				Only live births born before 6/1940			
Birth after 9/1939	0.0085***	0.0086***	0.026***	0.024***	0.011**	0.012***	0.031***	0.024**
	(0.0031)	(0.0032)	(0.0090)	(0.0090)	(0.0043)	(0.0044)	(0.011)	(0.011)
Observations	8426	8023	8023	8023	5689	5404	5404	5404
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Trend	No	No	Yes	Yes	No	No	Yes	Yes
Seasonality	No	No	No	Yes	No	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, parity, primipara, twinning status, infant's gender, marital status, a dummy for general ward and working status; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.. Only cases staying in the hospital at least days.

Table 3.15: Effect of war - Mortality - Death within seven days

Panel A	All observations				Born before 6/1940			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth after 9/1939	0.020***	0.020***	0.039**	0.038**	0.028***	0.030***	0.034*	0.029
	(0.0050)	(0.0050)	(0.015)	(0.016)	(0.0070)	(0.0069)	(0.020)	(0.021)
Observations	8669	8219	8219	8219	5836	5523	5523	5523
Panel B	Only live births				Only live births born before 6/1940			
Birth after 9/1939	0.0093***	0.0097***	0.029***	0.026***	0.014***	0.015***	0.029**	0.021*
	(0.0032)	(0.0033)	(0.0093)	(0.0094)	(0.0046)	(0.0047)	(0.011)	(0.012)
Observations	8343	7944	7944	7944	5625	5344	5344	5344
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Trend	No	No	Yes	Yes	No	No	Yes	Yes
Seasonality	No	No	No	Yes	No	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, parity, primipara, twinning status, infant's gender, marital status, a dummy for general ward and working status; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.. Only cases staying in the hospital at least days.

Temperature

In the first two months of 1940, Munich was hit by a particularly low temperatures (Stadtarchiv München 1939-1940). To rule out, that the effect we measure is in fact a shock caused by low temperatures, we include the monthly temperatures in Munich as additional control variables. The results are presented in Table 3.16. The estimated coefficients hardly change compared to the baseline estimates suggesting that temperature does not confound our baseline estimates.

Table 3.16: Effect of war - Mortality - Temperature

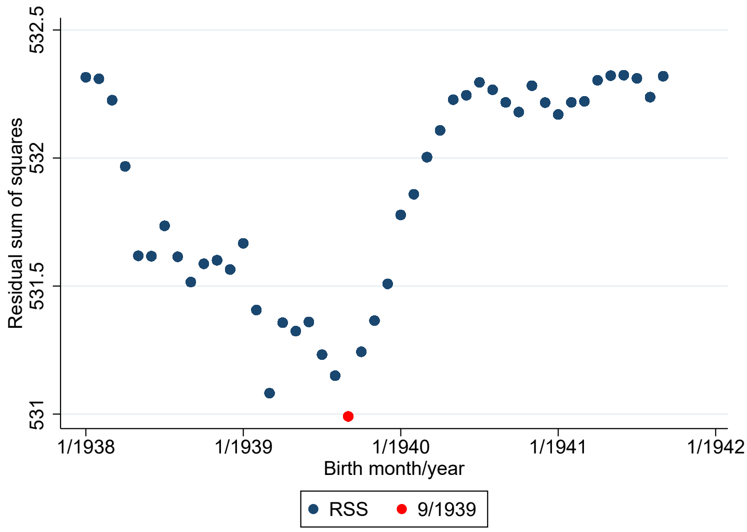
Panel A	All observations				Born before 6/1940			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth after 9/1939	0.025***	0.025***	0.044***	0.047***	0.035***	0.038***	0.035*	0.031
	(0.0053)	(0.0052)	(0.016)	(0.016)	(0.0075)	(0.0075)	(0.020)	(0.021)
Observations	8828	8363	8363	8363	5950	5626	5626	5626
Panel B	Only live births				Only live births born before 6/1940			
Birth after 9/1939	0.016***	0.016***	0.038***	0.040***	0.024***	0.026***	0.030**	0.026**
	(0.0035)	(0.0036)	(0.010)	(0.010)	(0.0053)	(0.0056)	(0.012)	(0.012)
Observations	8477	8071	8071	8071	5722	5435	5435	5435
Temperature	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Trend	No	No	Yes	Yes	No	No	Yes	Yes
Seasonality	No	No	No	Yes	No	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position, general ward and the average temperature in Munich for the current month; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

Structural break

In our empirical specification we estimate infant mortality as a function of maternal characteristics and time variables. In the regression adjustment we allow this function to differ between the prewar and the war period. To investigate whether such a structural break actually took place in September 1939, we shift our treatment indicator between January 1938 and September 1941. For each month we estimate Equation 3.1 without the war dummy separately to both sides of the threshold and calculate the sum of residual sum of squares. If no structural break occurred during our period of observation, the residual sum of squares will not exhibit any systematic pattern (Hansen 2001). However, Figure 3.13 depicts a clear trend. The residual sum of squares decreases when shifting the treatment indicator from January 1938 to September 1939 with September 1939 being a global minimum. When shifting the treatment indicator further into the war period, the residual sum of squares increases. This indicates that

Figure 3.13: Structural break analysis - Mortality



Notes: Residual sum of squares for infant mortality. We estimate the regression model $\texttt{Death}_i = \texttt{Controls}_i\beta + \epsilon_i$ separately for all births prior to month m and births in month m or later. m is shifted from January 1938 to September 1941. RSS denotes the combined sum of residual sum of squares.

the begin of WWII indeed marked a breakpoint changing the relationship between maternal characteristics and infant mortality.

3.5.4 Mechanism

In our setting we can rule out direct effects of the war like hunger, bombing or displacement. To explain the increase in perinatal mortality, we focus on two potential channels already present in fall 1939 - maternal stress and a decline in medical quality. First increased uncertainty and drafting of husbands are likely to increase stress levels of pregnant women. The literature has shown that uncertainty and maternal stress during pregnancy can affect the newborn negatively in the short-run (see Currie and Rossin-Slater 2013). Second the onset on the war did not only put a strain on individuals, but drafting of experienced physicians led to staff shortage in the hospital. All physicians working at the hospital in August 1939 were male and therefore potentially liable to military duty. Inexperienced graduates and unpaid trainees served as temporary replacements. In letters to the state administration, the director of the hospital maintained that the hospital routine threatened to break down and proper patient care was in jeopardy. Especially when the first physicians left in September 1939, it came as an unexpected shock for the hospital, while it could prepare for further drafts. Intelligence reports which captured and reported the public opinion to the Nazi government (since there wasn't any free press anymore) confirm that there was a severe shortage of doctors in several parts of the German Reich after the begin of the war which in turn lead to public dissatisfaction and turmoil (see Boberach 1984).

We cannot fully quantify these mechanisms. However, in the following we provide evidence that our results are mainly driven by a decline in medical quality. We show that the mortality effect is stronger, where medical quality should matter and furthermore we document a change in provision of certain medical procedures.

Unlike birth weight which is measured at birth and miscarriage, survival of life born children is partly under the control of the medical personnel. If live born children are disproportionally affected by the war, this will hint to a decline in medical care. Maternal stress on the other hand should lead to an increase in stillbirths and miscarriages. Panel B of Table 3.10 conducts the regression analysis for the sample of live born children. Given the low baseline mortality of live born children before the war of 1.8 % the effect of the war is surprisingly large. Between 9/1939 and 5/1940 the mortality of live born children almost doubles compared to the prewar period. Again we find that the jump around the threshold is larger than the differences in means.

Table 3.17: Effect of war - Mortality - Non-live births

Panel A: Stillbirth	Births				Births before 6/1940			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Birth after 9/1939	0.0098**	0.013***	0.012	0.012	0.012**	0.015***	0.014	0.013
	(0.0042)	(0.0043)	(0.014)	(0.014)	(0.0057)	(0.0058)	(0.018)	(0.019)
Observations	8828	8499	8499	8499	5950	5728	5728	5728
Panel B: Miscarriage	All observations				Observations before 6/1940			
Birth after 9/1939	0.0074	0.0036	-0.018	-0.023	-0.015*	-0.015*	0.018	0.013
	(0.0065)	(0.0066)	(0.021)	(0.021)	(0.0081)	(0.0081)	(0.025)	(0.027)
Observations	10022	9617	9617	9617	6689	6416	6416	6416
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Trend	No	No	Yes	Yes	No	No	Yes	Yes
Seasonality	No	No	No	Yes	No	No	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status and a dummy variable for general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by a set of month dummies.

The proportion of stillborn children also increases after the onset of WWII (see Panel A of Table 3.17). However, the effect is less than the increase in mortality of live born children and not robust to the inclusion of a time trend and seasonality effects. Therefore our overall effects seem to be driven by children who die in hospital after birth. While we did exclude miscarriages from the main analysis, we estimate the effect of war on the probability of miscarriage in Panel B of Table 3.17. We do not find any

evidence, that the onset of WWII lead to an increased number of miscarriages.
A shortage of physicians is likely to shift work from physicians to female midwives. Whereas midwives are able to supervise normal deliveries, only physicians can carry out surgeries like caesarean sections. We test whether women who should receive a caesarean section by modern standards are less likely to receive a caesarean section during the first months of the war. We construct a measure of whether a women has an indication for caesarean section based on guideline described in Mylonas and Friese (2015).[24] As shown in columns (1)-(2) of Table 3.18, the proportion of women with an indication for caesarean section does not change with the onset of the war. However, women with an indication are less likely to actually have a section performed. Instead we see the performance of another procedure. Symphysiotomy is an operation to widen the pelvis that can be carried out by non-specialist doctors and experienced midwives (see Monjok et al. 2012). It was frequently used in the 19th century, when caesarean section was a high risk for mothers. Due to negative consequences for maternal health today's WHO guidelines recommend the use of symphysiotomy only, when safe caesarean sectio is not available (WHO 2003). This result shows that the hospital replaced procedures in need of an experienced surgeon by simpler procedures. We also investigate how the use of medical procedures changes in less severe cases. We look at episiotomy, a simple procedure to prevent perineal tear. While perineal tear can be painful for the mother, it is not a live threatening condition. Columns (7)-(8) of Table 3.18 show a small decrease in the use of this procedure, but this is not reflected in a higher incidence of perineal tear.

[24] We assume a women to have an indication if one of the following conditions is present: Non regular fetal position, eclampsia, placenta previa, disproportion of pelvis and child, uterine rupture. We do not include the condition umbilical cord prolapse, since none of the cases with umbilical cord prolapse is treated with caesarean section in our sample.

Table 3.18: Effect of war - Medical procedures - Births before June 1940

	Indication		Caesarean sectio		Symphysiotomy		Episiotomy		Perineal tear	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Birth after 9/1939	-0.00044	-0.013	0.00021	0.017*	-0.0014**	0.0093	-0.0054	-0.031**	-0.0096	-0.014
	(0.0040)	(0.013)	(0.0013)	(0.0090)	(0.00066)	(0.0075)	(0.0042)	(0.015)	(0.0088)	(0.025)
Indication for caesarean			0.22***	0.22***	0.093***	0.093***				
			(0.041)	(0.041)	(0.029)	(0.029)				
War * Indication			-0.046*	-0.046*	0.045*	0.044*				
			(0.028)	(0.028)	(0.024)	(0.024)				
Observations	5626	5626	5626	5626	5626	5626	5626	5626	5626	5626
Trend + Seasonality	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: (Clustered) Standard errors in parentheses; Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; All regressions include the following controls: Social status, mother's age, marital status, working status, parity, primipara, twinning status, infant's gender and dummy variables for regular fetal position and general ward; Trend denotes a quadratic time trend fitted on each side of the threshold separately; Seasonality is captured by quarter of birth.

3.6 Conclusion

In this chapter, we investigate the effects of the onset of WWII on health at birth and perinatal mortality. We use a unique data set of historical birth records from Munich's largest birth hospital covering the period 1937–1941. Exploiting the onset of WWII as natural experiment, we show that the onset of the war had no effect on health at birth measured by birth weight, maturity and asphyxia. However, we document an increase in perinatal mortality. This effect is strongest at the beginning of the war and fades out gradually. Additional analyses reveal that perinatal mortality increases after the begin of the war for all social classes and especially for newborns below 2000 grams.
Since the data cover only the onset of WWII, we can rule out direct effects of the war, like hunger, bombings or flight. We discuss two potential mechanisms to explain the increase in mortality. On the one hand increased uncertainty and drafting of husbands are likely to increase stress levels of pregnant women and may therefore lead to this mortality increase. On the other hand according to letters from the head physician the drafting of experienced physicians led to severe staff shortage and later to a decrease in the quality of medical care due to the replacement with untrained medical students. To evaluate the importance of each mechanism we investigated whether war affected the proportion of women with an indication for a caesarean section, which was not the case. However women with an indication, are less likely to actually have a sectio performed. Instead the probability of other less complicated birth procedures increases which can be performed by auxiliary medical staff, but which are less safe. In combination all these results point to the deterioration of the quality of medical inputs (i.e. doctors) as the main driver of our results.

3.7 Appendix

3.7.1 HISCO-HISCLASS

As mentioned in the main text the birth journals - both from the general and private ward - contain parental status and/or occupation. This allows us to derive a measure of social classes which will be explained in this section. These specific occupations were originally recorded by hospital personnel with additional (grand-)parental socio-economic information in the medical files to ensure Aryan ancestry and the patient's health insurance among other things.[25] If fatherhood is known and stated in the medical files, usually the (civil) profession of the child's father (e.g. grocer, in German "Krämer") or her relation to him (e.g. grocer's wife, in German "Krämersfrau") was registered in the birth journals.[26]

In a first step we standardized the spelling of the occupations (to the male form) and separated non-occupational information. In the following step we assigned a numerical 5-digit code according to the "Historical International Classification of Occupations" (HISCO) which was developed by Van Leeuwen et al. (2002) and is provided as an online database called "History Of Work Information System".[27] HISCO combines information on occupational tasks and duties and forms a system of 1675 historical and international comparable occupations. It was developed upon ILO's modern-day "International Classification of Occupations"' from 1968 (ISCO68) and adjusted with 18th-20th century occupations from several countries in Europe and America. HISCO's hierarchical structure - similar to ISCO68 - into 9 major groups, 76 minor groups, 296 unit groups and finally 1675 occupation has descriptions for each level and therefore allows comparisons to modern-day occupational groups and professions as well. HISCO has three additional variables (Status, Relation, Product) from which the variable "Status" is the most important one. It contains information about supervisory tasks and skill levels within an occupation (e.g. master backer, journeyman baker, apprentice baker, baker's helper) which would otherwise be lost because HISCO codes only incorporate the raw definition of an occupation (e.g. baker). In a last step we translated the

[25] Due to data privacy regulations we were not able to use this valuable information.

[26] In very rare cases the relation of the pregnant woman to her father's occupation (e.g. grocer's daughter in German "Krämerstochter") was entered if she was too young to have a own job and most likely unwed. In other cases the female notation of an occupation was recorded (e.g. in German "Krämerin"). This is a sign that the pregnant woman is unwed, but in some cases it might just indicate her job. For some observations more and/or other non-occupational information is recorded, e.g. "unwed", "student" or "housewife". If just "housewife" was recorded, a cross-check with the medical files most often revealed the relevant occupation.

[27] http://historyofwork.iisg.nl [last accessed: 17 September 2017].

HISCO into HISCLASS codes, the measure of social status which we will later use in our empirical analysis.[28] The Historical International Social Class Scheme (HISCLASS) invented and explained by Van Leeuwen and Maas (2011) builds upon HISCO, assigns each occupational code one out of 12 social classes, and defines a social class as "a set of individuals with the same life chances" (Van Leeuwen and Maas 2011, p.18). These social classes are derived in the following step-wise procedure: First Van Leeuwen and Maas (2011) identified (1)"type of work" (manual vs. non-manual), (2) "skill level" (4 levels), (3) "supervisory tasks" (yes vs. no) and (4) economic sector (primary vs. other sectors) as the four relevant dimensions of social class through an intensive literature review of existing historical class/status schemes. Second they used the American Dictionary of Occupational Titles (DOT) to grade the 1675 HISCO occupations along these dimensions. Third if there is additional information over and above the simple occupation name (e.g. baker) in the "Status" variable mentioned before (e.g. master, apprentice, helper, etc.) this is taken into account by promoting or demoting individuals into a higher/lower social class respectively. Since the DOT was constructed for modern-day occupations these grades were adjusted with help of expert historians which was only necessary in a few cases and finally led to 12 distinct social classes. In our empirical analysis we rely on the previous literature and use a compressed 7-class version of HISCLASS (see Abramitzky et al. (2011) and Schumacher and Lorenzetti (2005) and references therein). This simplifies the interpretation of regression coefficients, attenuates possible coding errors and increases sample size within classes. Table 3.19 shows the original and compressed HISCLASS versions along with the underlying dimensions of social class and the number of observations in each class.

[28] For the actual translation we relied on a SPSS programme provided in the "History Of Work Information System" database (http://historyofwork.iisg.nl/docs/hisco_hisclass12_book@_numerical.inc [last accessed: 17 September 2017].) which we corrected and translated into a Stata. A commented do-file is available on request from the authors.

Table 3.19: Original and compressed HISCLASS

HISCLASS				Dimensions of social class				# of obs.
Original classification		Compressed classification		Type	Skill	Supervisory	Economic	in
Nr	Name	Nr	Name	of work	level	tasks	sector	dataset
1	Higher managers	1	Higher managers & professionals	non-manual	high	yes	mainly other	292
2	Higher professionals					no		397
3	Lower managers	2	Lower managers clerical and professionals, clerical and sales		medium	yes		554
4	Lower professionals and clerical and sales personnel					no		507
5	Lower clerical and sales personnel				low			968
6	Foremen	3	Skilled workers	manual	medium	yes		222
7	Medium skilled workers					no		2,014
8	Farmers and fishermen	4	Farmers and fishermen				primary	664
9	Lower skilled workers	5	Lower-skilled workers		low		other	1,343
10	Lower skilled farm workers	6	Farm workers				primary	118
11	Unskilled workers	7	Unskilled workers		unskilled		other	2,719
12	Unskilled farm workers	6	Farm workers				primary	115

Source: In style of Van Leeuwen and Maas (2011) and Schumacher and Lorenzetti (2005) **Notes:** Other economic sector refers to the industrial or service sector. Classes 1 and 3 of the original HISCLASS system contain only 3 occupations which are in the primary sector.

To adjust HISCO/HISCLASS to the specific Bavarian background and our dataset sometimes we had to refine or deviate from the suggested coding procedure by Van Leeuwen et al. (2002) and Van Leeuwen and Maas (2011). First and foremost we had to rely on the occupational information about the child's father/ pregnant woman's husband in most cases, because either the pregnant women were not working at all or their own job was not recorded in the birth journals. This is in contrast to Van Leeuwen and Maas (2011) who don't assign a HISCLASS code to them at all. Nevertheless we are confident that this measure captures the relevant social class of a family since - as mentioned in Section 3.2.1 - Nazi-propaganda promoted housewife-dom and the husband was the head of the most households. Second due to the fact that Munich is a state capital, there are a lot of public sector occupations which were strictly hierarchically ranked according to the "Führerprinzip" and comparable to military ranks.[29] This allowed us to use equivalent HISCLASS codes of military ranks as a benchmark for police, postal, railway, educational and other governmental HISCO codes and adjust the previously assigned HISCLASS codes if there were large discrepancies.

3.7.2 Birth records

Almost all birth records since the foundation of the hospital in 1884 have been preserved. Birth records span around four to eight pages and generally contain background information on the mother, information on the pregnancy, medical examinations, a labour protocol including detailed notes, characteristics of the newborn child and observations during childbed. A compressed version of the birth records is provided by two series of journals, called birth journals and main journals. Birth journals have been filled in by midwives shortly after childbirth. Main journals make a more official appearance, suggesting that they were kept by a hospital clerk. Both journals contain the birth number, name, age and parity of the mother, the date of birth, sex, length and weight of the child, and short notes on medical issues. Main journals additionally give the date of discharge and the fetal position. Birth journals include information on the socio-economic status, mostly in form of the occupation of either the father or the mother. Main journals are only available for the common section. We digitized the information contained in the main journal and birth journals for a period starting in November 1937 and ending in October 1941. Since main journals do not exist for the private section, the date of discharge and the fetal position were added from the birth records. Apart from birth records, parts of the correspondence of the management

[29] In English: "leader principle" (see Frei 2013)

of the hospital have been preserved in archives and the hospital itself. We use this material to corroborate our findings with qualitative evidence.

3.7.3 Additional tables and figures

Table 3.20: Live births 1938-1940

	Hospital	Munich	Bavaria
1938	2171	12164	168391
1939	2297	13028	179129
1940	2269	13741	174311

Source: Bayerisches Statistisches Landesamt (1937-1942)

Table 3.21: Descriptive statistics - Miscarriages

General characteristics	N	Mean	SD	Min	Max
Birth after 9/1939	1194	0.560	0.497	0	1
General ward	1194	0.956	0.204	0	1
Mother	N	Mean	SD	Min	Max
Age of mother	1194	29.775	6.364	14	48
Parity	1184	2.994	2.308	1	18
Status is wife	1194	0.680	0.467	0	1
Status is own job	1194	0.270	0.444	0	1
Status is single, divorced or widowed	1194	0.033	0.178	0	1
Social status	N	Mean	SD	Min	Max
Higher managers & professionals	1125	0.063	0.243	0	1
Lower managers & professionals, cleric	1125	0.276	0.447	0	1
Foremen & skilled workers	1125	0.238	0.426	0	1
Farmers	1125	0.028	0.166	0	1
Lower skilled workers	1125	0.156	0.363	0	1
Unskilled workers	1125	0.228	0.419	0	1
Farm workers	1125	0.012	0.107	0	1
Infant	N	Mean	SD	Min	Max
Male	174	0.667	0.473	0	1
Birth weight	146	389.322	272.975	20	1870
Length of infant	178	24.458	8.197	9	90
No. of infants	1194	1.020	0.140	1	2

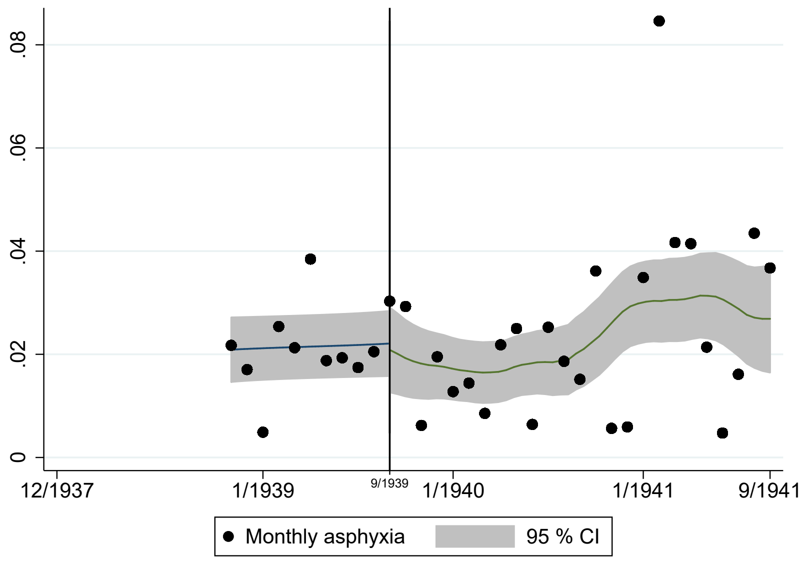
Notes: Descriptive statistics of miscarriages.

Table 3.22: Mean comparison - Miscarriages

General characteristics	Mean before war	Mean after war	Diff	SD	p	N before war	N after war
General ward	0.950	0.961	0.0107	0.012	0.371	525	669
Mother	Mean before war	Mean after war	Diff	SD	p	N before war	N after war
Age of mother	29.667	29.859	0.1928	0.371	0.603	525	669
Parity	3.033	2.964	-0.0689	0.135	0.611	518	666
Status is wife	0.632	0.717	0.0851**	0.027	0.002	525	669
Status is own job	0.310	0.238	-0.0728**	0.026	0.005	525	669
Status is single, divorced or widowed	0.036	0.030	-0.0063	0.010	0.544	525	669
Social status	Mean before war	Mean after war	Diff	SD	p	N before war	N after war
Higher managers & professionals	0.077	0.052	-0.0249	0.015	0.089	493	632
Lower managers & professionals, cleric	0.252	0.294	0.0428	0.027	0.111	493	632
Foremen & skilled workers	0.209	0.261	0.0522*	0.026	0.042	493	632
Farmers	0.037	0.022	-0.0144	0.010	0.151	493	632
Lower skilled workers	0.152	0.158	0.0061	0.022	0.780	493	632
Unskilled workers	0.260	0.203	-0.0571*	0.025	0.023	493	632
Farm workers	0.014	0.009	-0.0047	0.006	0.464	493	632
Infant	Mean before war	Mean after war	Diff	SD	p	N before war	N after war
Male	0.609	0.724	0.1149	0.071	0.109	87	87
Birth weight	392.877	385.767	-7.1096	45.336	0.876	73	73
Length of infant	24.333	24.585	0.2519	1.232	0.838	90	88
No. of infants	1.019	1.021	0.0019	0.008	0.819	525	669

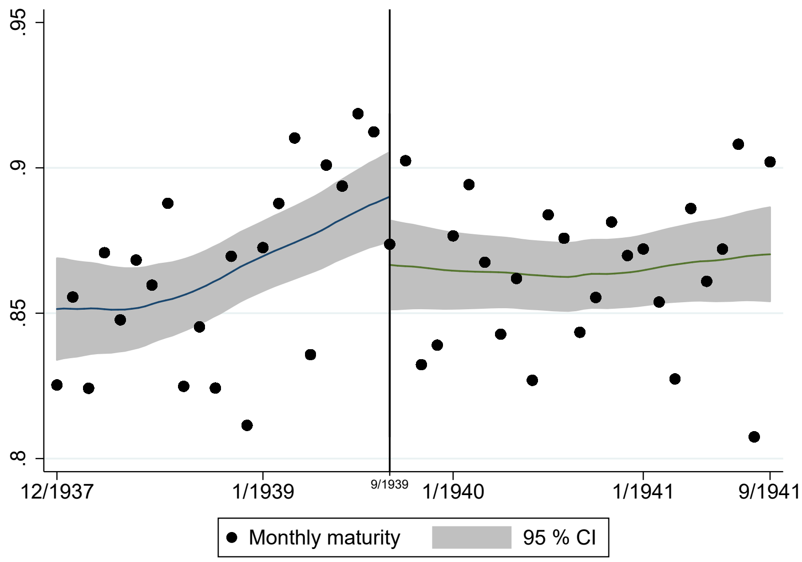
Notes: T-tests on the equality of means by war. Only miscarriages. Significance levels: ***$p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Figure 3.14: Raw asphyxia rates by month of birth



Notes: Asphyxia rates (monthly averaged) and local linear regressions with a ROT bandwidth and an Epanechnikov kernel separately for the prewar and the war period.

Figure 3.15: Raw average maturity by month of birth



Notes: Maturity (monthly averaged) and local linear regressions with a ROT bandwidth and an Epanechnikov kernel separately for the prewar and the war period.

Bibliography

Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos (2011). "Marrying up: The Role of Sex Ratio in Assortative Matching". *American Economic Journal: Applied Economics* 3(3), pp. 124–157.

Absolon, Rudolf (1960). *Wehrgesetz und Wehrdienst 1935-1945. Das Personalwesen in der Wehrmacht.* Schriften des Bundesarchivs 5. Boldt.

Aguiar, Mark, Mark Bils, Kerwin Kofi Charles, and Erik Hurst (2017). "Leisure Luxuries and the Labor Supply of Young Men". *NBER Working Paper No. 23552.*

Ahlfeldt, Gabriel, Pantelis Koutroumpis, and Tommaso Valletti (2017). "Speed 2.0: Evaluating access to universal digital highways". *Journal of the European Economic Association* 15(3), pp. 586–625.

Aizer, Anna and Janet Currie (2014). "The intergenerational transmission of inequality: Maternal disadvantage and health at birth". *Science* 344(6186), pp. 856–861.

Almond, Douglas and Janet Currie (2011). "Killing Me Softly: The Fetal Origins Hypothesis". *Journal of Economic Perspectives* 25(3), pp. 153–72.

Almond, Douglas, Joseph J. Doyle Jr., Amanda E. Kowalski, and Heidi Williams (2010). "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns". *The Quarterly Journal of Economics* 125(2), pp. 591–634.

Almond, Douglas, Joseph J Doyle, Amanda E Kowalski, and Heidi Williams (2011). "The role of hospital heterogeneity in measuring marginal returns to medical care: A reply to Barreca, Guldi, Lindo, and Waddell". *The Quarterly Journal of Economics* 126(4), pp. 2125–2131.

Amaral-Garcia, Sofia, Mattia Nardotto, Carol Propper, and Tommaso Valletti (2017). "Internet and Health Choices: Evidence from the UK". *Unpublished.*

Atella, Vincenzo, Edoardo Di Porto, and Joanna Kopinska (2016). "Stress, Famine and The Fetal Programming: The Long Term Effect of WWII in Italy". *Tor Vergata University, CEIS Research Paper Series, Vol. 14, Issue 9, No. 385.*

AWMF (2007). "S2k-Leitline der Gesellschaft für Neonatologie und pädiatrische Intensivmedizin: Frühgeborene an der Grenze der Lebensfähigkeit, AWMF Leitlinien-Register Nr. 024-019".

AWMF (2012). "S2k-Leitlinine der Deutschen Gesellschaft für Kinderchirurgie: Kurzstreckige Ösophagusatresie, AWMF Leitlinien-Register Nr. 006-045".

AWMF (2014). "S3-Leitlinie der Deutschen Gesellschaft für Ernährungsmedizin: Parenterale Ernährung in der Kinder- und Jugendmedizin, AWMF Leitlinien-Register Nr. 073-023".

Barker, David J (1990). "The fetal and infant origins of adult disease". *British Medical Journal* 301(6761), p. 1111.

Barreca, Alan I, Melanie Guldi, Jason M Lindo, and Glen R Waddell (2011). "Saving Babies? Revisiting the effect of very low birth weight classification". *The Quarterly Journal of Economics* 126(4), pp. 2117–2123.

Barros, Pedro and Gisele Braun (2016). "Upcoding in a national health service: The evidence from Portugal". *Health Economics* 26(5), pp. 600–618.

Bauer, Karl et al. (2006). "Empfehlungen für die strukturellen Voraussetzungen der perinatologischen Versorgung in Deutschland". *Zeitschrift für Geburtshilfe und Neonatologie* 210(01), pp. 19–24.

Bauernschuster, Stefan, Oliver Falck, and Ludger Woessmann (2014). "Surfing alone? The internet and social capital: Evidence from an unforeseeable technological mistake". *Journal of Public Economics* 117(C), pp. 73–89.

Bayerisches Statistisches Landesamt, ed. (1937-1942). *Zeitschrift des Bayerischen Statistischen Landesamts.*

Behrman, Jere R and Mark R Rosenzweig (2004). "Returns to Birthweight". *Review of Economics and Statistics* 86(2), pp. 586–601.

Bellou, Andriana (2015). "The impact of Internet diffusion on marriage rates: evidence from the broadband market". *Journal of Population Economics* 28(2), pp. 265–297.

Bertschek, Irene, Wolfgang Briglauer, Kai Hüschelrath, Benedikt Kauf, and Thomas Niebel (2015). "The Economic Impacts of Broadband Internet: A Survey". *Review of Network Economics* 14(4), pp. 201–227.

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

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, and Magne Mogstad (2013). "Broadband Internet: An Information Superhighway to Sex Crime?" *The Review of Economic Studies* 80(4), pp. 1237–1266.

Bhuller, Manudeep, Tarjei Havnes, Jeremy McCauley, and Magne Mogstad (2017). "How the Internet Changed the Market for Print Media". *Unpublished.*

Billari, Francesco C, Osea Giuntella, and Luca Stella (2017a). "Broadband Internet, Digital Temptations, and Sleep". *Working Paper.*

Billari, Francesco C, Osea Giuntella, and Luca Stella (2017b). "Does Broadband Internet Affect Fertility?" *IZA Discussion Paper No. 10935.*

Boberach, Heinz (1984). *Meldungen aus dem Reich: die geheimen Lageberichte des Sicherheitsdienstes der SS: 1938-1945.* Pawlak Verlag.

Breining, Sanni, N. Meltem Daysal, Marianne Simonsen, and Mircea Trandafir (2015). "Spillover Effects of Early-Life Medical Interventions". *IZA Working Paper Nr. 9086.*

Briglauer, Wolfgang, Niklas Dürr, Oliver Falck, and Kai Hüschelrath (2016). "Does State Aid for Broadband Deployment in Rural Areas Close the Digital and Economic Divide?" *ZEW Discussion Paper No. 16-064.*

Bundesnetzagentur (2010). *Annual report 2009.*

Bundesnetzagentur (2017). *Annual report 2016.*

Campante, Filipe R, Ruben Durante, and Francesco Sobbrio (forthcoming). "Politics 2.0: The multifaceted effect of broadband internet on political participation". *Journal of the European Economic Association.*

Canzian, Giulia, Samuele Poy, and Simone Schüller (2015). "Broadband Diffusion and Firm Performance in Rural Areas: Quasi-Experimental Evidence". *FBK-IRVAPP Working Paper No. 2015-10.*

Carlson, Kyle (2015). "Fear itself: The effects of distressing economic news on birth outcomes". *Journal of Health Economics* 41, pp. 117–132.

Currie, Janet (2011). "Inequality at Birth: Some Causes and Consequences". *American Economic Review* 101(3), pp. 1–22.

Currie, Janet and Maya Rossin-Slater (2013). "Weathering the storm: Hurricanes and birth outcomes". *Journal of Health Economics* 32(3), pp. 487–503.

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

Czernich, Nina (2012). "Broadband Internet and Political Participation: Evidence for Germany". *Kyklos* 65(1), pp. 31–52.

Dafny, Leemore S (2005). "How do hospitals respond to price changes?" *The American Economic Review* 95(5), pp. 1525–1547.

Deaton, Angus (2013). *The great escape: health, wealth, and the origins of inequality.* Princeton University Press.

DellaVigna, Stefano and Eliana La Ferrara (2015). "Economic and Social Impacts of the Media". In: *Handbook on the Economics of the Media.* Ed. by Robert G Picard and Steven S Wildman.

Dettling, Lisa J (2017). "Broadband in the labor market: the impact of residential high-speed internet on married women's labor force participation". *ILR Review* 70(2), pp. 451–482.

DiNardi, Mike, Melanie Guldi, and David Simon (2017). "Body weight and internet access: Evidence from the rollout of broadband providers". *Working paper.*

Evans, Richard J. (2004). *Das dritte Reich.* Dt. Verl.-Anst.

Falck, Oliver (2017). "Does broadband infrastructure boost employment?" *IZA World of Labor: 341.*

Falck, Oliver, Robert Gold, and Stephan Heblich (2014a). "E-lections: Voting Behavior and the Internet". *American Economic Review* 104(7), pp. 2238–65.

Falck, Oliver, Robert Gold, and Stephan Heblich (2014b). "E-lections: Voting Behavior and the Internet: Appendix". *American Economic Review* 104(7), pp. 2238–65.

Falck, Oliver, Alexandra Heimisch, and Simon Wiederhold (2016). "Returns to ICT Skills". *CESifo Working Paper No.5720.*

Fallwell, Lynne (2013). *Modern german midwifery: 1885 - 1960.* Studies for the Society for the Social History of Medicine 13. Pickering & Chatto.

Floris, Joel, Kaspar Staub, and Ulrich Woitek (2016). "The Benefits of Intervention: Birth Weights in Basle 1912-1920". *University of Zurich, Department of Economics, Working Paper Series No. 236.*

Frei, Norbert (2013). *Der Führerstaat. Nationalsozialistische Herrschaft 1933 bis 1945.* Beck.

Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti (2017). “Internet and Politics: Evidence from UK Local Elections and Local Government Policies”. *Working Paper.*

Gebhardt, Georg (2010). “The impact of the Internet on retail competition: Evidence from technological differences in Internet access”. *SFB/TR 15 Discussion Paper No. 345.*

Gemeinsamer Bundesausschuss (2013). “Beschluss über eine Änderung der Vereinbarung über Maßnahmen zur Qualitätssicherung der Versorgung von Früh-und Neugeborenen vom 20 Juni 2013”.

Goebel, Jan and Bernd Pauer (2014). “Datenschutzkonzept zur Nutzung von SOEPgeo im Forschungsdatenzentrum SOEP am DIW Berlin”. *Zeitschrift für amtliche Statistik Berlin-Brandenburg* 3, pp. 42–47.

Grossman, Michael (1972). “On the Concept of Health Capital and the Demand for Health”. *Journal of Political Economy* 80(2), pp. 223–255.

Guldi, Melanie and Chris M. Herbst (2017). “Offline effects of online connecting: the impact of broadband diffusion on teen fertility decisions”. *Journal of Population Economics* 30(1), pp. 69–91.

Hack, Maureen, Daniel J Flannery, Mark Schluchter, Lydia Cartar, Elaine Borawski, and Nancy Klein (2002). “Outcomes in young adulthood for very-low-birth-weight infants”. *New England Journal of Medicine* 346(3), pp. 149–157.

Hansen, Bruce E (2001). “The new econometrics of structural change: Dating breaks in US labor productivity”. *The Journal of Economic Perspectives* 15(4), pp. 117–128.

Hjort, Jonas and Jonas Poulsen (2017). “The Arrival of Fast Internet and Employment in Africa”. *NBER Working Paper No. 23582.*

Hummer, Michael, Thomas Lehner, and Gerald Pruckner (2014). “Low birth weight and health expenditures from birth to late adolescence”. *The European Journal of Health Economics* 15(3), pp. 229–242.

Januleviciute, Jurgita, Jan Erik Askildsen, Oddvar Kaarboe, Luigi Siciliani, and Matt Sutton (2016). “How do Hospitals Respond to Price Changes? Evidence from Norway”. *Health Economics* 25(5), pp. 620–636.

Jürges, Hendrik (2013). “Collateral damage: The German food crisis, educational attainment and labor market outcomes of German post-war cohorts”. *Journal of Health Economics* 32(1), pp. 286–303.

Jürges, Hendrik and Juliane Köberlein (2015). “What explains DRG upcoding in neonatology? The roles of financial incentives and infant health”. *Journal of Health Economics* 43, pp. 13–26.

Kesternich, Iris, Bettina Siflinger, James Smith, and Joachim Winter (2014). “The Effects of World War II on Economic and Health Outcomes across Europe”. *The Review of Economics and Statistics* 96(1), pp. 103–118.

Kesternich, Iris, Bettina Siflinger, James Smith, and Joachim Winter (2015). “Individual Behavior as a Pathway between Early-Life Shocks and Adult Health: Evidence from Hunger Episodes in Post-War Germany”. *The Economic Journal* 125(588), F372–F393.

Kuntz, Benjamin and Thomas Lampert (2016). “Social Disparities in Maternal Smoking during Pregnancy”. *Geburtshilfe und Frauenheilkunde* 76(03), pp. 239–247.

Lange, Mirjam RJ (2017). “Tariff diversity and competition policy: Drivers for broadband adoption in the European Union”. *DICE Discussion Paper No 262.*

Lelkes, Yphtach, Gaurav Sood, and Shanto Iyengar (2017). “The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect”. *American Journal of Political Science* 61(1), pp. 5–20.

Lindeboom, Maarten and Reyn Van Ewijk (2015). “Babies of the War: The effect of war exposure early in life on mortality throughout life”. *Biodemography and Social Biology* 61(2), pp. 167–186.

Lumey, Lambert H, Aryeh D Stein, and Ezra Susser (2011). “Prenatal famine and adult health”. *Annual Review of Public Health* 32, pp. 237–262.

MacKie-Mason, Jeffrey K. and Hal Varian (1994). “Economic FAQs About the Internet”. *Journal of Economic Perspectives* 8(3), pp. 75–96.

Mang, Constantin (2016). “Market Consequences of ICT Innovations”. PhD thesis. ifo Institut.

McIntire, Donald D., Steven L. Bloom, Brian M. Casey, and Kenneth J. Leveno (1999). “Birth Weight in Relation to Morbidity and Mortality among Newborn Infants”. *New England Journal of Medicine* 340(16), pp. 1234–1238.

Monjok, Emmanuel, Ita B Okokon, Margaret M Opiah, Justin A Ingwu, John E Ekabua, and Ekere J Essien (2012). “Obstructed labour in resource-poor settings: the need for revival of symphysiotomy in Nigeria”. *African Journal of Reproductive Health* 16(3), pp. 93–100.

Mylonas, Ioannis and Klaus Friese (2015). “Indications for and Risks of Elective Cesarean Section”. *Deutsches Ärzteblatt International* 112(29-30), p. 489.

Nardotto, Mattia, Tommaso Valletti, and Frank Verboven (2015). “Unbundling the incumbent: Evidence from UK broadband”. *Journal of the European Economic Association* 13(2), pp. 330–362.

Nolte, André (2016). “Sex Crime, Murder, and Broadband Internet Expansion - Evidence for German Municipalities”. *Working paper.*

Overmanns, Rüdiger (2009). *Deutsche militärische Verluste im Zweiten Weltkrieg.* 3rd ed. Beiträge zur Militärgeschichte 46. Oldebourg.

Poy, Samuele and Simone Schüller (2016). “Internet and Voting in the Web 2.0 Era: Evidence from a Local Broadband Policy”. *CESifo Working Paper No.6129.*

Putnam, Robert D (2001). *Bowling alone: The collapse and revival of American community.* Simon and Schuster.

Quintana-Domeque, Climent and Pedro Ródenas-Serrano (forthcoming). “The Hidden Costs of Terrorism: The Effects on Health at Birth”. *Journal of Health Economics.*

Reichsminister des Inneren (1939). “Runderlass, betr. Hausentbindungen und Anstaltsentbindungen vom 6. September 1939”. *Reichsgesundheitsblatt* 14(63), p. 873.

Robert-Koch-Institut, ed. (2014). *Gesundheitliche Lage der Männer in Deutschland.* Beiträge zur Gesundheitsberichterstattung des Bundes. Robert Koch-Institut.

Robert-Koch-Institut, ed. (2015). *Gesundheit in Deutschland.* Gesundheitsberichterstattung des Bundes. Robert Koch-Institut.

Salm, Martin and Ansgar Wübker (2015). "Do Hospitals Respond to Increasing Prices by Supplying Fewer Services?" *Ruhr Economic Papers Nr. 567.*

Schiman, Jeffrey C., Robert Kaestner, and Anthony T. Lo Sasso (2017). "Early Childhood Health Shocks and Adult Wellbeing: Evidence from Wartime Britain". *NBER Working Paper No. 23763.*

Schreyögg, Jonas, Reinhard Busse, Matthias Bäuml, Jonas Krämer, Tilman Dette, and Alexander Geissler (2014). "Forschungsauftrag zur Mengenentwicklung nach § 17b Abs. 9 KHG - Endbericht". Hamburg Center for Health Economics.

Schumacher, Reto and Luigi Lorenzetti (2005). ""We Have No Proletariat": Social Stratification and Occupational Homogamy in Industrial Switzerland, Winterthur 1909/10–1928". *International Review of Social History* null (Supplement S13), pp. 65–91.

Sensch, Jürgen (2006). *histat-Datenkompilation online: Grunddaten zur historischen Entwicklung des Gesundheitswesens in Deutschland von 1876 bis 1999.* Data from: GESIS Datenarchiv, Köln. histat. Study number 8209.

Shigeoka, Hitoshi and Kiyohide Fushimi (2014). "Supplier-induced demand for newborn treatment: Evidence from Japan". *Journal of Health Economics* 35, pp. 162–178.

Silverman, Elaine and Jonathan Skinner (2004). "Medicare upcoding and hospital ownership". *Journal of Health Economics* 23(2), pp. 369–389.

Smith, James P. (1999). "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status". *Journal of Economic Perspectives* 13(2), pp. 145–166.

Stadtarchiv München, ed. (1939-1940). *Kriegswirtschaftsberichte.*

Statistisches Reichsamt, ed. (1933-1940). *Statistisches Jahrbuch für das Deutsche Reich.*

Stauber, M. (2012). "Herausforderungen – 100 Jahre Bayerische Gesellschaft für Geburtshilfe und Frauenheilkunde". In: ed. by C. Anthuber, M.W. Beckmann, J. Dietl, F. Dross, and W. Frobenius. Thieme. Chap. Vergangenheitsbewältigung in der bayerischen Gynäkologie – Erfahrungen an der I. Universitätsfrauenklinik München, pp. 237–256.

Stein, Aryeh D, Patricia A Zybert, Margot Van de Bor, and LH Lumey (2004). "Intrauterine famine exposure and body proportions at birth: the Dutch Hunger Winter". *International Journal of Epidemiology* 33(4), pp. 831–836.

Süß, Winfried (2003). *Der "Volkskörper" im Krieg. Gesundheitspolitik, Gesundheitsverhältnisse und Krankenmord im nationalsozialistischen Deutschland 1939-1945.* Oldenbourg.

Torche, Florencia (2011). "The Effect of Maternal Stress on Birth Outcomes: Exploiting a Natural Experiment". *Demography* 48(4), pp. 1473–1491.

Usborne, Cornelie (2011). "Social Body, Racial Body, Woman's Body. Discourses, Policies, Practices from Wilhelmine to Nazi Germany, 1912-1945". *Historical Social Research* 36(2), pp. 140–161.

Uysal, S. Derya (2015). "Doubly Robust Estimation of Causal Effects with Multivalued Treatments: An Application to the Returns to Schooling". *Journal of Applied Econometrics* 30(5), pp. 763–786.

Van den Berg, Gerard J., Pia R. Pinger, and Johannes Schoch (2016). "Instrumental Variable Estimation of the Causal Effect of Hunger Early in Life on Health Later in Life". *Economic Journal* 126(591), pp. 65–506.

Van Ewijk, Reyn and Maarten Lindeboom (2016). "Why People Born During World War II are Healthier". *Gutenberg School of Management and Economics & Research Unit "Interdisciplinary Public Policy", Discussion Paper No. 1619.*

Van Leeuwen, Marco HD and Ineke Maas (2011). *HISCLASS: A historical international social class scheme.* Leuven University Press.

Van Leeuwen, Marco HD, Ineke Maas, and Andrew Miles (2002). *HISCO: Historical international standard classification of occupations.* Leuven University Press.

Verzulli, Rossella, Gianluca Fiorentini, Matteo Lippi Bruni, and Cristina Ugolini (2016). "Price Changes in Regulated Healthcare Markets: Do Public Hospitals Respond and How?" *Health Economics.*

Wagner, Gert, Joachim Frick, and Jürgen Schupp (2007). "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements". *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften* 127(1), pp. 139–169.

World Health Organization, ed. (2000). *Obesity: preventing and managing the global epidemic.* World Health Organization.

World Health Organization, ed. (2003). *Managing Complications in Pregnancy and Childbirth: A guide for midwives and doctors.* World Health Organization.

World Health Organization, ed. (2006). *Neonatal and perinatal mortality: country, regional and global estimates.* World Health Organization.

Zander, Josef and Elisabeth Goetz (1986). "Hausgeburten und klinische Entbindung im Dritten Reich (Über eine Denkschrift der Deutschen Gesellschaft für Gynäkologie aus dem Jahre 1939)". In: *Zur Geschichte der Gynäkologie und Geburtshilfe: Aus Anlaß des 100jährigen Bestehens der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe.* Ed. by Lutwin Beck. Springer.