Physicians and the production of health: Returns to health care during the mortality transition

Helge Liebert*, ‡ Beatrice Mäder†, ‡

September 25, 2020

Abstract

This paper investigates the returns to health care provision during the mortality transition. We construct a new panel data set covering German municipalities from 1928 to 1936. The endogeneity of health care supply is addressed by using the expulsion of Jewish physicians from health insurance schemes as exogenous variation in regional physician supply ratios. Increases in the supply of physicians substantially reduce infant mortality, stillbirths and common childhood diseases. Mortality effects are larger where hospital infrastructure is absent or capacity is limited, suggesting human capital and infrastructure are complementary inputs for health production. Our results emphasize diminishing returns to health care provision. Using a semiparametric control function approach, we find that the marginal returns to physicians are highly nonlinear and decreasing. The estimates are consistent with historical trends in infant mortality and health care supply over the 20th century.

Keywords: infant mortality, physicians, health care supply, mortality transition, childhood diseases, diminishing returns, semiparametric IV

JEL classification: I10, I18, N34

*Corresponding author. Department of Economics, University of Zürich. E-mail: helge.liebert@uzh.ch

†Department of Economics, University of St. Gallen and Polynomics. E-mail: beatrice.maeder@unisg.ch.

‡We thank Simone Balestra, Amitabh Chandra, Beatrix Eugster, Claudia Goldin, Roland Hodler, Rafael Lalive, Michael Lechner, Nicole Maestas, Dina Pomeranz, Anthony Strittmatter, Nico Voigtländer, Hans-Joachim Voth, Joachim Winter and Nicolas Ziebarth for helpful comments. The paper has also benefited from comments by seminar participants at the Department of Health Care Policy at Harvard, University of St. Gallen, University of Luzern, University of Göttingen, the European Workshop on Econometrics and Health Economics, the annual meeting of the Verein für Socialpolitik and the Swiss Society of Economics and Statistics. The paper was awarded the Young Economist Award of the Swiss Society of Economics and Statistics 2018.

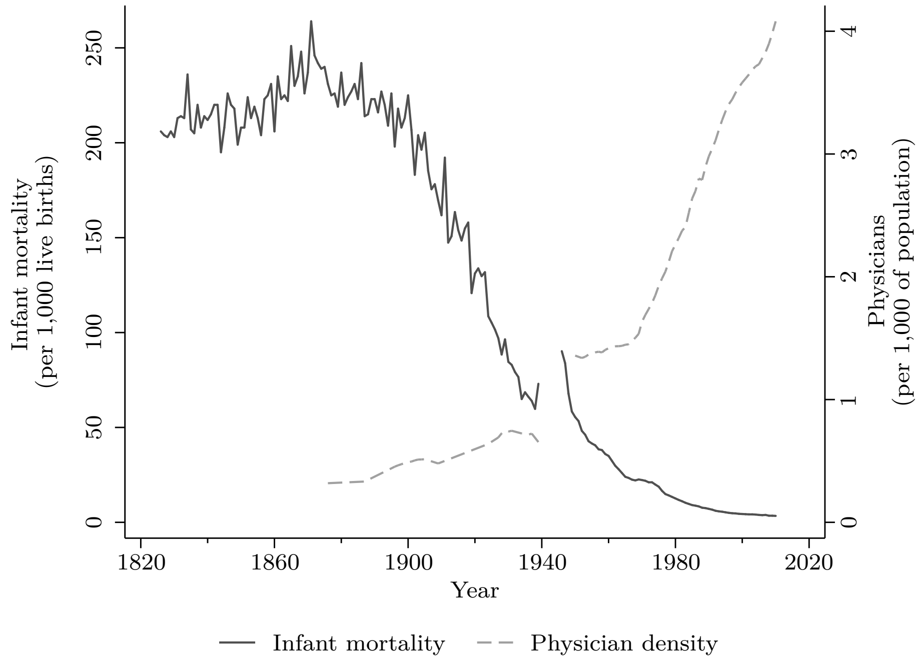
1 Introduction

The reduction in infant mortality among industrialized countries since the mid-19th century constitutes an unprecedented development in human history. Before, infant mortality rates in Europe had essentially been unchanged since the middle ages (Dyer 1989, Shahar 1990, Mitchell 2013, Santow 2001). In 1900, 225 children out of 1,000 born died within their first year of life in Germany. In 1950, it was less than 60 and in 2015, the rate was down to 3.1, a decrease of more than 98% over a century (cf. Figure 1). This mortality transition can be observed across all of western Europe and the United States (Santow 2001, Caselli 2015). The accompanying improvements in life expectancy have been shown to be important for long-run economic growth and development (Acemoglu and Johnson 2007, Bloom et al. 2014). In contrast to this, infant mortality rates still remain high in many developing economies. Of 4.5 million infant deaths worldwide in 2015, 99% occurred in developing countries (You et al. 2015).

The mortality decline in developed economies coincides with important public health developments. Standards of living, nutrition and public hygiene started improving at the end of the 19th century; health care supply and public health infrastructure were expanded substantially (Preston 1980, Loudon 1991, Cutler et al. 2006, Frohman and Brook 2006, Alsan and Goldin 2018). Between 1900 and 1950, the ratio of physicans per population increased more than twofold (cf. Figure 1). Physicians provide pre- and postnatal care, attend birth, administer medication and encourage health-related behavior and compliance with hygienic standards (cf. Stanton and Clemens 1987, Fewtrell et al. 2005, Terza et al. 2008).

In this paper, we analyze the role of physicians for the production of health in the early 20th century, towards the end of the mortality transition. We establish a causal link between health care supply and health by tracing exogenous shocks to the supply of physicians. We find that improvements in health care supply can substantially reduce infant mortality, stillbirths and mortality from common childhood diseases. Moreover, we investigate how health care supply and infrastructure interact as joint inputs for the production of health. Our findings indicate that physical hospital infrastructure and

Figure 1: Infant mortality in Germany, 1826–2010



Notes: The graphs shows the historical development of infant mortality and physician density in German territories between 1826–2014. Infant mortality is measured as the number of children dying within the first year of life per 1,000 live births. Sources: Data is collected from Gehrmann (2012), Statistisches Reichsamt (1884–1940) and Statistisches Bundesamt (1944-2015) [Federal statistical office].

physician human capital are complements. Physicians matter especially in those regions where access to infrastructure is limited.

To pinpoint where improving coverage is most effective, we provide non-parametric estimates of the marginal returns to physicians. In doing so, we evaluate the premise of diminishing returns to health care provision, featured commonly in models of health care demand following Grossman (1972). Improving medical care may be especially effective in regions where coverage is sparse; whereas increasing supply in regions where levels are already high may increase non-vital health care consumption, but do little to improve vital health outcomes. For this purpose, we develop a semiparametric instrumental variables (IV) estimation approach. We combine a control function approach with a partially linear model in the spirit of Robinson (1988) and Baltagi and Li (2002) to derive a non-parametric estimate of the dose-response function. We find that health effects are broadly positive, but subject to rapidly diminishing returns—positive effects are limited to a narrow support region of total supply. Physicians have large positive effects on health especially when baseline coverage is sparse.

Since the supply of physicians is endogenous to health care demand, a simple comparison of health outcomes between regions will lead to biased estimates. To identify the causal effect of physicians on mortality, we utilize a series of discriminatory policies introduced in Germany in 1933 which banned Jews and those with left-wing political affiliations from public positions and severely limited the professional activity of Jewish physicians. Within a system of universal public health insurance, laws were introduced which excluded Jewish and other physicians from being reimbursed for rendering medical services, leading many physicians to emigrate to other countries (cf. Figure 1). Jews were disproportionately over-represented in the medical profession, with up to 17% of physicians considered Jewish by the Nazi definition in 1933 (Kröner 1989), compared to less than 0.7% in the overall population. The distribution of the Jewish population varies across German municipalities, providing variation in treatment exposure. The variation induced by the occupational restrictions allows us to solve the problem of positive selection.

Our analysis relies on a large sample of administrative data covering the period from 1928 to 1936. Detailed information on causes of death and disease incidence allows us to examine specifically which medical conditions physicians can influence and also how this effect interacts with health care infrastructure. We carefully show that health effects are not driven by other confounding factors or policies at the time. We also show that due to the disparity in population size, it is impossible that the mortality effects we document are driven by higher mortality among the Jewish population alone.

Our findings suggest that one additional doctor per 1,000 of population reduces infant mortality by about 18 cases per 1,000 live births. This corresponds to about a 23% reduction in baseline mortality. While this effect is sizeable, an additional physician is also a large increase in coverage—in our sample, increasing the coverage ration by one additional doctor is approximately equivalent to doubling the supply of physicians. A one standard deviation increase in the number of physicians translates to 0.45 standard deviation decrease in infant mortality.

In addition, our results show that reductions occur for deaths from inflammatory bowel diseases and stillbirths, which are among the main causes for mortality in developing countries today. We also find mortality reductions for viral diseases like measles, influenza and bronchitis. Fatalities due to premature birth and congenital issues, for which medical treatment is difficult, are unaffected. Mortality from these health conditions accounts almost completely for the remaining infant mortality in developed countries today.

Regarding effect heterogeneity, we find that mortality effects are larger in municipalities with limited hospital capacity or without specialized hospital infrastructure. Using the semiparametric IV estimation method, we demonstrate that mortality effects are highly nonlinear and disappear after a coverage ratio of about two physicians per 1,000 of population. Our results align with the historical development of infant mortality in Europe and the US over the 20th century—infant mortality rates have been largely stable since the 1980s, when most countries reached comparable coverage ratios.

Our work in this paper contributes to several distinct branches of literature in economic history, health economics and development. First, our results add to the literature on the value of health insurance and large social programs in general. Several recent studies have evaluated the introduction of health insurance schemes and found broad health and mortality effects (e.g. Card et al. 2008, Finkelstein and McKnight 2008). Evaluating the introduction of Medicaid in the 1960s, Goodman-Bacon (2018) finds large reductions in infant and child mortality for at-risk populations. While these studies focus on the effects of health care provision at the *extensive* margin, we study the effects of providing additional health care resources at the *intensive* margin within a system of universal health insurance. Our results suggest that the effect of additional supply depends crucially on the existing level of provision.

Second, our results add to the broad literature on the effects of health care services and infrastructure on health. Many papers have documented broadly positive effects of the availability of health care infrastructure on child health (e.g. Lavy et al. 1996, Frankenberg 1995). Similar positive effects have been documented for increases in supply-side financing (e.g Gruber et al. 2014) and improved public infrastructure due to understanding the germ theory of disease (Cutler et al. 2006, Preston 1975). We contribute by studying how the availability of hospital infrastructure interacts with the provision of health care supply at the margin. Our results indicate that human capital and infrastructure are complements and only partially substitutable.

Third, we contribute directly to the literature on the effects of physicians on infant mortality and health in general. Although previous work has established a connection between physician supply and health, causal evidence on how the supply of physicians affects health outcomes is limited. Most studies rely on cross-country comparisons, often focusing on single cross-sections (e.g. Kim and Moody 1992, Hertz et al. 1994, Anand and Bärnighausen 2004). Some papers use a panel data approach to control for the influence of time-invariant unobservable factors (e.g. Farahani et al. 2009, Bhargava et al. 2011, Chauvet et al. 2013). The results from these studies vary. Some find a negative relationship between physicians and mortality, but many results are inconclusive. Micro data evidence is limited.Aakvik and Holmås (2006) use a dynamic panel approach to estimate the effect of general practitioner coverage on total mortality in Norwegian municipalities from 1986 to 2001, but do not find a significant relationship.

A limiting factor of these studies is their inability to account for time-varying endogenous changes in the supply of physicians. Moreover, sample sizes are often small and the data quality questionable (see Hill et al. 2007, Farahani et al. 2009). Most low- and middle-income countries do not have well-functioning vital registration systems and estimates of the infant mortality rate often rely on a mixture of sources and reporting systems. Our paper addresses both of these issues. Using exogenous variation in the supply of physicians due to discriminatory measures allows us to obtain a causal estimate of the effect of an additional physician on infant mortality. Furthermore, population statistics and vital registration systems were already well established in Germany in the early 20th century, allowing us to rely on a comparatively large administrative dataset of German municipalities.

Fourth, we provide empirical evidence on the shape of the health production function. Diminishing marginal returns are a key feature of theoretical models (Grossman 1972), but real world empirical evidence is limited. Our estimated dose-response curve can be considered a partial estimate of the health care production function (Reinhardt 1972, Thurston and Libby 2002). Health care provision is shown to be especially effective at low levels of baseline provision, but subject to rapidly diminishing marginal returns. The nonparametric results also provide an intuitive explanation for many of the previous null results in the literature. Most developed countries reached levels of physician supply for which we find no effect any more in the early 1980s, and infant mortality rates have been mostly flat since (cf. Figure A3). Remaining infant mortality today is largely composed of critical congenital conditions, for which we find no effect at any margin of supply provision.

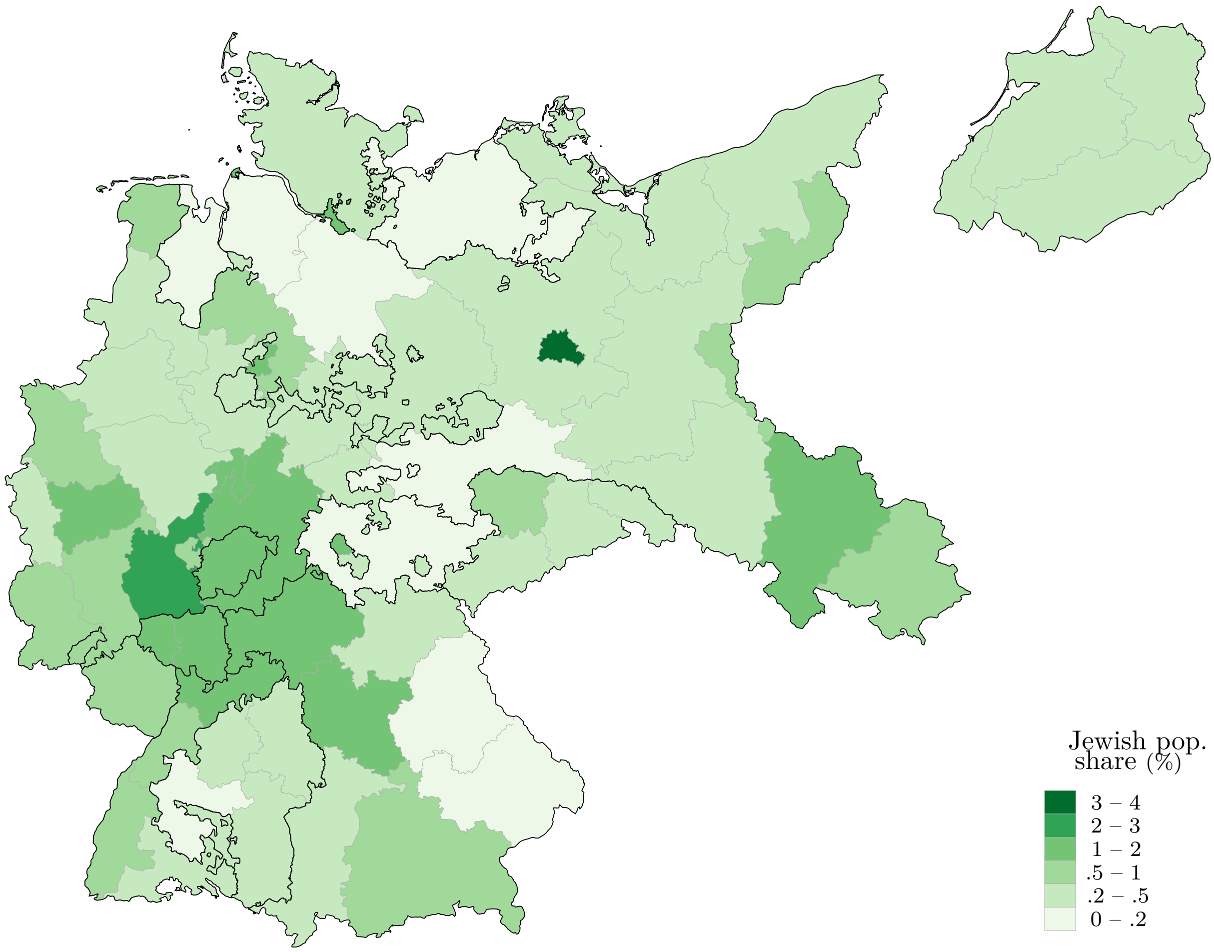
Finally, our results add to the literature on negative human capital shocks and the consequences of large-scale discriminatory policies. Akbulut-Yuksel and Yuksel (2015) and Waldinger (2010, 2012) document negative effects of the Jewish persecution on educational attainment and research productivity. Waldinger (2016) documents that the expulsion of Jewish scientists and the resulting human capital loss substantially reduced long-run output. We document the severe consequences of physician brain drain and human capital loss in the health domain. Between 1933 and 1936, the occupational policies in Germany reduced the number of practicing physicians by up to one fifth, and the supply shortages resulting from discrimination are linked to substantial increases in mortality.

The remainder of the paper is structured as follows: section 2 provides the institutional background for our analysis, section 3 reviews the data, section 4 discusses our empirical strategy and develops the estimation methods, section 5 presents our results and sensitivity checks and section 6 concludes.

2 Institutional background

The 1933 German census registered 499,682 residents of Jewish faith. Overall, this constituted only a small share of the then total German population of about 65 Million, 0.77%. Since the census registered only citizens who confessed to the Jewish faith, this measure constitutes a lower bound for the share of the population considered Jewish by the extended definition introduced by the National Socialist Party. The Jewish population was distributed across all German regions, but concentrated in urban regions with over

Figure 2: Jewish population share in German regions in 1933



Notes: The graph shows the Jewish population share in 1933 across German sub-state administrative districts (*Regierungsbezirke*). District borders are shaded grey. State borders (*Länder*) are colored in black.

70% of Jews residing in larger cities compared to 30% of the entire population.

Figure 2 shows the spatial distribution of the Jewish population across German districts. The relatively smallest Jewish communities are located in rural areas in the north and the south. The largest Jewish communities lived in Frankfurt am Main (4.71%), Berlin (3.78%) and Breslau (3.23%). However, even though Jews constituted only a small part of the total population, they were disproportionately overrepresented in skilled occupations, especially in the medical profession. 10.9% of all doctors were of Jewish faith and approximately 17% were considered Jewish by the Nazis. In some municipalities, these figures where considerably higher. In Berlin, around 52% of all doctors were considered non-Aryan in 1933 (Kröner 1989).

Shortly after seizing power in 1933, the Nazi government introduced discriminatory policies that limited the professional activity of Jewish citizens. On April 7, 1933, the NSDAP passed the *Gesetz zur Wiederherstellung des Berufsbeamtentums* [Law for the restoration of the professional civil services]. The law decreed that civil servants of non-Aryan descent or with dubious past political activity could not be trusted to be loyal to the state and were to be placed in retirement effective immediately.[1] An implementation decree further specified that it was sufficient to have one Jewish grandparent to be considered of non-Aryan descent. Exceptions were granted to those who had been employed since before August 1, 1914 or who had served in the First World War. However, this privilege could be declared exempt if a person was judged to be politically unreliable, an exception that was invoked frequently. This meant that non-Aryan doctors were forced into retirement at universities, publicly funded hospitals and all public medical institutions. By May 6, 1933 two executive orders extended the law to ordinary state employees, forcing resident physicians out of work as well.

On April 22, the law was followed by the *Verordnung über die Zulassung von Ärzten zur Tätigkeit bei den Krankenkassen* [Decree on the accreditation of doctors for health insurance funds], which withdrew the licence to practice for the compulsory health insurance fund from Jewish and other non-Aryan physicians. The same exceptions as above applied, but could again be declared void if the physician could be shown to have been active as a communist. The term communist was interpreted broadly and regularly included social democrats (Kröner 1989, Leibfried and Tennstedt 1980). Shortly after, an agreement between the association of German doctors and the association of private health insurance providers from July 1933 declared that bills from non-Aryan doctors were only reimbursed if they were subject to the exception clause or if the patient himself was of non-Aryan descent (Böhle 2003). This made it impossible for patients to be reimbursed for medical expenses when visiting Jewish physicians, and effectively deprived Jewish doctors from the possibility of treating privately insured patients as well.

[1] The repercussions of the discriminatory laws and the subsequent dismissal of Jewish scientists from German universities have been investigated in a series of papers by Fabian Waldinger, who examines the consequences for educational outcomes (Waldinger 2010), academic productivity and peer effects (Waldinger 2012), the role of human and physical capital for the creation of scientific knowledge (Waldinger 2016) and the link between German Jewish Emigrés and US inventions (Moser et al. 2014).

Finally, a directive by the *Reichsärztekommissar* [Federal commissioner of physicians] from August 1933 specified that doctors of Aryan and non-Aryan descent were no longer allowed to stand-in for one another nor to refer patients to each other or to consult (Beddies et al. 2014). This also applied to non-Aryan physicians who profited from the exception clause and was especially harmful for specialists, who depended on referrals. According to the *Reichsvertretung der Juden in Deutschland* [Representation of German Jews], the directive rendered it virtually impossible for non-Aryan physicians to work in private hospitals and made the license for compulsory health insurance practice often worthless for those who still had one (Kröner 1989).

Germany has had universal public health insurance since 1883. With few exceptions, people are enrolled in a statutory health insurance plan, which provides a standardized level of coverage through one of the public or semi-private occupational insurance funds. Physicians treat patients on a fee-for-service basis and are reimbursed by the health insurance funds. At the time of our study, health insurance was compulsory for all employees with an income of less than 3,600 Reichsmark. This meant that only the very affluent could opt out as average income in 1927 was 1,742 Reichsmark (Klingenberger 2001). Treatment for cash payments was rare, unusual and restricted to a small minority of wealthy people. By barring Jewish physicians and those with left-wing political associations from submitting bills for reimbursement to the public insurance funds, the laws deprived physicians of their primary source of income.

The only remaining options for doctors whose health insurance licenses were revoked were to emigrate or to go into private practice. However, Klingenberger (2001) estimates that in 1930 there were only around 600,000 private patients in Germany, compared to a total population of around 65 million. This suggests that private practice could only provide a means of existence for a limited number of physicians and would be associated with a severe reduction in income. Considering their future in Germany, many Jewish physicians opted to emigrate (Kröner 1989). Likewise, physicians who faced oppression for their political views or were uneasy with the political development also left. The Jewish physicians who remained in Germany finally lost their medical licence in September 1938.

The effects of migration can be seen in Figure 1, where the historically increasing trend in the number of physicians breaks in the 1930s and the supply of physicians decreases.

Importantly for the purposes of our study, the occupational laws we consider did not apply to other health personal like nurses, but only affected physicians. Only in September 1938, the *Gesetz zur Ordnung der Krankenpflege* [Law on the Regulation of Nursing Care] and accompanying ordinances decreed that Jewish nurses where solely allowed to treat Jewish patients or work in Jewish hospitals. These regulations stayed in place until 1945, meaning that nursing was one of the few professions formally open to Jewish women in National Socialism (Steppe 1997, 2013). Before 1938, occupational restrictions in the health care profession were limited to physicians. Moreover, there is no evidence that the share of Jewish practicitioners was as skewed among nurses as it was among physicians, limiting any potential impact.

In light of the escalating discrimination and persecution, we limit our study to the years 1936 and prior. In 1938, when Jewish physicians' medical licences were revoked, preparations for war were already under way. A year later, on September 1, 1939, Germany invaded Poland, initiating World War II and the Holocaust. Ghettos where set up to segregate Jews from the rest of the population. In the following years, thousands of detention sites were established all over German-occupied Europe. Deportation to forced labor and specialized extermination camps, mass shootings and pogroms commenced in 1941. By the end of the war in May 1945, six million European Jews had been systematically murdered.

3 Data

The dataset is assembled from a variety of historical sources. The population and mortality figures are taken from the *Reichsgesundheitsblatt* [Health bulletin] (eds. 1927–1936), a yearly statistical publication by the health ministry of the German Reich. The data covers all municipalities with a population above 15,000. It tracks yearly changes in the municipal population. Available information includes total population figures, the number of births, marriages, and deaths. Mortality information is available in detailed categories in accordance with the International List of Causes of Death (ILCD-4, approximately corresponding to the International Classification of Diseases ICD-10 in use since 1990). The disease categories were expanded and partially redefined in 1930, restricting the coverage period for some variables. Panel attrition is less than 3% since the population threshold of 15,000 was not adhered to strictly. We restrict the sample to those municipalities with full coverage for a balanced panel (this choice is inconsequential for the results). Around 7% of the municipalities in our sample merge or split during the observation period. The affected municipalities have been harmonized to the municipal structure in place in 1933. For reasons outlined in the previous section, we limit our sample to 1936 and before and do not consider later data.

The main analysis focuses on infant mortality, i.e. the number of children dying within the first year after birth (excluding stillbirths) and the number of stillbirths. We also look at a few specific causes of death for infants that are listed in the data, inflammatory bowel diseases and premature birth or congenital defects. Additional outcomes are the population disease incidence figures for a number of common childhood diseases. While mortality is an extreme health outcome, it is historically well-measured and an unambiguous indicator of poor health that is not subject to varying measurement standards. The analysis uses incidence rates to track mortality effects. Infant mortality is measured as deaths in year t per 1,000 live births in $t-1$, stillbirths are recorded per 1,000 births in $t-1$ and disease mortality incidence per 1,000 of population in $t-1$. Similarly, the physician coverage ratio is expressed as physicians per 1,000 of population in $t-1$. These expressions are commonly used in epidemiology and allow comparisons with other work. Relative incidence is the more interesting target parameter in this context and also helps to abstract from trends in birth rates. We use lagged values for the reference population instead of current ones to contain endogenous feedback effects.

In addition, the unadjusted mortality figure distributions feature overdispersion and a long right tail typical for count data. Rescaling reduces skewness and alleviates these issues to some degree, improving the viability of standard linear models. Figure A1 compares the original and the scaled distributions for selected variables and illustrates that in many cases, anchoring mortality incidence to a reference population leads to a distribution that is approximately normal. For some variables, especially those with a comparatively large share of zero observations, the scaled distribution remains right-skewed. We address this issue in section 5.6.

Throughout the paper, we mostly work with rates for dependent and independent variables. For some specifications, we also use log-transformed counts (or the inverse hyperbolic sine transformation, e.g. Burbidge et al. 1988, MacKinnon and Magee 1990, Bellemare and Wichman 2020). For the regression analysis, we prefer the rates over logged counts for interpretability and to avoid both scaling issues and the log-of-zero problem, as some of the outcomes and exposure measures are sparse and have a higher rate of zeros in the data. We demonstrate in the sensitivity analysis that this choice is inconsequential for our results.

Information about the number of physicians is taken from the *Reichs-Medizinal-Kalender/Verzeichnis der deutschen Ärzte und Heilanstalten* [Register of German physicians and hospitals], a yearly publication listing physicians in Germany for each municipality. The register includes the full name, address, specialization and the year the medical licence was acquired. The publication also tracks physicians who cease practising and émigrés between years. We observe yearly emigration counts of physicians for every municipality. Prior to 1933 we observe only 145 physicians who emigrate. From 1933 to 1936 this number increases to over 2700 physicians. We use these emigration statistics as our main measure of policy exposure. Note that we do not digitize all individual-level data, only municipality-level counts. One advantage of the emigration measure is that it also includes other physicians (e.g. those affiliated with social-democratic or socialist/communist organizations) which left the country due to the occupational restrictions.

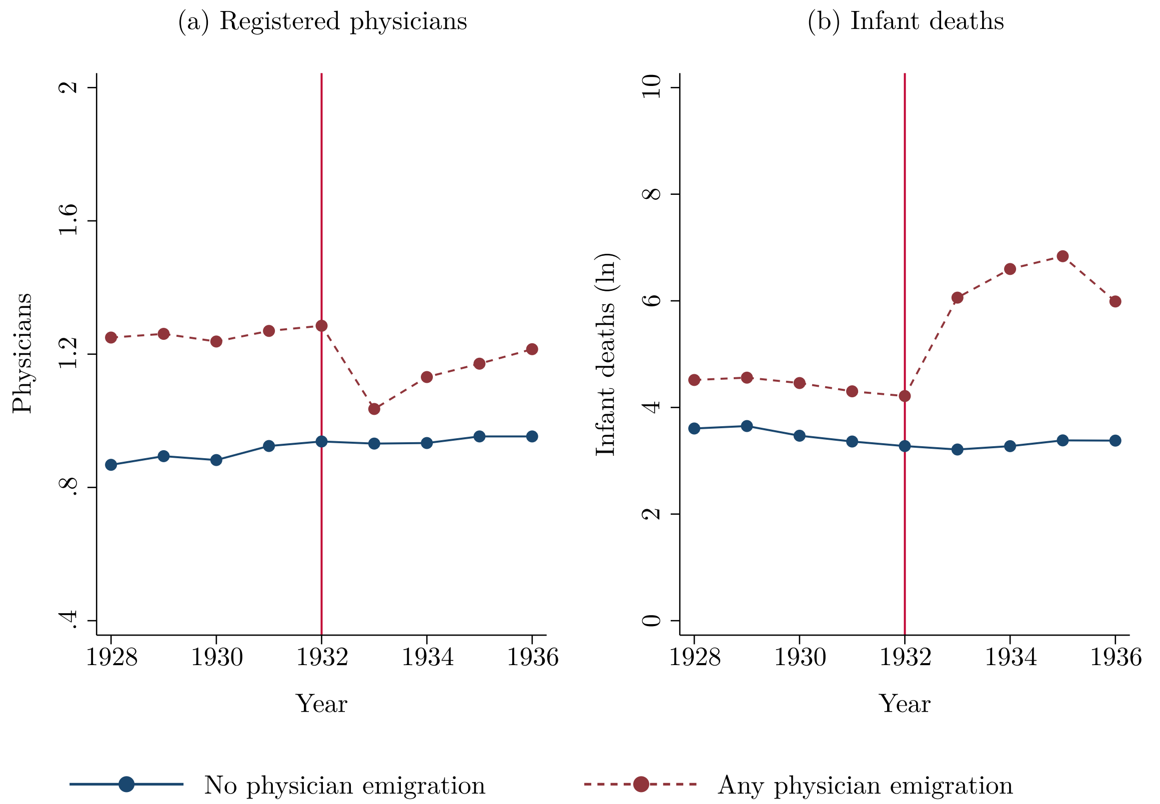
In addition to emigration, we construct two alternative measures of policy exposure in 1933. Editions of the Reichs-Medizinal-Kalender after 1936 explicitly tag Jewish physicians (considered Jewish by the extended definition). From this, we construct a proxy measure of Jewish physicians per municipality. Furthermore, we know the exact amount of people of Jewish faith in each municipality from the official population census conducted in 1933, just when the discriminatory measures where implemented. We use both the number of Jewish physicians remaining after 1936 and the local Jewish population in 1933 to proxy for the number of Jewish physicians expelled in 1933. Note that even if these variables only proxy for policy exposure with measurement error, this is inconsequential as long as they correctly represent exposure in the cross-sectional dimension. We discuss this point in more detail in the next section.

Although it would be very interesting to study different physician specializations, a more detailed analysis is precluded by the fact that medical practice was much less developed compared to today. In the early 20th century, most physicians trained as general practitioners and did not specialize in other domains. From aggregate data, we know that in 1933, 68% of all physicians in Germany were general practitioners. Of the 30% who did specialize, most were trained in internal medicine or surgery. Overall, only 2.6% of physicians were trained in pediatrics, and 3.5% in gynaecology. These physicians mostly worked in larger cities with university-affiliated hospitals. Births were generally attended by general practitioners/family physicians, not specialists. Our setting predates the pronounced specialization of modern medicine, precluding a more detailed analysis.

Instead, we focus on investigating the role of hospital infrastructure. To measure hospital infrastructure provision, we extract information about the prevalence and type of hospitals in each municipality from the 1933 edition of the Reichs-Medizinal-Kalender. Moreover, we also measure the capacity of hospitals by collecting information on the number of beds available at each facility. We further enrich the data with municipality-specific information from the population census, official labor statistics and election results.

The final sample is a balanced panel comprising 2,853 observations in 317 municipalities between 1928–1936. Our dataset covers 29,168,080 people in 1933, about 45% of the total German population at the time. By construction, the sample is selected on larger, more populous and urban municipalities. Descriptive statistics are given in Table A1 in the appendix. Although persons of Jewish faith number only 0.6% of the population, Jewish physicians account for about 8% of all physicians in our sample. This differs slightly from

Figure 3: Trends in physician coverage and infant mortality



Note: The graphs shows the development of the number of registered physicians per 1,000 of population and log infant deaths over time. Plotted are yearly mean values. Municipalities are partitioned into groups according to whether there is any outmigration of physicians. To account for size, yearly means for the group with positive migration are weighted by migrant frequency.

the 10% reported to be of Jewish faith in official aggregate statistics due to the timing of the measurement and the difference in the definition of who is considered Jewish. For every 1,000 live births, about 75 children die before reaching one year of age. The physician coverage ratio is approximately 1 per 1,000 of population.

To illustrate the dynamics in the data, Figure 3 depicts unconditional trends in the municipal physician coverage ratio per 1,000 of population and infant mortality over time, partitioned by groups of municipalities with and without any outmigration of physicians. The graph shows a distinct drop in the number of physicians in 1933 in municipalities with a higher share of Jewish residents, and a corresponding rise in infant mortality. Note that this type of analysis does not fully reflect our empirical approach. We only rely on partitioning into groups for illustration. Since municipalities are affected to varying degrees in our setting, we prefer accounting for exposure directly in the analysis.

Although we make use of all three exposure measures, we consider emigration to be the best measure of treatment exposure since it is time-varying. As such it measures treatment exposure more fine-grained and not only at a single point in time. From the graph, it appears that the number of physicians partially recovers after 1934. Although the discriminatory policies cause a decline in the number of licensed physicians, vacant spots can be filled again. While mobility of established practices is likely to be limited; younger physicians who finished their residency may open up practice in municipalities where positions are left vacant. This recovery is not accounted for when using time-invariant measures of policy exposure. However, new residents would have opened up practice eventually anyways—if they are more likely to do so in municipalities with vacant spots, this will only bias estimates downwards.

4 Empirical strategy

In the first part of this section, we discuss the main modeling approach and estimation, compare the merits of different exposure measures and address potential concerns regarding measurement error. In the next part, we discuss the explicit identification assumptions and potential threats to their validity. In the last part, we develop a more flexible estimation approach that imposes fewer functional form restrictions and permits isolating the dose-response function.

4.1 Identification and linear model

Identifying the effect of physician coverage on health outcomes is hampered by a series of selection problems. Most importantly, health care supply is endogenous to health care demand. Physicians tend to locate in places where their services are in demand, leading to a positive association between sickness prevalence and health care services. The endogenous locational choice of practice has long been recognized. The geographic distribution of physicians has and continues to receive a lot of attention, from economics, medical science and policy makers alike (e.g. Cooper et al. 1977, Newhouse et al. 1982, Fruen and Cantwell 1982). Positive selection is strengthened by the fact that physicians can potentially influence the demand for their services (e.g. Arrow 1963, McGuire 2000).

In addition, positive selection also occurs in more government-regulated health care markets. Traditionally, many continental European governments with public health insurance have regulated licensing to allocate medical services where they are deemed to be needed. In our context, the market for physicians can be considered one of managed competition. These features of health care markets make it notoriously difficult to identify the degree to which changes in health care coverage influence the health of the population.

We are using exogenous shocks to the supply of physicians to identify the causal effect of physician coverage on population mortality incidence. The 1933 law causes a drop in the number of registered physicians (cf. Figure 3). The size of the impact can be measured using the different policy exposure variables outlined in the previous section. We discuss the advantages and disadvantages of the different measures in the context of our analysis in more detail below.

As a first step, we conduct a reduced-form event study analysis. This analysis helps us ascertain the direct effect of the policy change on infant mortality, and also allows us to assess the importance of pre-treatment changes which would invalidate the interpretation of our main estimates. We estimate the following equation:

$$y_{it} = \sum_{\substack{j \neq \tilde{t} \\ j=\underline{t}}}^{\bar{t}} \beta_j z_i + \eta_i + \delta_t + \epsilon_{it} \ , \tag{1}$$

where $\left[\underline{t}, \bar{t}\right] = \left[1929, 1936\right]$ are the period limits, z_i is a municipality-level scaler measuring exposure to the policy and η_i and δ_t are municipality and year fixed effects. We use the different exposure measures outlined in the previous section. We choose the year $\tilde{t} = 1932$, one year before the policy was implemented, as the reference period. Our results replicate when choosing earlier years as reference periods, or a broader combination of reference periods.

However, this reduced-form approach does not isolate the effect of health care provision, i.e. the effect of an additional physician on health. For the remainder of the empirical analysis, we are using an instrumental variables approach to trace the exposure to variations in health care supply.[2] Consider a regression equation of mortality incidence y on physician supply s in municipality $i = 1, \ldots, N$ in year $t = 1, \ldots, T$,

$$y_{it} = \beta s_{it} + \eta_i + \delta_t + \epsilon_{it} \ , \tag{2}$$

where η_i and δ_t are municipality and year fixed effects and $T \ll N$. Estimates for β_1 from this equation are typically biased upwards due to positive selection. We are augmenting the structural equation (2) with the first stage regression

$$s_{it} = \alpha z_{it} + \psi_i + \theta_t + \nu_{it} \ , \tag{3}$$

where z_{it} measures exposure during the period of the policy. This approach both accounts for endogeneity in the overall number of physicians and the intensity of exposure to the policy itself.

The triangular system outlined in equations 2 and 3 using physician emigration or an alternative measure for exposure is our preferred model specification. We estimate the model using two-stage least squares. Further extensions are estimated using a control function approach, i.e., two-stage residual inclusion, estimated as a single GMM system specification to avoid the generated regressor problem for correct inference. We also rely on control function estimation for the semiparametric model we develop in section 4.3.

Our main estimates rely on variation in the number of registered physicians induced by the emigration of physicians. We observe the yearly number of physicians emigrating per municipality, i.e. the number of physicians who give up their residency and move to a foreign country. Note that we explicitly use emigration *of physicians*, not overall emigration. Since emigration is a time-variant measure, it also accounts for dynamic effects

[2] We explicitly choose to outline the setting in an IV framework since the treatment variable (number of physicians) and exposure (policy affects a varying amount of physicians) are *continuous* measures. Popular difference-in-difference approaches rely on the comparison of distinct groups defined by *discrete* exposure variables.

when vacancies are filled again. Moreover, the emigration data also captures non-Jewish physicians who were forced to leave the country due to the occupational restrictions.

We are aware that emigration is a downstream measure with regard to the policy and can be considered potentially endogenous. However, this would only invalidate our analysis if the response to the policy change among physicians is heterogeneous in a way that is related to the change in outcome. Unless there is severe heterogeneity of this kind in the propensity to migrate between physicians as a response to the policy, the fact that emigration is a downstream measure of the policy does not invalidate the use of physician emigration as a regressor and emigration can be considered a valid proxy of the exposure to the policy. Strictly speaking, using emigration as an instrument requires that the propensity to migrate as a response to the policy is homogeneous and unrelated to unobservable time-variant factors influencing mortality. If this assumption holds, emigration is preferable to the policy itself or other time-invariant measures of the policy impact, as it offers richer variation for identification.

With this in mind, we also employ alternative measures of exposure to the policy. We use the two proxies for the number of Jewish physicians expelled in 1933 as alternative instruments, i.e., the number of Jewish physicians remaining after 1936 and the local Jewish population in 1933. In this case, exposure z_{it} is equal to the interaction between the respective proxy for Jewish physicians and the period indicator $\mathbb{1}\{t \geq 1933\}$. Hence, these measures only proxy for treatment intensity at a single point in time and therefore offer less identifying variation compared to our preferred specification. They also do not account for the fact that physicians with certain political affiliations were also banned from practicing and left the country as well.

Although both variables are only imperfect proxies, when using them to measure exposure, the measurement error is not going to affect the estimates if the propensity to remain in 1936 or to become a physician is constant within the Jewish population. For the analysis, it does not matter if the exposure measure suffers from measurement error in the level, as long as it is still comparable cross-sectionally. More general, the measurement error in the proxy will not confound the estimates unless it is correlated with municipal unobservable factors that are related to health.

Ultimately, we find that results from using all three measures are very similar, suggesting they all capture policy exposure. Using a binary transformation (any physician emigration) also leads to estimates of similar magnitude as our main specification.

4.2 Threats to identification

We discuss the assumptions implicit to the estimation procedure in turn. Monotonicity (or in the linear case, homogeneity in the form of a common α), i.e., that emigration of physicians can only influence the total number of physicians in one direction or not at all is likely fulfilled: Outmigration reduces the number of registered physicians. It is unlikely that excess migration of physicians in the opposite direction from other municipalities or countries would immediately overcompensate for the reduction. Another possibility is that if the market for physicians were fully saturated and non-registered German physicians working in other professions now overcompensate for the reduction due to emigration. These scenarios do not seem plausible.

Relevance is an empirical issue. Emigration of physicians naturally reduces the count of registered doctors. Our first stage results show a significant negative effect that is statistically indistinguishable from a one-to-one relation. This result is as expected und suggests the instrument is sufficiently relevant.

The exclusion restriction required for identification necessitates that emigration of physicians only influences mortality by changing the overall supply of physicians. In other words, changes in the instrument (i.e., the propensity of physicians to migrate/the level of Jewish physicians/the level of Jewish population) within a municipality over time only influence population health by affecting the total number of physicians available. This means physician migration cannot influence mortality directly and must be unrelated to other unobserved time-variant factors that influence mortality. While the first mechanism—physician emigration directly affecting health—can credibly be ruled out, we discuss possible confounding factors via the second mechanism in more detail.

One potential confounding factor would be the influence the laws had on other professions.

However, as discussed previously, other health care professions were not affected by the laws and did not feature a similar overrepresentation of Jewish individuals. Other affected professions (university professors, public officials) were both smaller in numbers and are unrelated to health.

Another concern is that discrimination and a hostile environment may have led to a higher mortality among the Jewish population and that this is driving our results. We consider this issue in our sensitivity analysis and show that it is virtually impossible that this mechanism is driving our results. Considering the large disparity in population share among physicians and the general population (0.7%), no reasonable parameter scenario of birth and death rates can generate our results. The Jewish population would have to have birth rates multiple times that of the rest of the population combined with 100% infant death rates. We also find that areas with larger Jewish populations are not more likely to vote for the NSDAP (cf. balance tests below). Nazi policies also generally aimed at improving population health.

More generally, a valid concern is that a violation may be caused by non-random assignment of the instrument, i.e. if the spatial distribution of Jewish physicians is correlated with other features that impact population health in a way that is not absorbed by municipality or time fixed effects. Jewish communities were more prevalent in certain regions within Germany and also within metropolitan areas. Should health and population dynamics differ substantially between more and less populous or urbanized regions, the exclusion restriction may be violated such that $\mathrm{E}[z_{it}\epsilon_{it}] \neq 0$. This concern is partly alleviated by the fact that our sample is already preselected on more populous, urban municipalities. Nevertheless, we address the issues of population dynamics and regional confounding factors in detail in the robustness checks and show that our results are unlikely to be driven by these factors. We also develop several more general approaches to check instrument validity, showing that the exposure measures neither predict past outcomes nor affect several placebo mortality outcomes that are unlikely to be influenced by physicians. A placebo instrument (emigration of psychiatrists) also does not have any effect on mortality. Finally, we consider specifications including the interaction of year fixed effects and state-level, province-level or district-level fixed effects or additional linear trends to rule out that our results are driven by other location-based factors that may be correlated with emigration and health.

Moreover, we find that municipalities with larger shares of Jewish inhabitants do not differ from other areas. Strictly, balance in observable characteristics is not a test for time-variant confounding unobservable factors. However, the absence of such confounding factors and the exclusion restriction are more credible if municipalities are similar in observable characteristics ex-ante. In Table A2, we contrast the means for mortality outcomes and political and socio-demographic characteristics for municipalities with below and above Jewish population shares prior to the introduction of the occupational restrictions. We find no difference between municipalities in any of the major outcomes we look at with the exception of a small difference in the pneumonia mortality rate. Moreover, municipalities are very similar with regard to political preference. Importantly, there is no difference in the share of votes for the Nazi Party (Nationalsozialistische Deutsche Arbeiterpartei, NSDAP). Only in the middle of the political spectrum there is a minor difference. Voters in municipalities with larger Jewish populations vote slightly more for the Centre Party (Zentrum), while voters in other municipalities cast more votes for the Social Democrats (Sozialdemokratische Partei Deutschlands, SPD) or abstain from voting. Regarding other characteristics, areas with larger relative Jewish population are slightly more populous. Since all variables are expressed relative to base populations, this should not matter for the results. Still, we explicitly address population dynamics in the robustness checks. In addition, the groups do not show any differences in population growth prior to 1933. There are also no differences in key labor market characteristics, with both groups having similar rates of labor force participation, female labor force participation, total unemployment and individuals on basic social assistance. This makes us confident that our identifying assumptions are credible and met.

4.3 Semiparametric model

Diminishing returns matter for vital outcomes like mortality. A simple linear specification as given in (2) may not fully capture the mortality effects of physicians. The effect of physicians is most likely nonlinear, as an additional doctor in a region with sparse coverage can prevent more infant deaths compared to an additional doctor in a saturated region. After a certain coverage ratio has been reached, mortality effects are likely to taper off. Most deaths resulting from treatable conditions will have been prevented and the remaining cases are subject to conditions which are increasingly difficult to treat. While morbidity effects may persist, systematic fatality reductions beyond a lower bound of medically difficult cases are unlikely. We investigate the hypothesis of diminishing returns by employing a semiparametric estimation approach using a partially linear model. We combine the Baltagi and Li (2002) semiparametric fixed-effects estimator with a control function (e.g. Heckman and Robb 1985) to derive a flexible estimate of the effect of physicians on infant mortality.

Consider a general panel model of the form

$$y_{it} = f(s_{it}) + \eta_i + \epsilon_{it} \tag{4}$$

$$s_{it} = g(z_{it}) + \psi_i + \nu_{it} \; . \tag{5}$$

Time fixed effects are dropped for notational convenience. To control for the endogeneity in the physician supply s_{it}, we add the control function $\hat{\nu}_{it}$ for s_{it} obtained from (3) to the model as a linear term,

$$y_{it} = f(s_{it}) + \hat{\nu}_{it}\rho + \eta_i + \zeta_{it} \; . \tag{6}$$

The control function approach assumes that the correlation between the structural error ϵ_{it} and the first stage error ν_{it} can be described as a linear relationship $\epsilon_{it} = \nu_{it}\rho + \zeta_{it}$ with $\mathrm{E}[\nu_{it}\zeta_{it}] = 0$. The endogenous variation in s_{it} is corrected using the estimated control function $\hat{\nu}_{it}$.

We estimate this model using the first difference approach described in Baltagi and Li (2002), as the conventional Robinson (1988) double residual estimator cannot accommodate the municipality-specific intercepts. To eliminate the fixed effects, take first differences of (6) with respect to time,

$$\Delta y_{it} = \Big\{ f(s_{it}) - f(s_{it-1}) \Big\} + \Delta \hat{\nu}_{it} \rho + \Delta \zeta_{it} \ . \tag{7}$$

Baltagi and Li (2002) show that $f(s)$ can be approximated by a power series $p^k(s)$, and $\Big\{ f(s_{it}) - f(s_{it-1}) \Big\}$ by $\Big\{ p^k(s_{it}) - p^k(s_{it-1}) \Big\} \gamma$. Equation (7) can be rewritten as

$$\Delta y_{it} = \Big\{ p^k(s_{it}) - p^k(s_{it-1}) \Big\} \gamma + \Delta \hat{\nu}_{it} \rho + \Delta \zeta_{it} \ . \tag{8}$$

We use cubic B-splines to approximate $f(s_{it})$ and $f(s_{it-1})$. Having estimated the series terms, γ and ρ can be estimated with least squares from (8). The parameters and the nuisance estimates can be used to fit the fixed effects $\hat{\eta}_i$ by residualizing and averaging within panel units. Together, the estimates can be used to get the partialled-out residuals

$$\hat{u}_{it} = y_{it} - \hat{\nu}_{it} \hat{\rho} - \hat{\eta}_i = f(s_{it}) + \zeta_{it} \ . \tag{9}$$

We then fit $f(\cdot)$ by regressing the residual $\hat{u}_{it}$ on s_{it} using local polynomial regression. The first derivative of the obtained function with respect to s is the desired marginal effect, $\mathrm{d}y/\mathrm{d}s = \mathrm{d}f(s)/\mathrm{d}s$.

Valid inference in this multi-stage estimation approach needs to account for the estimation error of the plug-in estimates. We employ a wild cluster boostrap to obtain asymmetric Bootstrap-t confidence intervals for the residualized outcome and the marginal effect (cf. Cameron et al. 2008, Davidson and MacKinnon 2010).

5 Results

In the first part of this section, we investigate the direct effects of different policy exposure variables on the supply of health care and infant mortality, estimating different first-stage regression specifications and a reduced-form event study analysis. We then proceed to estimate the effect of physicians on different infant mortality outcomes and the incidence of various childhood diseases using the main approach outlined in the previous section. Next, we investigate how changes in the supply of physicians interact with the availability of hospital infrastructure. We then use our semiparametric modeling approach to investigate diminishing returns to health care provision. Finally, we conduct a variety of sensitivity analyses and robustness check to ensure the validity of our results.

5.1 The policy's effects on physician supply and infant mortality

As a first step, we estimate the effect of exposure to the policy on the supply of physicians. Results in Table 1 are based on different variants of equation 3. The estimates in panel (a) are based on a specification using the rate of physicians per 1,000 of population as an outcome and the different exposure measures. We find that one additional physician emigrating after 1933 reduces the number of physicians by about 1.1. We cannot reject the null hypothesis that the coefficient is equal to one, indicating a one-to-one relationship as expected. If we measure exposure as the number of jewish physicians post 1933, we find that one additional jewish physician leads to a reduction of about 0.4 physicians. Corrected for order of magnitude, the estimate for exposure measured as overall Jewish population is of similar size.

In panel (b), we consider the same specification but using the log of physicians as an outcome. Note that there are no municipalities without physicians in our data, so we avoid the log-of-zero problem. We find that all estimates are very similar to those in panel (a), suggesting that logging the treatment variables has about the same effect as normalizing it as a rate over population. In panel (c), we estimate a log-log specification, using the inverse hyperbolic sine transformation as some of the exposure variables have zero counts

Table 1: The effect of emigration on the supply of physicians

(a) Dependent variable: Registered physicians per 1,000 of population (rate) (level-level)			
# of physicians emigrating	−1.149***		
	(0.211)		
# of Jewish physicians		−0.380***	
		(0.096)	
Jewish population in 1933			−0.004***
			(0.001)
Year FE	✓	✓	✓
Municipality FE	✓	✓	✓
First stage F-stat	29.69	15.60	14.88
N municipalities	317	317	317
N	2853	2853	2853

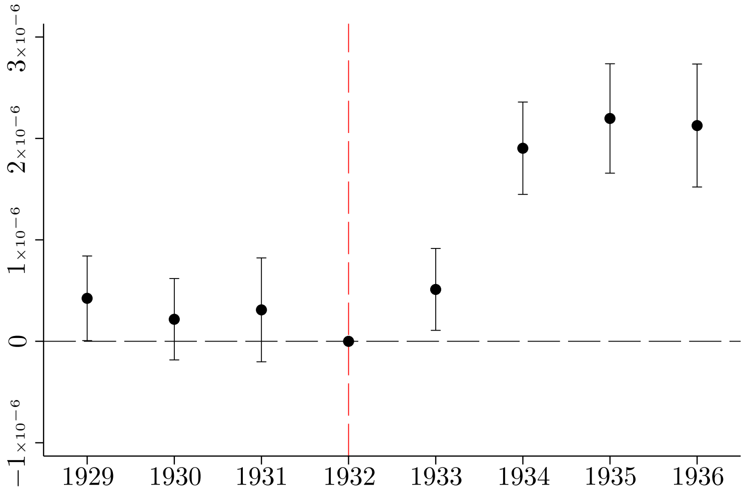
(b) Dependent variable: Registered physicians (ln) (log-level)			
# of physicians emigrating	−1.116***		
	(0.187)		
# of Jewish physicians		−0.362***	
		(0.086)	
Jewish population in 1933			−0.004***
			(0.001)
Year FE	✓	✓	✓
Municipality FE	✓	✓	✓
First stage F-stat	35.69	17.66	16.70
N municipalities	317	317	317
N	2853	2853	2853

(c) Dependent variable: Registered physicians (ln) (log-log)			
# of physicians emigrating (inv. hyp. sine)	−0.051***		
	(0.007)		
# of Jewish physicians (inv. hyp. sine)		−0.047***	
		(0.005)	
Jewish population in 1933 (inv. hyp. sine)			−0.036***
			(0.004)
Year FE	✓	✓	✓
Municipality FE	✓	✓	✓
First stage F-stat	54.58	104.94	67.99
N municipalities	317	317	317
N	2853	2853	2853

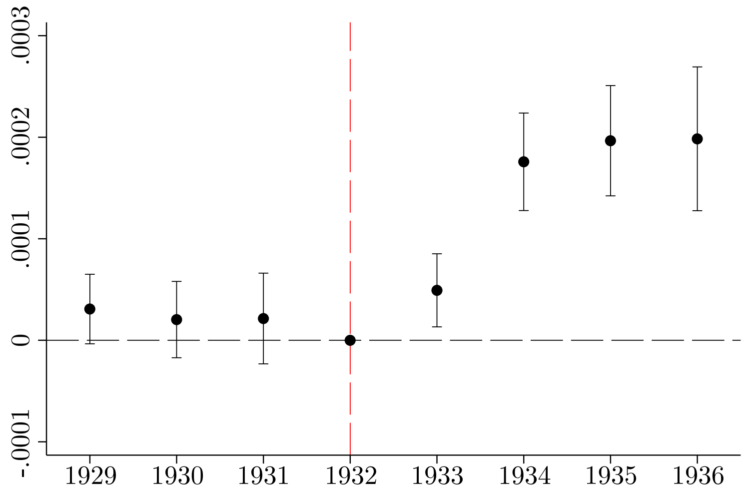
Notes: First stage estimates for different exposure variables using a linear regression model. The dependent variable in panel (a) is the rate of registered physicians per 1,000 of population in $t-1$ and the instrumental variables are measured in levels (level-level specification). The dependent variable in panel (b) is the log of the number of registered physicians and the instrumental variables are measured in levels (level-level specification). Estimates in panel (a) are scaled by 10^3. The dependent variable in panel (b) is the log of the number of registered physicians, the dependent variables are transformed using the inverse hyperbolic sine transformation to approximate the natural log in the presence of zero counts (log-log specification). Included in all specifications are a full set of year and municipality dummies. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

Figure 4: Event Study Analysis

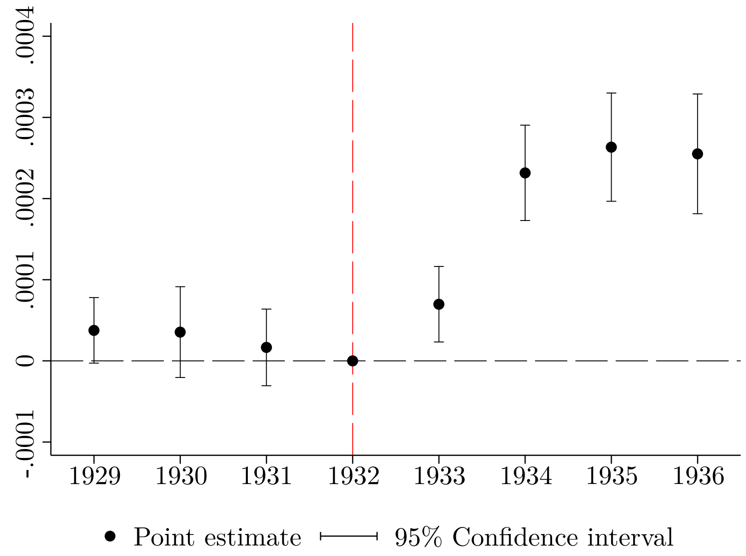
(a) Infant mortality (ln), exposure: year x 1933 jewish population



(b) Infant mortality (ln), exposure: year x jewish physicians



(c) Infant mortality (ln), exposure: year x emigration



Notes: Estimates are based on a specification including year dummies interacted with the respective exposure measure. The excluded reference period is highlighted. The graph plots the coefficient point estimates and the corresponding 95% confidence intervals over time.

in the data. We find that a one percent increase in emigration leads to about a five percent decrease in the total number of physicians. Similarly, a one percent increase in the number of Jewish physicians after 1933 also leads to a five percent decrease in the overall number. The effect for a one percent increase in the Jewish population is about four percent. Since these relative effect estimates are all of about the same size and order of magnitude, we are confident that our variables are relevant and all measuring the same underlying exposure.

Given the similarity of findings, for most of the regression analysis, we prefer using rates for comparability as some outcome measures are sparser and have a higher rate of zeros. As we document in Figure A1, normalizing variables by population works well. We show in the robustness checks that using various other log-linear regression approaches which avoid the log-of-zero problem lead to the same results.

Next, we conduct a reduced-form event study analysis to ascertain the direct effect of the policy exposure on infant mortality. Figure 4 shows the results from estimations based on the reduced-form specification 1. We find that all exposure measures lead to higher infant mortality post 1932. All three series of estimates show the same pattern and are about the same size and magnitude (after accounting for scaling when considering population). Comparing panel (c) to the other measures in panels (a) and (b), we find that estimates for emigration as the exposure measure are slightly larger and more precise. Importantly, none of the estimates prior to 1933 are significant, indicating there are no pre-trend effects preceding the policy that could invalidate the remainder of our results. The post-ban effects are comparable in size to a simple single-coefficient reduced-form estimation approach.

Given the intuitive one-to-one first-stage relationship and more precise reduced-form, emigration is our preferred exposure measure. Nevertheless, for the remainder of the paper, we show results for all three measures for transparency and to stress that they are all similar.

5.2 The effect of physicians on infant mortality and stillbirths

The main results are presented in Table 2. The reported coefficients can be interpreted as the effect on mortality if the coverage ratio increases by one additional physician per 1,000 people. Since the mean coverage ratio in the sample is 1.05, the estimates also approximately indicate the effect of doubling the coverage ratio.

The first panel in Table 2 shows the results when directly regressing health outcomes on physician coverage. For any outcome, the effect is indistinguishable from zero. The next panel shows our preferred specification, instrumenting the physician coverage ratio with the emigration figures. Looking at column (1), we find that one additional physician reduces overall infant mortality by about 18 cases per 1,000 live births. The average yearly infant mortality prior to 1933 is about 80. In relative terms, doubling the coverage ratio reduces infant mortality by about 23%. While this effect is large, doubling the number of physicians is also a drastic intervention. In terms of standard deviations, a one standard deviation increase in the number of physicians would lead to reduction in mortality of about 0.45 standard deviations.

For infant deaths caused by inflammatory bowel diseases (ulcerative colitis, enteritis, diarrhea or other ulcerations of intestines), we also find a significant negative effect. Inflammatory bowel diseases are a prominent cause of death for infants, especially in conditions of sub-standard hygiene and contaminated drinking water. This effect is also large in relative terms, presumably since gastrointestinal diseases were already known as a prominent cause of infant mortality in the 1930s.

In contrast, the estimates in column (3) show that increasing the physician coverage ratio has no effect on infants dying from premature birth or congenital defects. This is an intuitive result. Medical complications arising from premature birth and congenital debility are notoriously difficult to treat and account almost completely for the remaining cases of infant mortality in developed economies today. Note that the number of observations for this set of regressions is lower due to a shorter coverage period in the data.

Regarding stillbirths, our results in column (4) show that an additional physician reduces the number of stillbirths by about 8 cases. In relative terms, this corresponds to a decrease

Table 2: Mortality estimates

	(1)	(2)	(3)	(4)
	Infant mortality			
	Total	Infl. bowel diseases	Prem. birth, cong. debility	Stillbirths
	OLS			
# of registered physicians	2.71	−0.10	−0.43	−2.85
	(3.61)	(0.93)	(3.35)	(2.12)
	IV: Emigration			
# of registered physicians	−18.79***	−7.11***	−2.48	−8.42***
	(4.17)	(0.98)	(5.09)	(2.24)
First stage F-stat.	29.70	29.70	36.18	36.18
	IV: Jewish physicians in 1933			
# of registered physicians	−13.62**	−4.18***	−0.92	−6.11***
	(6.41)	(1.20)	(4.02)	(1.64)
First stage F-stat.	15.60	15.60	17.79	17.79
	IV: Jewish population in 1933			
# of registered physicians	−13.10**	−4.41***	−0.29	−6.88***
	(6.39)	(1.14)	(4.31)	(1.78)
First stage F-stat.	14.89	14.89	17.05	17.05
Year fixed effects	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓
N municipalities	317	317	317	317
N	2853	2853	1902	1902

Notes: Infant mortality variables are measured per thousand live births in $t-1$, stillbirths per thousand total births in $t-1$. Registered physicians are measured per thousand of population in $t-1$. Included instruments are a full set of year and municipality dummies. Excluded instruments are given in the respective paragraph header if applicable. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

of about 27% (prior to 1933 there were on average 30 stillbirths per 1,000 births per municipality).

Comparing the results using emigration as the instrument (second panel) to the alternative instruments (panels three and four), results are very similar. All estimates are significant and results are comparable in magnitude, albeit consistently slightly smaller. This may be due to the fact that the instruments are time invariant (after 1933), and although they factor in the magnitude of the policy change in 1933, they cannot account for the selection of physicians into municipalities with vacancies in later years. In addition, the emigration measure also covers variation due to physicians who were affected by the occupational restrictions and left the country for political reasons.

Table 3: Disease incidence

	Incidence rate by cause of death per 1,000 of population						
	Bronchitis	Influenza	Measles	Pneumonia	Diphteria	Scarlet fever	Typhoid fever
# of registered Physicians	−0.047* (0.028)	−0.045** (0.018)	−0.013*** (0.005)	−0.325 (0.213)	0.008 (0.018)	−0.001 (0.005)	−0.003 (0.003)
Year FE	✓	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓	✓
First stage F-stat.	36.18	29.70	29.70	29.70	29.70	29.70	29.70
Unconditional mean	0.15	0.10	0.02	0.75	0.08	0.01	0.01
N municipalities	317	317	317	317	317	317	317
N	1902	2853	2853	2853	2853	2853	2853

Notes: All dependent variables given as incidence rates per thousand of population in $t-1$. Registered physicians are measured per thousand of population in $t-1$. Included instruments are a full set of year and municipality dummies. Excluded instrument is yearly emigration of physicians. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

5.3 Disease incidence

We extend our analysis to mortality rates from a set of diseases that, although not limited to, are a common cause of death for children. All variables are stated as population incidence per 1,000 persons. Bronchitis, diphteria, influenza and pneumonia are among the more frequent causes of death with between 0.1 to 0.8 deaths per 1,000 individuals.

The results for disease incidence are given in Table 3. We find that one additional physician can decrease mortality due to bronchitis by about 4.7 cases per 100,000. The effect for influenza mortality is of comparable magnitude with a reduction of 4.5 cases. For measles, we find that incidence is reduced by about 4.5 cases per 100,000. Given the comparatively low baseline incidence, these effects are substantial. We cannot reject a zero effect for diphteria, scarlet fever and typhoid fever.

We find negative effects mostly for viral diseases. Viral diseases cannot be treated by antibiotics and with the exception of measles, vaccinations for the viral diseases we consider are not widespread and do not offer lasting protection. Treatment for these diseases typically consists of non-steroidal anti-inflammatory drugs like Aspirin to reduce fever and inflammation. These drugs were known and readily available in pharmacies in the 1930s. In addition to this, physicians recommend sufficient rest, fluids and nutrition. They also offer behavioral advice that limits infections and reduces the spread of the diseases.

However, we do not want to put too much emphasis on this result, as it may also be driven by a combination of power and the set of diseases for which data is available. Most widespread bacterial diseases have very low fatality rates, and the bacterial diseases we observe are both rare and have very low overall mortality rates. The diseases we find effects for are mostly those with larger baseline incidence. It might be that for some of the low-incidence diseases, the noise in the data overpowers the signal and our sample size is insufficient to pick up the effect. The only exception of a more prevalent bacterial disease in our data is pneumonia, for which the p-value is 0.12, suggesting that power may be an issue when trying to pick up effects for rare conditions.[3]

Another potential explanation is that effects are partially driven by a quality of care mechanism. It is unlikely that people at the margin are going to forego medical care completely given universal health insurance, but they may receive lower quality care due to congestion or delay seeking treatment. Physicians which are less pressed on time are more likely to diagnose these diseases early and correctly, and also more likely to give more exhaustive behavioral advice on how to prevent infections in the first place. At the same time, mothers with children who develop symptoms of sickness may potentially delay visits to the physician in anticipation of long waiting times, and only seek treatment after the infant's health condition has worsened.

5.4 Hospital infrastructure

In this section we analyse how the provision of physicians interacts with the availability of health care infrastructure. To do so, we augment our preferred model for infant mortality from subsection 5.2 using emigration as the instrument. We add an interaction of the physicians variable with various infrastructure proxies to the model. Specifically, we consider whether the municipality lacks a hospital of a certain type of specialization, and whether the bed capacity of the public hospitals in a municipality is below the median or in either of the lower terciles. All infrastructure variables are measured in 1933 at the start of the employment restrictions. To allow for consistent estimation and inference of

[3]Note that pneumonia can be caused by both viral and bacterial mechanisms, although most of the infections at the time are likely to have been bacterial.

Table 4: Infant mortality and hospital infrastructure

	(1)	(2)	(3)	(4)	(5)	(6)
	Infant mortality					
# of physicians	−19.24***	−19.18***	−16.83***	−16.07***	−18.70***	−18.78***
	(5.04)	(5.09)	(5.65)	(5.80)	(5.13)	(5.14)
Hospital type						
# of physicians x no public hospital	−0.09					
	(7.66)					
# of physicians x no university hospital		−1.57				
		(5.61)				
# of physicians x no infant hospital			−13.65*			
			(7.03)			
# of physicians x no maternity hospital				−17.88**		
				(7.30)		
Hospital capacity						
# of physicians x $\mathbb{1}\left\{\text{beds} \leq p(50)\right\}$					−10.66*	
					(6.29)	
# of physicians x $\mathbb{1}\left\{p(33) \leq \text{beds} \leq p(66)\right\}$						−1.54
						(6.78)
# of physicians x $\mathbb{1}\left\{\text{beds} \leq p(33)\right\}$						−12.64*
						(7.64)
Year fixed effects	✓	✓	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓	✓	✓
N municipalities	317	317	317	317	317	317
N	2853	2853	2853	2853	2853	2853

Notes: Infant mortality variables are measured per thousand live births in $t-1$, stillbirths per thousand total births in $t-1$. Registered physicians are measured per thousand of population in $t-1$. Included instruments are a full set of year and municipality dummies. Excluded instrument is yearly emigration of physicians. Results are based on a control function specification estimated using a GMM system of equations. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

the interaction coefficient without additional instruments, we estimate the model using a control function approach. The control function approach is implemented as a single GMM system to obtain consistent standard errors. The results are shown in Table 4.

Considering the results for hospital type, we do not find any additional effect for physicians in municipalities which lack a public hospital. This is not surprising, as public hospitals are very common and more than 75% of municipalities have at least one public hospital. We then look at more specific specializations. Looking at column (2), we also do not find any additional effect of physicians if the municipality does have a university-affiliated hospital. However, we do find additional effects for physicians in municipalities which do not have hospitals specializing in infant or maternal care (columns 3 and 4). This implies that losing a physician in a municipality without specialized hospital infrastructure carries a larger mortality penalty compared to a situation where this type of infrastructure is in place. This result shows that health care infrastructure is only a partial substitute for physician manpower, and suggests that both inputs are complements for health production.

We confirm this finding by looking at an alternative infrastructure measure, hospital capacity. For this analysis, we distinguish between municipalities which are at the lower end of the hospital bed capacity distribution and those that are not. Again, we find that the effect of physicians is larger in municipalities which have comparatively worse infrastructure. Notably, the effect seems to be especially large in municipalities whose hospital bed capacities are below the 50th and 33rd percentile. The effects are of comparable magnitude as those for women and children hospitals, albeit slightly smaller. Taken together, these results suggest that clinic infrastructure matters for the effect of health care provision, but cannot replace physicians and human capital input.

5.5 Nonlinear mortality effects

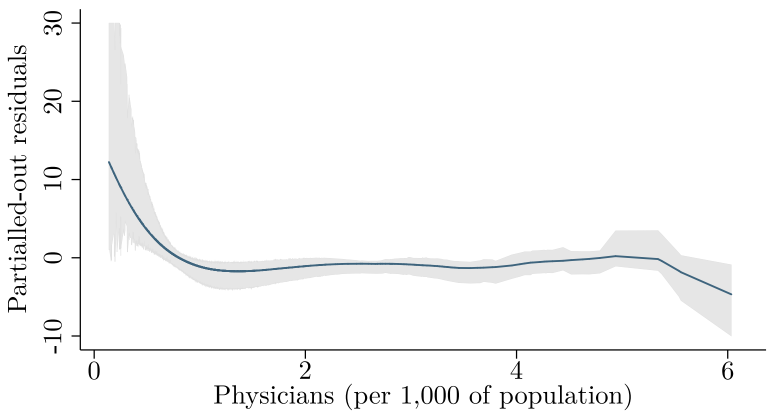
In this part we evaluate possible non-linear marginal effects. We conjecture that there are diminishing returns to health care provision. The marginal patient is likely to differ when approaching the zero-mortality lower bound. As the number of physicians increases, the remaining mortality cases will be characterized by diseases with higher frailty for which treatment is less effective. Results for the semiparametric model developed in 4.3 are presented in Figure 5.

Panel (a) plots an estimate of the conditional mean function using the partialled-out residuals. This function provides an estimate of the average residual infant mortality by number of physicians after controlling for endogenous selection of physicians, municipality and time fixed effects. Residual infant mortality is positive for smaller values of physician coverage and decreases as the number of physicians decreases. For large regions of the support of physicians, residual infant mortality is indistinguishable from zero. We derive the dose-response function as the estimated first derivative of the residualized outcome function. Panel (b) shows our main result, the marginal effect of increasing the physician coverage ratio on infant mortality.

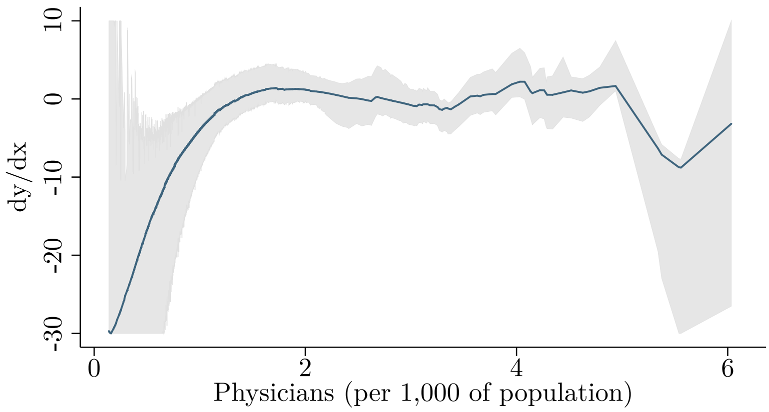
We find that physicians can strongly reduce infant mortality when coverage is sparse, but that the effect declines as coverage increases. Importantly, mortality effects are

Figure 5: Semiparametric analysis

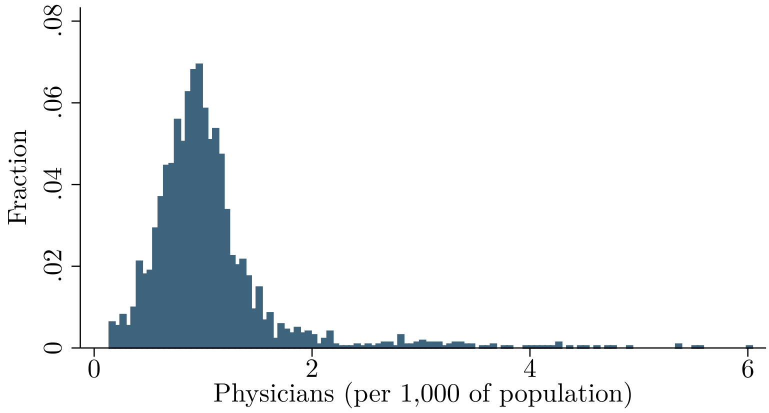
(a) Residual infant mortality



(b) Marginal effect



(c) Density



Notes: Figure (a) plots the partialled-out residuals of the infant mortality rate by physician density; accounting for the control function of the physician coverage ratio, individual and time fixed effects following the framework outlined by Baltagi and Li (2002). Observations with values outside the scale of the y-axis have been cropped from the scatterplot to improve exposition. The nonparametric fit was generated using local polynomial regression with a polynomial of degree 4, an Epanechnikov kernel function and a bandwidth of 1.3 chosen by Silverman's rule-of-thumb. Bootstrap-t confidence intervals shown in the graph are computed using the wild cluster boostrap. Figure (b) plots the marginal effects by physician density, i.e. the first derivative of the above function for each level of the independent variable. Shaded in grey in both plots are the 95% confidence intervals. Figure (c) plots the distribution of the physician density to better illustrate sparse regions in the support.

restricted to a specific interval of low coverage. After the coverage ratio reaches about 1.8 physicians per 1,000 of population, mortality effects subside. The estimated marginal effect is indistinguishable from zero for higher coverage ratios, in a region where the data is still dense (cf. panel c). Physicians most likely still influence morbidity, general health and quality of life in the population. However, after the population coverage ratio exceeds two physicians per 1,000, any remaining cases of infant mortality are unrelated to physician coverage and most likely medically difficult cases.

Our results mirror the development of the industrialized world. In most developed countries, infant mortality rates have barely changed since the 1980s, when physician levels reached a comparable magnitude (cf. section A.3). Together with positive selection, this also offers a comprehensive expanation for the prevalence of null results in studies which investigate the effect of physicians using recent data from developed countries.

5.6 Sensitivity analysis

We conduct a series of robustness and specification checks to show that our results are valid and not driven by alternative mechanisms or the choice of a specific model. In this section, we outline these tests and present their results. First, we focus on violations of the main identifying assumption, the instrument's exclusion restriction, and devise checks for both general violations and specific mechanisms. Second, we focus on alternative explanations for our results and more general modeling and specification concerns.

5.6.1 Exclusion restriction

The discussion in section 4 raises the point that a violation of the exclusion restriction could occur if districts with a higher rate of Jewish physicians evolve differently over time in a way that is related to mortality. For example, if the municipal population structure changes differentially for different levels of the instrument, population dynamics could also differ over time. Since we are restricting the outcome to mortality figures, any potential dynamic confounding mechanism will necessarily manifest in a municipality's population dynamics. To address this issue, we replicate the main results from Table 2 in Table A3 in the appendix, adding the logarithm of population and population growth in $t-1$ as covariates to control for population dynamics. The results replicate those in Table 2 almost exactly. In fact, precision improves for some of the estimates. The findings are robust to the inclusion of higher order polynomials as well. In addition, changes in population health may not occur instantaneously, but only with slight delay. We have tested this extensively, and our results are replicated almost exactly with a variety of alternative lag structures in the treatment variable.

Next, we consider another check to show that our results are not driven by unobservable time-variant trends in health outcomes. If there are other factors that influence mortality in a municipality over time, it is unlikely that these factors are restricted only to a single municipality, but most likely also influence other municipalities nearby due to regional concentration. We utilize this fact and extend our analysis by allowing for different year fixed effects at the state level or additional linear time trends on different regional levels. The results are shown in Table A4 and Table A5 and correspond to our main results in Table 2. First, in Table A4 we allow the year fixed effects to vary at the state level. This should pick up any variation at the regional level without imposing additional functional form. All coefficients are very similar to our main results, suggesting that potentially confounding regional unobservable trends are absent.

Since the number of parameters needed to estimate increases substantially for smaller regional units, we then proceed to allowing for linear trends at different regional levels in addition to the year fixed effects in Table A5. We decrease the size of the regional trend unit incrementally. Estimates in panel (a) are based on a specification including state-level time trends, those in panel (b) on a specification including province-level time trends, and those in panel (c) on a specification including vote district-level time trends. All estimates are very similar and comparable in magnitude to our main results. Some of the estimates are slightly more noisy when accounting for trends within smaller regional units due to the larger number of parameters, but all of them are still comparable in magnitude and the confidence intervals include the original estimates. This corroborates that our results are not driven by regional-level unobservable trends over time. Similarly, the event study-results in Figure 4 also suggest that our results are not driven by deviating within-municipality trends in mortality prior to 1933.

As mentioned in section 3, we also report additional general checks for confounding trends. The results are given in Table A6. First, we check whether future instrument values (post-occupational restrictions) can predict past outcomes (pre-occupational restrictions). The results are given in panel (a). We do not find any evidence that future instrument values are related to past outcomes, all coefficients are indistinguishable from zero at conventional significance levels, providing no evidence of possible confounding.

Second, we conduct robustness checks using a placebo instrument. We gathered data on emigration of psychiatrists and re-estimate our main specification using this measure as an instrument. The results are given in panel (b), Table A6, and are as expected: variation in the number of physicians induced by the number of psychiatrists emigrating does not influence any of our main infant mortality outcomes, contrary to physician emigration. All estimates are insignificant at conventional levels.

Third, we conduct a series of placebo checks. We repeat our main specification for outcomes which should not be affected by the instrument. The results are given in panels (c)-(e) of Table A6, separately for each instrument measure. We test whether physicians affect the mortality incidence of strokes, deaths due to old age, and those classified as dying of unknown reasons. As expected, we do not find any evidence that physicians influence the number of deaths due to these conditions. These results suggest that our findings are unlikely to be driven by differential underlying trends in health and health behavior, as these would most likely also impact these outcomes. Moreover, our instrument seems to influence the correct fundamental measure we are interested in, physician coverage.

Another concern is substitution, i.e. that health care demand is satisfied by visiting out-of-municipality physicians. If this behavior were to occur, it would likely bias our results towards zero. However, we believe that such bias is absent or marginal at best. The municipalities we consider are sufficiently large such that outside substitution is likely to be negligible in comparison. In our data, we only consider mortality of individuals registered within the municipality in question. From the limited vital statistics for children of outside residents we have, we cannot discern any noticeable trend in the regions were there are few Jewish physicians.

Finally, an alternative mechanism that could explain our findings is if higher mortality is restricted to Jewish families. In that case, mortality effects may simply be caused by other discriminatory measures instead of the reduced quality of health care. Unfortunately, we are unable to distinguish mortality by population groups. However, a simple back-of-the-envelope calculation reveals that it is impossible that the sizeable mortality effects are restricted to Jewish communities alone because of their disproportionately small population share. Assuming equal birth ratios, we find that for every 1,000 live births among the Jewish population, 2,930 infants would need to die to account for the effect in our main specification. This means that three times as many children as were actually born according to the Jewish population rate would have to die, and not a single Jewish baby would survive infancy. This scenario is impossible, and any feasible scenario would involve unrealistic parameter values. This suggests that the effect cannot be driven by mortality among the Jewish population.

More generally, while the general environment grew more hostile towards Jewish citizens after 1933, they still retained their citizenship status and full health insurance. Only beginning in late 1935, the *Nürnberger Gesetze* limited Jewish citizenship and marriage rights. After 1936, when preparations for war began, further policies essentially removed Jews' citizenship rights gradually and targeted discriminatory measures against anybody with Jewish heritage intensified. Following the events of the *Novemberpogrome* in 1938, Nazi politics escalated from discrimination to systematic persecution, deracination and dispossession. Mass arrests without charges or proceedings commenced. The war started in 1939, the Holocaust began shortly after. For these reasons, we do not include data after 1936 in our analysis. The results are also robust to whether the year 1936 is included in the analysis or not.

5.6.2 Model choice and functional form

As discussed in sections 3 and 4, the distribution of some of the dependent variables also raises the concern that the simple linear model may not be appropriate, since some variables remain right-skewed after computing incidence rates. We repeat our main analysis using an exponential conditional mean model outlined in Appendix subsection A.1. This type of model provides a straightforward approach for IV estimation in a nonlinear setting using GMM. Exponential models also offer intuitive interpretation due to constant relative effects. Like linear models and related quasi-maximum likelihood methods, the model is consistent if the mean is corrrectly specified. It is also preferable to log-transforming the dependent variable in our case, as it avoids the log-of-zero problem. Dropping panel units with any zero counts leads to attrition, and adding an arbitrary constant is not a satisfactory solution. Results are given in Table A7.[4]

We choose a simple Poisson fixed effects model for the naive specification (Table A7, first row), which is consistent even if the distribution were misspecified. Comparing the results, all findings from our main analysis carry through. The naive model does not offer conclusive results, suffering from positive selection. Results for the instrumental variable models are extremely similar. For most variables, the relative effects are almost identical to those obtained from the linear specification, indicating that the linear approximation for the mean works relatively well.

Similarly, we are relying strongly on the linearity of the control function in the semi-parametric analysis. However, theoretically, any non-linear transformation of the control function can be used as a control function as well. As a further robustness check, we are repeating the semiparametric analysis including higher-order polynomials of the control function to see whether the results change. The results for the marginal effect are shown in Figure A2. The estimates appear to be very stable, making us confident that our findings reflect a causal relationship in the population and are not driven by spurious variation or

[4]For the main outcome, infant mortality, taking logs is feasible as there are no observations which take the value zero. Our main results are unchanged in this case (cf. Table A8). Alternatively, using the inverse hyperbolic sine transformation (e.g. Burbidge et al. 1988) also gives comparable results for all outcomes.

the linear functional form of the first stage.

A related concern is that all dependent variables are expressed as ratios to describe relative incidence. This is standard in the literature on health and development and in medical studies. We believe the advantages of this approach outweigh the disadvantages, meriting its application. The normalization allows for better interpretability and preserves comparability to standard measures used in other studies. As discussed previously, it reduces the distributional skewness of the raw count data (cf. Figure A1), improving the viability of standard linear models. To mitigate concerns, all birth and population counts used in the denominator of the dependent variable are lagged by one period to rule out contemporaneous effects. To further investigate this issue, we reestimate our main model using log infant mortality as the outcome, and sequentially add log population and log live births as control variables (Table A8). Compared to the model without population controls, including log population increases the coefficient slightly whereas including live births in the model reduces the coefficient slightly. Since the effect is present in the unconditional model and any changes are small, we believe that using relative incidence figures to express population mortality is the correct approach.

6 Discussion and conclusion

We analyze the effect of changes in physician supply on infant mortality and disease incidence during the mortality transition. Our results indicate substantial mortality reductions. Increasing the coverage ratio from one to two physicians per 1,000 of population decreases infant mortality by about 23% relative to the pre-1933 level. We find a similarly large effect for stillbirths. Increasing health care supply also reduces mortality due to inflammatory bowel conditions and common childhood diseases, specifically measles, influenza and bronchitis. The Nazis' dismissal of Jewish physicians resulted in a 10% drop in physician levels. A simple back-of-the-envelope calculation suggests this lead to more than 2,200 additional infant deaths each year, even in a conservative scenario and disregarding non-linearities.

Moreover, we find that mortality due to premature births and congenital conditions remains unaffected. Medical treatment for these conditions is difficult and they account for the majority of infant mortality in developed economies today. We also find that mortality effects are larger in regions without specialized hospital infrastructure or with limited hospital capacities, indicating that infrastructure and physician human capital are complementary inputs for health production.

In sum, our findings are consistent with Goodman-Bacon (2018), who analyzes the introduction of Medicaid in the 1960s and its effects on infant mortality. Focusing on health insurance coverage, he also finds large reductions in infant mortality driven by permitting at-risk population groups access to health care and relates the infant mortality reductions to improved acute care at birth. We complement these results by focusing on the intensive margin of care provision, its interaction with health care infrastructure and the shape of the dose-response relationship.

In doing so, we provide evidence for diminishing marginal returns to health care provision and underscore the importance of basic health care provision. Our semiparametric estimation approach reveals that the effect of increasing the physician coverage ratio on mortality is highly nonlinear. Mortality reductions are large in regions where coverage is sparse, but decrease quickly when coverage increases. Effects are only present in a narrow coverage region. After a ratio of about two physicians per 1,000 of population, mortality effects disappear.

This result highlights the important historical role of physicians and lines up with trends over the 20th century. Our estimate coincides with the physician coverage most developed countries reached in the late 1980s, and infant mortality has remained relatively stable at a low level since then. Much of the remaining infant mortality is accounted for precisely by those conditions for which we find no effect. In addition, the nonlinear relationship also provides an explanation why more recent studies analysing developed countries often fail to detect a significant relation between physician supply and mortality. Variations in physician coverage in developed countries post-1990 largely occur in regions of the support where physicians have no substantial influence on mortality anymore.

The historical setting of our analysis provides reliable data and a clean policy experiment for identification. Still, it is important to consider possible caveats regarding external validity. Although our analysis focuses on mortality, health effects are unlikely to be restricted to fatalities alone. Higher disease mortality is caused by a general increase in disease prevalence among the population. Increased morbidity in an infant age will reduce general well-being and can often lead to prolonged spells of ill-health and lower life expectancy. We are unable to capture morbidity effects with our analysis due to a lack of data, but they most likely exist. Quantifying effects on disease prevalence and morbidity remains a task for future research.

Another limitation of our model is that we implicitly assume that physicians are homogeneous. In practice, physicians are heterogeneous with regard to skills and specialization. Employing a local average treatment effect (Angrist and Imbens 1995) interpretation of our results, we estimate the effect of changes in the physician coverage induced by emigration. Considering that physicians and the effect on mortality may be heterogeneous, our local estimate may not necessarily coincide with the population average treatment effect if the physicians who emigrate are not a random draw from the skill distribution. If only the better performing physicians emigrate, we may overestimate the effect to some degree, although this would require substantial heterogeneity in very basic skills among physicians. Moreover, if such heterogeneity exists, there is no a priori reason why it would be correlated with religious faith or political convictions.

It would be very interesting to further investigate the role of different physician specializations, however, the majority of physicians in the early 20th century were general practitioners and not clinical specialists. Our time period predates the excessive specialization of modern medical training. In 1933, 75% of physicians trained in family or internal medicine, only 2.6% specialized in pediatrics and 3.5% in gynaecology. Those that specialized typically worked at university research hospitals. These features preclude a more detailed analysis of medical specializations.

We are hesitant to generalize our findings to the present day context and draw quantitative conclusions from the historical analysis for present-day health policy in developing countries. Medical technology and competing risks have changed substantially over the last century. Increasing the supply of physicians in countries with similar mortality rates today will not have the same effect it did in 1930. Historical data can provide only a minimal benchmark: Since technological progress increases the treatment efficiency of physicians, our estimates may be thought of as a lower bound. In appendix A.3, we provide a detailed discussion of historical developments and highlight differences and parallels between 20th century Germany and developing countries today.[5]

Still, our analysis suggests that establishing and maintaining a level of baseline health care coverage has historically been vital to prevent infant mortality and an important complement to other public health policies. In light of this, it is important to ask what factors are preventing many developing countries from experiencing a similarly drastic mortality transition as did the US, Japan and European countries in the early 20th century.

While we do not want to generalize our estimates quantitatively, the qualitative insights offered by the non-linear shape of the health care production function offer some perspective. The finding that diminishing returns matter for health care provision is generally relevant for policy design. Many countries are experiencing regional shortages of physicians and are considering supply-side regulation or incentive schemes to ensure sufficient regional provision. Our analysis underscores that ensuring basic health care supply is important, as the costs of underprovision and overprovision are not symmetric.

Similarly, it is important to prevent the breakdown of basic service provision resulting from negative supply shocks. In many of the least developed countries, where child mortality and health issues are already grave problems, current physician supply ratios are maintained by humanitarian aid and foreign health professionals. This supply is volatile and resources may be withdrawn due to budget reallocation or conflict hazards. Hospital infrastructure is often lacking or insufficient, exacerbating negative health effects. As a development policy, training more physicians may not be the most cost-effective option to improve public health, when improving sanitation and health-related behavior are cheaper

[5] We document how medical technology has changed, in which dimensions infant mortality and causes of death are comparable, and in which they are not. Interestingly, recommendations provided by development organizations for the treatment of newborns today are remarkably similar to those historically provided by physician organizations in Germany (cf. Frohman and Brook 2006).

(and possibly more effective) alternatives. Nevertheless, our results emphasize the need to ensure sufficient access to basic health care when service provision is low or non-existent.

References

Aakvik, A. and Holmås, T. H. (2006). Access to primary health care and health outcomes: The relationships between GP characteristics and mortality rates. *Journal of Health Economics* 25(6), 1139–1153.

Acemoglu, D. and Johnson, S. (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy* 115(6), 925–985.

Akbulut-Yuksel, M. and Yuksel, M. (2015). The long-term direct and external effects of Jewish expulsions in Nazi Germany. *American Economic Journal: Economic Policy* 7(3), 58–85.

Alsan, M. and Goldin, C. (2018). Watersheds in Child Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880–1920. *Journal of Political Economy* 127(2), 586–638.

Anand, S. and Bärnighausen, T. (2004). Human resources and health outcomes: Cross-country econometric study. *The Lancet* 364(9445), 1603–1609.

Andrews, K., Brouillette, D. and Brouillette, R. (2008). Infant mortality. *In* M. M. Haith and J. B. Benson (eds.), *Encyclopedia of Infant and Early Childhood Development*, San Diego, Academic Press, 343–359.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90(430), 431–442.

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

Baltagi, B. and Li, D. (2002). Series estimation of partially linear panel data models with fixed effects. *Annals of Economics and Finance* 3(1), 103–116.

Beddies, T., Doetz, S. and Kopke, C. (2014). *Jüdische Ärztinnen und Ärzte im Nationalsozialismus – Entrechtung, Vertreibung, Ermordung.* Berlin, Boston, De Gruyter Oldenbourg.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics* 82(1), 50–61.

Bhargava, A., Docquier, F. and Moullan, Y. (2011). Modeling the effects of physician emigration on human development. *Economics & Human Biology* 9(2), 172–183.

Bloom, D. E., Canning, D. and Fink, G. (2014). Disease and development revisited. *Journal of Political Economy* 122(6), 1355–1366.

Böhle, I. (2003). *Private Krankenversicherung (PKV) im Nationalsozialismus – Unternehmens- und sozialgeschichtliche Studie unter besonderer Berücksichtigung der Deutschen Krankenversicherung AG (DKV).* Frankfurt am Main, Mabuse Verlag.

Burbidge, J. B., Magee, L. and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association* 83(401), 123–127.

Cameron, A., Ewen, M., Ross-Degnan, D., Ball, D. and Laing, R. (2009). Medicine prices, availability, and affordability in 36 developing and middle-income countries: A secondary analysis. *The Lancet* 373(9659), 240–249.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics* 90(3), 414–427.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data.* No. 53 in Econometric Society Monograph, New York, Cambridge University Press, 2nd ed.

Card, D., Dobkin, C. and Maestas, N. (2008). The impact of nearly universal insurance coverage on health care utilization: Evidence from medicare. *American Economic Review* 98(5), 2242–2258.

Caselli, G. (2015). Mortality, epidemiological, and health transitions. *In* J. D. Wright (ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, Oxford, Elsevier, 857–862.

Chauvet, L., Gubert, F. and Mesplé-Somps, S. (2013). Aid, remittances, medical brain drain and child mortality: Evidence using inter and intra-country data. *The Journal of Development Studies* 49(6), 801–818.

Cooper, J. K., Heald, K. and Samuels, M. (1977). Affecting the supply of rural physicians. *American Journal of Public Health* 67(8), 756–759.

Cutler, D., Deaton, A. and Lleras-Muney, A. (2006). The determinants of mortality. *Journal of Economic Perspectives* 20(3), 97–120.

Davidson, R. and MacKinnon, J. G. (2010). Wild bootstrap tests for IV regression. *Journal of Business & Economic Statistics* 28(1), 128–144.

Dupas, P. (2011). Health behavior in developing countries. *Annual Review of Economics* 3(1), 425–449.

Dyer, C. (1989). *Standards of Living in the Later Middle Ages: Social Change in England C.1200-1520*. Cambridge Medieval Textbooks, Cambridge University Press.

Farahani, M., Subramanian, S. and Canning, D. (2009). The effect of changes in health sector resources on infant mortality in the short-run and the long-run: A longitudinal econometric analysis. *Social Science & Medicine* 68(11), 1918–1925.

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L. and Colford Jr., J. M. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: A systematic review and meta-analysis. *The Lancet Infectious Diseases* 5(1), 42–52.

Finkelstein, A. and McKnight, R. (2008). What did Medicare do? The initial impact of Medicare on mortality and out of pocket medical spending. *Journal of Public Economics* 92(7), 1644–1668.

Frankenberg, E. (1995). The effects of access to health care on infant mortality in Indonesia. *Health Transition Review* 5(2), 143–163.

Frohman, L. and Brook, S. S. (2006). Prevention, welfare, and citizenship: The war on tuberculosis and infant mortality in Germany, 1900-1930. *Central European History* 39(3), 431–481.

Fruen, M. A. and Cantwell, J. R. (1982). Geographic distribution of physicians: Past trends and future influences. *Inquiry* 19(1), 44–50.

Gehrmann, R. (2012). Infant mortality in Germany in the 19th century. *Comparative Population Studies* 36(4), 839–868.

Goodman-Bacon, A. (2018). Public insurance and mortality: Evidence from medicaid implementation. *Journal of Political Economy* 126(1), 216–262.

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

Gruber, J., Hendren, N. and Townsend, R. M. (2014). The Great Equalizer: Health Care Access and Infant Mortality in Thailand. *American Economic Journal: Applied Economics* 6(1), 91–107.

Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions. *Journal of Econometrics* 30(1), 239–267.

Hertz, E., Hebert, J. R. and Landon, J. (1994). Social and environmental factors and life expectancy, infant mortality, and maternal mortality rates: Results of a cross-national comparison. *Social Science & Medicine* 39(1), 105–114.

Hill, K., Lopez, A. D., Shibuya, K. and Jha, P. (2007). Interim measures for meeting needs for health sector data: Births, deaths, and causes of death. *The Lancet* 370(9600), 1726–1735.

Howitt, P., Darzi, A., Yang, G.-Z., Ashrafian, H., Atun, R., Barlow, J., Blakemore, A., Bull, A. M. J., Car, J., Conteh, L., Cooke, G. S., Ford, N., Gregson, S. A. J., Kerr, K., King, D., Kulendran, M., Malkin, R. A., Majeed, A., Matlin, S., Merrifield, R., Penfold, H. A., Reid, S. D., Smith, P. C., Stevens, M. M., Templeton, M. R., Vincent, C. and Wilson, E. (2012). Technologies for global health. *The Lancet* 380(9840), 507–535.

Jeffreys, D. (2008). *Aspirin: the remarkable story of a wonder drug.* Bloomsbury Publishing USA.

Kim, K. and Moody, P. M. (1992). More resources better health? A cross-national perspective. *Social Science & Medicine* 34(8), 837–842.

Klingenberger, D. (2001). *Die Friedensgrenze zwischen gesetzlicher und privater Krankenversicherung – Ökonomische und metaökonomische Kriterien einer optimierten Aufgabenabgrenzung zwischen Sozial- und Individualversicherung.* Regensburg, Transfer Verlag.

Kröner, H.-P. (1989). Die Emigration deutschsprachiger Mediziner im Nationalsozialismus. *Berichte zur Wissenschaftsgeschichte* 12(4), A1–A44.

Lavy, V., Strauss, J., Thomas, D. and de Vreyer, P. (1996). Quality of health care, survival and health outcomes in Ghana. *Journal of Health Economics* 15(3), 333–357.

Lawn, J. E., Cousens, S., Zupan, J., Team, L. N. S. S. et al. (2005). 4 million neonatal deaths: When? Where? Why? *The Lancet* 365(9462), 891–900.

Lawn, J. E., Mwansa-Kambafwile, J., Horta, B. L., Barros, F. C. and Cousens, S. (2010). Kangaroo mother care to prevent neonatal deaths due to preterm birth complications. *International Journal of Epidemiology* 39, i144–i154.

Leibfried, S. and Tennstedt, F. (1980). *Berufsverbote und Sozialpolitik 1933 – Die Auswirkungen der nationalsozialistischen Machtergreifung auf die Krankenkassenverwaltung und die Kassenärzte.* Bremen, Universität Bremen, 2 ed.

Loudon, I. (1991). On maternal and infant mortality 1900–1960. *Social History of Medicine* 4(1), 29–73.

MacKinnon, J. G. and Magee, L. (1990). Transforming the dependent variable in regression models. *International Economic Review* 31(2), 315–339.

Malkin, R. A. (2007). Design of health care technologies for the developing world. *Annu. Rev. Biomed. Eng.* 9, 567–587.

McGuire, T. G. (2000). Physician agency. *In* A. J. Culyer and J. P. Newhouse (eds.), *Handbook of Health Economics*, Elsevier, Vol. 1, Part A of *Handbook of Health Economics*, chap. 9, 461–536.

Mitchell, B. R. (2013). *International historical statistics.* London, Palgrave Macmillan.

Moser, P., Voena, A. and Waldinger, F. (2014). German Jewish émigrés and US invention. *American Economic Review* 104(10), 3222–3255.

Moss, W., Darmstadt, G. L., Marsh, D. R., Black, R. E. and Santosham, M. (2002). Research priorities for the reduction of perinatal and neonatal morbidity and mortality in developing country communities. *Journal of Perinatology* 22(6), 484–95.

Mullahy, J. (1997). Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *The Review of Economics and Statistics* 79(4), 586–593.

Newhouse, J. P., Williams, A. P., Bennett, B. W. and Schwartz, W. B. (1982). Where have all the doctors gone? *JAMA* 247(17), 2392–2396.

Perry, L. and Malkin, R. (2011). Effectiveness of medical equipment donations to improve health systems: How much medical equipment is broken in the developing world? *Medical & Biological Engineering & Computing* 49(7), 719–722.

Peters, D. H., Garg, A., Bloom, G., Walker, D. G., Brieger, W. R. and Hafizur Rahman, M. (2008). Poverty and access to health care in developing countries. *Annals of the New York Academy of Sciences* 1136(1), 161–171.

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies* 29(2), 231–248.

Preston, S. H. (1980). Causes and consequences of mortality declines in less developed countries during the twentieth century. *In Population and Economic Change in Developing Countries*, University of Chicago Press, 289–360.

Reichs-Medizinal-Kalender (1898–1937). *Reichs-Medizinal-Kalender für Deutschland*. Leipzig, Thieme.

Reichsgesundheitsamt (1926–1945). *Reichsgesundheitsblatt*. Berlin, Reichsverlagsamt.

Reinhardt, U. (1972). A production function for physician services. *The Review of Economics and Statistics* 54(1), 55–66.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica* 56(4), 931–954.

Ruiz-Peláez, J. G., Charpak, N. and Cuervo, L. G. (2004). Kangaroo mother care, an example to follow from developing countries. *British Medical Journal* 329(7475), 1179–1181.

Santow, G. (2001). Mortality, epidemiological, and health transitions. *In* N. J. Smelser and P. B. Baltes (eds.), *International Encyclopedia of the Social & Behavioral Sciences*, Oxford, Pergamon, 10071–10075.

Shahar, S. (1990). *Childhood in the Middle Ages*. Routledge.

Silva, J. M. C. S. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics* 88(4), 641–658.

Stanton, B. F. and Clemens, J. D. (1987). An educational intervention for altering water-sanitation behaviors to reduce childhood diarrhea in urban Bangladesh: II. A randomized trial to assess the impact of the intervention on hygienic behaviors and rates of diarrhea. *American Journal of Epidemiology* 125(2), 292–301.

Statistisches Bundesamt (1951). *Statistisches Jahrbuch für die Bundesrepublik Deutschland 1951*.

Statistisches Reichsamt (1884–1940). *Statistisches Jahrbuch für das Deutsche Reich*.

Steppe, H. (1997). *"...den Kranken zum Troste und dem Judenthum zur Ehre...": zur Geschichte der jüdischen Krankenpflege in Deutschland*. Frankfurt am Main, Mabuse-Verlag.

Steppe, H. (2013). *Krankenpflege im Nationalsozialismus*. Frankfurt am Main, Mabuse-Verlag, 10th ed.

Terza, J. V., Kenkel, D. S., Lin, T.-F. and Sakata, S. (2008). Care-giver advice as a preventive measure for drinking during pregnancy: Zeros, categorical outcome responses, and endogeneity. *Health Economics* 17(1), 41–54.

The World Bank (2016). *World Development Indicators*.

Thurston, N. K. and Libby, A. M. (2002). A production function for physician services revisited. *The Review of Economics and Statistics* 84(1), 184–191.

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany. *Journal of Political Economy* 118(4), 787–831.

Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *The Review of Economic Studies* 79(2), 838–861.

Waldinger, F. (2016). Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge. *Review of Economics and Statistics* 98(5), 811–831.

WHO (2016). Global health observatory. Data retrieved from Global Health Observatory, [15.07.2016],
`http://www.who.int/gho/en/`.

Windmeijer, F. A. G. and Santos Silva, J. M. C. (1997). Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics* 12(3), 281–94.

You, D., Hug, L., Ejdemeyr, S. and Beise, J. (2015). Levels and trends in child mortality. Report 2015. Estimates developed by the UN inter-agency group for child mortality estimation. *Working paper*, Unicef.

Appendix

A.1 Exponential model

The model specified in equations (2) and (3) may also be inadequate because of the non-negative nature of the dependent variable. Although rescaling reduces the long right tail in the original count distribution for many variables, this does not work equally well in all cases (cf. Figure A1). Especially when the original variable has generally low incidence that is spread over a comparatively large reference population, the transformed data still resembles typical count data distributions, but is not discrete anymore.

We propose an alternative model specification to address this issue as a robustness check. In addition to the triangular model outlined in subsection 4.1, we estimate an exponential conditional mean model with multiplicative error structure (e.g. Mullahy 1997, Windmeijer and Santos Silva 1997)

$$\begin{aligned} y_{it} &= \exp(\beta s_{it} + \eta_i + \delta_t + \epsilon_{it}) \\ &= \exp(\beta s_{it} + \eta_i + \delta_t)\nu_{it} \\ &= \mu_{it}\nu_{it} \ . \end{aligned} \tag{10}$$

For $\mathrm{E}[y_{it} \mid s_{it}] = \mu_{it}$, the error term in (10) must have a conditional unit mean, i.e. it must be that $\mathrm{E}[\nu_{it} \mid s_{it}] = 1$. This implies that

$$\mathrm{E}\left[\frac{y_{it} - \mu_{it}}{\mu_{it}} \;\middle|\; s_{it}\right] = 0 \ , \tag{11}$$

which will typically be violated due to endogeneity in the number of physicians s_{it}. Instead, we assume a conditional moment condition based on the orthogonal instrument z_i which satisfies

$$\mathrm{E}\left[\frac{y_{it} - \mu_{it}}{\mu_{it}} \;\middle|\; z_{it}\right] = 0 \ . \tag{12}$$

Using moment conditions based on (12), estimation by GMM is straightforward. Even though non-linear, the model features constant relative effects that are intuitive to interpret as incidence rate ratios. The exponentiated coefficients express the multiplicative change in the dependent variable given a unit increase in the independent variable. Since we only specify one conditional moment, the recursive interpretation of equations (2) and (3) is lost. In fact, this limited-information approach requires less assumptions than two-stage least-squares, as we make no assumptions about the distribution of s_{it} given z_{it} (cf. Cameron and Trivedi 2013). The model is also preferable to the simple method of log-transforming the dependent variable, as it avoids the log-of-zero problem and other issues associated with log-transformations (e.g. Silva and Tenreyro 2006). Like quasi-maximum likelihood models, it is consistent if the mean is correctly specified.

A.2 Tables and Figures

Table A1: Descriptive statistics

	Median	Mean	SD	Min	Max	N	Description
Infant mortality	72.308	75.080	24.076	8.772	229.358	2853	Total yearly municipal mortality rate of children dying before one year of age. Measure per thousand live births in $t-1$.
Inflammatory bowel diseases	3.083	4.529	5.502	0.000	53.640	2853	Yearly municipal mortality rate of children dying from Colitis, Enteritis, diarrhea or other ulceration of intestines before one year of age. Measured per thousand live births in $t-1$.
Premature birth, congenital debility	35.461	37.454	15.233	0.000	99.768	1902	Yearly municipal mortality rate of children dying from congenital debility, malformations or as a consequence of premature birth before one year of age. Measured per thousand live births in $t-1$.
Stillbirths	28.215	29.709	10.917	0.000	80.000	1902	Yearly municipal rate of stillborn children. Measured per thousand births in $t-1$.
Measles	0.000	0.018	0.041	0.000	0.638	2853	Yearly municipal mortality rate due to Measles. Incidence rate measured per thousand of population $t-1$.
Scarlet fever	0.000	0.014	0.028	0.000	0.427	2853	Yearly municipal mortality rate due to Scarlet fever. Incidence rate measured per thousand of population $t-1$.
Diphteria	0.045	0.076	0.108	0.000	1.285	2853	Yearly municipal mortality rate due to Diphteria. Incidence rate measured per thousand of population $t-1$.
Influenza	0.069	0.102	0.114	0.000	1.144	2853	Yearly municipal mortality rate due to Influenza. Incidence rate measured per thousand of population $t-1$.
Bronchitis	0.127	0.153	0.124	0.000	0.966	1902	Yearly municipal mortality rate due to Bronchitis. Incidence rate measured per thousand of population $t-1$.
Pneumonia	0.708	0.746	0.298	0.000	3.119	2853	Yearly municipal mortality rate due to Pneumonia. Incidence rate measured per thousand of population $t-1$.
Typhoid fever	0.000	0.009	0.022	0.000	0.445	2853	Yearly municipal mortality rate due to Typhoid fever. Incidence rate measured per thousand of population $t-1$.
Physicians	0.944	1.048	0.613	0.139	6.034	2853	Registered physicians coverage ratio. Measured per thousand of population in $t-1$.
Physician emigration	0	0.955	15.258	0	622	2853	Emigration of physicians
Jewish physicians	1	12.385	104.575	0	1817	2853	Jewish physicians
Jewish physicians, percent	5.263	7.923	10.634	0.000	87.095	2853	Share of Jewish physicians over all physicians in 1933 in percent.
1933 Jewish population	0.156	1.292	9.303	0.001	160.564	2853	Jewish population in 1933 in thousands.
1933 Jewish population, percent	0.451	0.629	0.643	0.004	4.713	2853	Share of Jewish population based on 1933 Jewish and total population in percent.
Population	32.004	91.598	270.226	15.192	4339.641	2853	Total municipal population in thousands.

Notes: All statistics are based on the largest estimation sample covering the years 1928–1936. Sources: *Reichsgesundheitsblatt* [Health bulletin] (eds. 1928–1936), Reichsgesundheitsamt [Federal health ministry]. *Volks-, Berufs- und Betriebszählung* [Census] 1933, Statistisches Reichsamt [Federal statistical office]. *Reichsmedizinalkalender – Verzeichnis der deutschen Ärzte und Heilanstalten* [Register of German physicians and hospitals] (eds. 1928-1936), Thieme Verlag.

Table A2: Pre-treatment covariate balance statistics

	Jewish pshare$_{1933} < Q_2$	Jewish pshare$_{1933} \geq Q_2$	Difference
	(1)	(2)	(2) - (1)
(a) Mortality outcomes			
Infant mortality	70.822	71.345	0.523
	(22.284)	(21.645)	(2.467)
Infl. bowel diseases	3.945	4.050	0.105
	(4.861)	(4.191)	(0.509)
Premat. birth/congen. debility	32.948	32.899	−0.049
	(15.903)	(13.154)	(1.636)
Stillbirths	30.846	30.027	−0.819
	(12.756)	(10.505)	(1.310)
Infant mortality/1,000 pop.	1.087	1.080	−0.007
	(0.495)	(0.471)	(0.054)
Measles	0.010	0.016	0.006
	(0.024)	(0.040)	(0.004)
Scarlet fever	0.007	0.009	0.001
	(0.028)	(0.018)	(0.003)
Diphteria	0.064	0.061	−0.003
	(0.108)	(0.075)	(0.010)
Influenza	0.107	0.117	0.010
	(0.105)	(0.101)	(0.012)
Bronchitis	0.172	0.172	−0.001
	(0.148)	(0.100)	(0.014)
Pneumonia	0.653	0.742	0.088***
	(0.240)	(0.243)	(0.027)
Typhoid fever	0.006	0.007	0.001
	(0.020)	(0.015)	(0.002)
(b) Municipality characteristics: Vote shares			
Vote share: NSDAP	0.311	0.319	0.008
	(0.083)	(0.090)	(0.013)
Vote share: SPD	0.248	0.206	−0.042***
	(0.092)	(0.082)	(0.013)
Vote share: KPD	0.171	0.165	−0.006
	(0.074)	(0.069)	(0.010)
Vote share: Zentrum	0.095	0.153	0.058***
	(0.130)	(0.150)	(0.020)
Vote share: DNVP	0.097	0.088	−0.008
	(0.057)	(0.041)	(0.007)
Vote share: DVP	0.027	0.025	−0.002
	(0.017)	(0.017)	(0.002)
(c) Municipality characteristics: Population and Employment			
Population (ln)	10.378	10.961	0.583***
	(0.637)	(1.113)	(0.102)
Population growth	0.616	0.691	0.076
	(1.477)	(1.656)	(0.176)
Labor force participation	0.764	0.771	0.006
	(0.073)	(0.058)	(0.007)
Unemployment rate	0.148	0.139	−0.009
	(0.046)	(0.043)	(0.006)
Social assistance	0.040	0.042	0.001
	(0.020)	(0.020)	(0.003)

Notes: All statistics are based on a 1931 cross-section of municipalities prior to treatment. Vote shares are based on the general election November 1932. Sources: *Reichsgesundheitsblatt* [Health bulletin] (eds. 1928–1936), Reichsgesundheitsamt [Federal health ministry]. *Volks-, Berufs- und Betriebszählung* [Census], Statistisches Reichsamt [Federal statistical office].

Table A3: Robustness: Population dynamics

	(1)	(2)	(3)	(4)
	Infant mortality			
	Total	Infl. bowel diseases	Prem. birth, cong. debility	Stillbirths
	OLS			
# of registered physicians	−1.24	−0.62	0.81	−3.00
	(3.63)	(0.93)	(3.66)	(2.24)
ln(population) ($t-1$)	−16.57	−3.24	27.19**	6.46
	(11.67)	(2.69)	(13.15)	(8.59)
Population growth (%, $t-1$)	0.82***	0.09***	0.10	0.24
	(0.13)	(0.04)	(0.23)	(0.15)
	IV: Emigration			
# of registered physicians	−18.34***	−6.75***	−4.26	−8.85***
	(3.54)	(0.99)	(5.02)	(2.22)
ln(population) ($t-1$)	−26.25**	−6.71**	23.33*	2.00
	(11.18)	(2.75)	(13.33)	(7.77)
Population growth (%, $t-1$)	0.93***	0.13***	0.14	0.29**
	(0.13)	(0.04)	(0.23)	(0.14)
First stage F-stat.	32.62	32.62	39.70	39.70
	IV: Jewish physicians in 1933			
# of registered physicians	−12.58**	−3.84***	−2.42	−6.20***
	(5.63)	(1.11)	(4.04)	(1.65)
ln(population) ($t-1$)	−22.99**	−5.06*	24.72*	4.02
	(11.39)	(2.67)	(12.75)	(7.67)
Population growth (%, $t-1$)	0.89***	0.11***	0.13	0.27*
	(0.13)	(0.04)	(0.22)	(0.14)
First stage F-stat.	16.77	16.77	18.97	18.97
	IV: Jewish population in 1933			
# of registered physicians	−12.40**	−4.10***	−1.91	−6.96***
	(5.51)	(1.08)	(4.25)	(1.78)
ln(population) ($t-1$)	−22.88**	−5.21*	25.12**	3.44
	(11.39)	(2.67)	(12.71)	(7.66)
Population growth (%, $t-1$)	0.89***	0.11***	0.12	0.27**
	(0.13)	(0.04)	(0.22)	(0.14)
First stage F-stat.	16.08	16.08	18.02	18.02
Year fixed effects	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓
N municipalities	317	317	317	317
N	2853	2853	1902	1902

Notes: Infant mortality variables are measured per thousand live births in $t-1$, stillbirths per thousand total births in $t-1$. Registered physicians are measured per thousand of population in $t-1$. Included instruments are a full set of year and municipality dummies. Excluded instruments are given in the respective paragraph header if applicable. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

Table A4: Robustness: Including state x year fixed effects

	Infant mortality			
	Total	Infl. bowel diseases	Prem. birth, cong. debility	Stillbirths
	IV: Emigration			
# of registered physicians	−20.63***	−7.49***	−2.94	−7.95***
	(4.28)	(1.02)	(5.78)	(2.24)
	IV: Jewish physicians in 1933			
# of registered physicians	−16.01**	−4.53***	−1.88	−5.26***
	(6.31)	(1.22)	(4.05)	(1.82)
	IV: Jewish population in 1933			
# of registered physicians	−15.29**	−4.77***	−1.32	−6.12***
	(6.22)	(1.18)	(4.24)	(1.86)
State x year fixed effects	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓
N municipalities	317	317	317	317
N	2853	2853	1902	1902

Notes: Infant mortality variables are measured per thousand live births in $t-1$, stillbirths per thousand total births in $t-1$. Registered physicians are measured per thousand of population in $t-1$. Included instruments are a full set of year effects and linear time trends on the state-/province-/district-level. Excluded instruments are given in the respective paragraph header. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

Table A5: Robustness: Including linear regional trends

(a) State-level linear trends

	Infant mortality			
	Total	Infl. bowel diseases	Prem. birth, cong. debility	Stillbirths
	IV: Emigration			
# of registered physicians	−20.36***	−7.67***	−2.71	−8.31***
	(4.06)	(1.05)	(5.50)	(2.17)
	IV: Jewish physicians in 1933			
# of registered physicians	−15.44**	−4.64***	−1.19	−6.01***
	(6.36)	(1.23)	(4.40)	(1.75)
	IV: Jewish population in 1933			
# of registered physicians	−14.84**	−4.91***	−0.75	−6.83***
	(6.26)	(1.18)	(4.55)	(1.80)
State-level linear trend	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓
N municipalities	317	317	317	317
N	2853	2853	1902	1902

(b) Province-level linear trends

	Infant mortality			
	Total	Infl. bowel diseases	Prem. birth, cong. debility	Stillbirths
	IV: Emigration			
# of registered physicians	−18.71**	−10.50***	−1.11	−10.47***
	(7.84)	(2.02)	(8.97)	(2.91)
	IV: Jewish physicians in 1933			
# of registered physicians	−10.64	−5.11***	2.37	−7.55***
	(9.33)	(1.82)	(7.02)	(2.46)
	IV: Jewish population in 1933			
# of registered physicians	−7.69	−4.93***	3.48	−8.49***
	(9.44)	(1.87)	(7.20)	(2.62)
Province-level linear trend	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓
N municipalities	317	317	317	317
N	2853	2853	1902	1902

(c) Vote district-level linear trends

	Infant mortality			
	Total	Infl. bowel diseases	Prem. birth, cong. debility	Stillbirths
	IV: Emigration			
# of registered physicians	−18.46**	−9.77***	−1.11	−10.39***
	(7.64)	(2.22)	(8.85)	(2.81)
	IV: Jewish physicians in 1933			
# of registered physicians	−12.30*	−4.65**	2.13	−8.49***
	(7.43)	(1.93)	(6.68)	(1.87)
	IV: Jewish population in 1933			
# of registered physicians	−8.94	−4.56**	3.53	−9.01***
	(8.02)	(2.06)	(7.07)	(2.21)
District-level linear trend	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓
N municipalities	317	317	317	317
N	2853	2853	1902	1902

Notes: Infant mortality variables are measured per thousand live births in $t-1$, stillbirths per thousand total births in $t-1$. Registered physicians are measured per thousand of population in $t-1$. Included instruments are a full set of year effects and linear time trends on the state-/province-/district-level. Excluded instruments are given in the respective paragraph header if applicable. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

Table A6: Robustness: IV placebo checks

(a) Predicting past outcomes with future IV

	Infant mortality in 1929		
Emigration	−0.015		
	(0.011)		
Jewish population		0.000	
		(0.000)	
Jewish physicians			−0.003
			(0.003)

(b) Placebo instrument: Main infant mortality outcomes
(IV: Psychiatrist emigration)

	Infant mortality	Infl. bowel	Prem. birth
# of registered physicians	6.869	0.604	8.420
	(14.771)	(2.322)	(14.553)

(c) Placebo outcome: Non-mitigable diseases
(IV: Emigration of physicians)

	Stroke	Old age	Unknown
# of registered physicians	−0.219	−0.020	−0.063
	(0.203)	(0.134)	(0.118)

(d) Placebo outcome: Non-mitigable diseases
(IV: Jewish population)

	Stroke	Old age	Unknown
# of registered physicians	−0.406	0.058	−0.002
	(0.274)	(0.124)	(0.095)

(e) Placebo outcome: Non-mitigable diseases
(IV: Jewish physicians)

	Stroke	Old age	Unknown
# of registered physicians	−0.367	0.049	−0.002
	(0.258)	(0.121)	(0.090)

Notes: Results in panel (a) are based on a linear regression of cross-sectional infant mortality in 1929, pre-dating the occupational restrictions, on the instruments in 1934, post-dating the restrictions. Panel (b) repeats the main specification from Table 2 for all infant mortality outcomes, using an indicator for emigration of psychiatrists instead of emigration of physicians as the instrument. Panels (c) to (e) repeat the main specification for each instrument for outcomes which are unlikely to be affected by changes in physicians induced by the instrument. Infant mortality is measured per thousand live births in $t-1$, incidence for other conditions per thousand of population in $t-1$. Registered physicians are measured per thousand of population in $t-1$. In panels (e) to (e), included instruments are a full set of year and municipality dummies. Excluded instruments are given in the respective paragraph header if applicable. Standard errors clustered at the municipality level given in parentheses.

Table A7: Robustness: Exponential model mortality estimates

	(1)	(2)	(3)	(4)
	Infant mortality			
	Total	Infl. bowel diseases	Prem. birth, cong. debility	Stillbirths
	Poisson			
# of registered physicians	1.02	0.83	0.96	0.90
	(0.05)	(0.17)	(0.10)	(0.07)
	IV: Emigration			
# of registered physicians	0.77***	0.32***	0.92	0.78***
	(0.04)	(0.09)	(0.12)	(0.06)
	IV: Jewish physicians in 1933			
# of registered physicians	0.83**	0.39***	1.01	0.84***
	(0.06)	(0.10)	(0.13)	(0.05)
	IV: Jewish population in 1933			
# of registered physicians	0.84**	0.38***	1.03	0.82***
	(0.06)	(0.10)	(0.14)	(0.05)
Year fixed effects	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓
N municipalities	317	317	317	317
N	2853	2853	1902	1902

Notes: Results from exponential model instrumental variable estimation by GMM. Estimates are reported as incidence rate ratios, indicating the multiplicative change in the dependent variable given a unit increase in the physician coverage ratio. The null hypothesis in all tests is that the coefficient takes value one. Infant mortality variables are measured per thousand live births in $t-1$, stillbirths per thousand total births in $t-1$. Registered physicians are measured per thousand of population in $t-1$. Included instruments are a full set of year and municipality dummies. Excluded instruments are given in the respective paragraph header if applicable. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

Table A8: Robustness: Linear model using log-transformed infant mortality counts

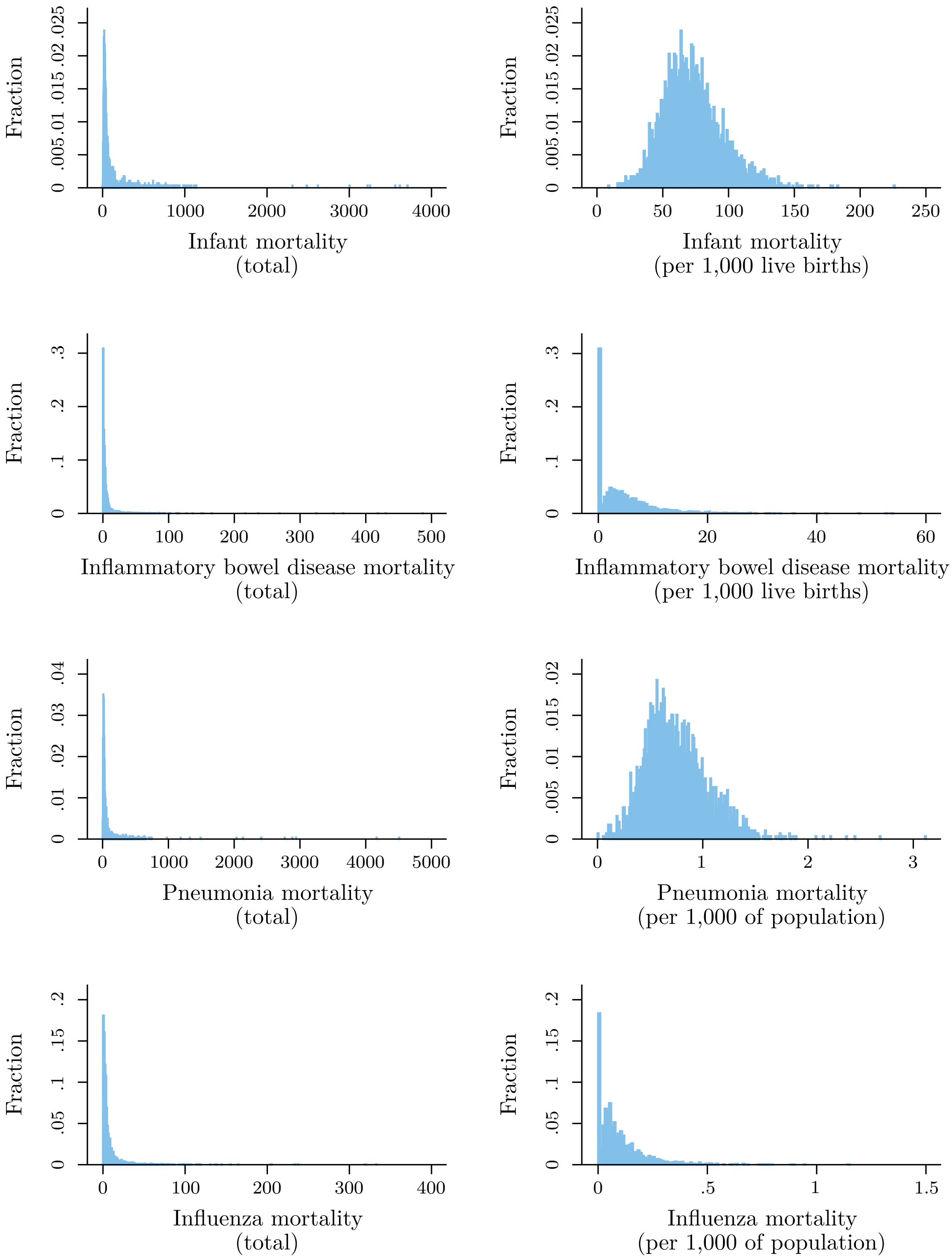
	(1)	(2)	(3)	(4)
	Log total infant mortality			
	IV: Emigration			
# of registered physicians	−0.47***	−0.51***	−0.37***	−0.38***
	(0.08)	(0.08)	(0.06)	(0.06)
First stage F-stat.	29.71	32.77	27.24	31.35
	IV: Jewish physicians in 1933			
# of registered physicians	−0.37***	−0.42***	−0.31***	−0.33***
	(0.09)	(0.10)	(0.08)	(0.08)
First stage F-stat.	15.61	16.83	15.30	17.10
	IV: Jewish population in 1933			
# of registered physicians	−0.34***	−0.39***	−0.26***	−0.28***
	(0.11)	(0.11)	(0.09)	(0.09)
First stage F-stat.	14.89	16.11	14.03	15.76
Control: ln population		✓		✓
Control: ln live births			✓	✓
Year fixed effects	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓
N municipalities	317	317	317	317
N	2853	2853	2853	2853

Notes: Dependent variable is log total infant mortality. Registered physicians are measured per thousand of population in $t-1$. Included instruments are a full set of year and municipality dummies. Excluded instruments are given in the respective paragraph header if applicable. Standard errors clustered at the municipality level given in parentheses. *, **, *** denote significance at the 0.1, 0.05, 0.01 level respectively.

Figure A1: Comparison of count and reference-scaled distributions

(a) Raw counts

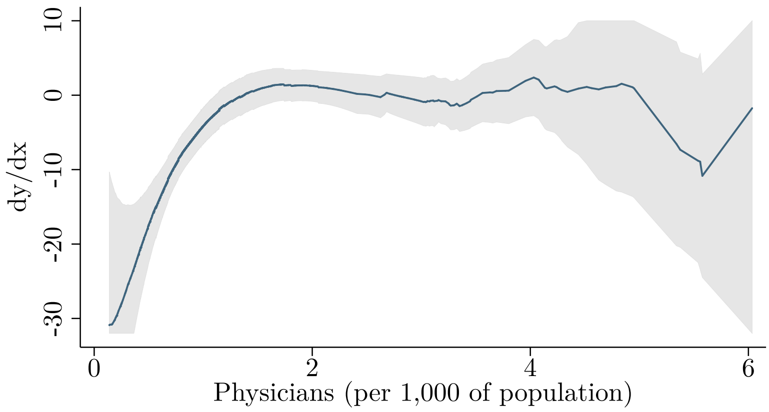
(b) Per reference population



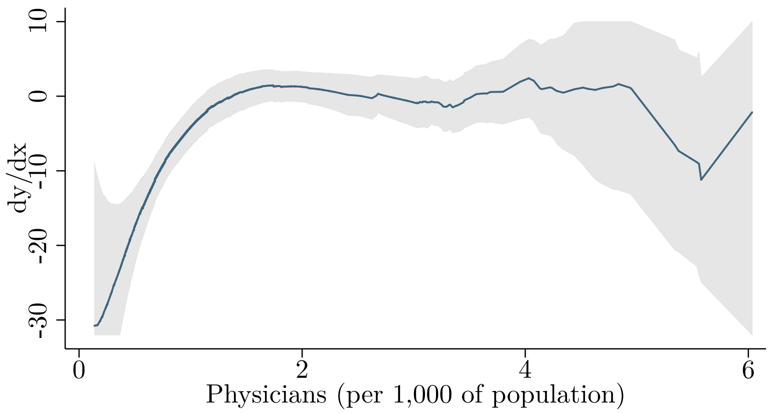
Notes: Column (a) shows the distribution of the original unscaled counts, column (b) the distribution of the same variables scaled per reference population. Each row depicts the same base variable. All plots are based on the (largest) main estimation sample.

Figure A2: Robustness: Control function and semiparametric analysis

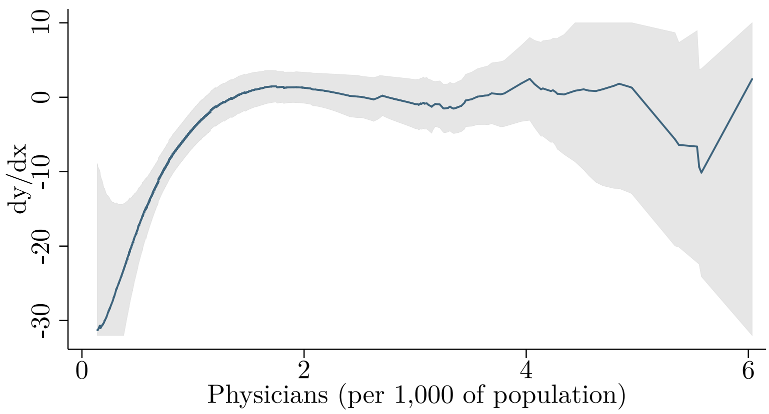
(a) Marginal effect (control function: linear)



(b) Marginal effect (control function: second order polynomial)



(c) Marginal effect (control function: third order polynomial)



Notes: Figures plot the marginal effect of physician density on infant mortality. Shaded in grey are the 95% confidence intervals. Panel (a) is based on an specification including only the linear control function term, panels (b) and (c) on a specification including a second/third order polynomial of the control function.

A.3 Infant mortality in historical context

This section draws a comparison between developing countries today and Germany in the first half of the 20th century. We focus on infant mortality rates, causes of death and the disease environment, health care supply and treatment technology. We show that there are both pronounced differences and important similarities within these dimensions. We do not want to argue that our results generalize to present-day developing countries. Instead, we want to be clear about the dimensions in which the situation in these countries differs from that during the time period we study.

Since the 1950s, infant mortality in many low and middle income countries has decreased substantially. Infant mortality in South Asia, the Middle East and Sub-Saharan Africa has decreased from around 160 infant deaths per 1,000 live births in the 1960s to between 40 and 60 in 2015 (see Figure A3). These changes resemble the development of infant mortality rates in industrialized economies a century ago (e.g. Cutler et al. 2006). In Germany, 207 infant deaths per 1,000 live births were registered in 1990. By 1933, this number had fallen markedly to 77 cases (cf. Figure A3).

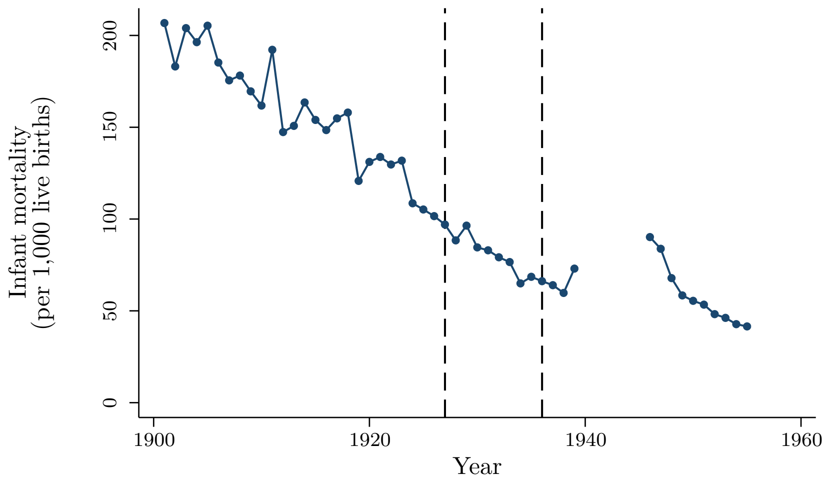
Both in developing countries today as well as in 20th Germany, this decrease was mainly driven by a reduction in mortality of infants which had survived the first month. Consequently, the share of neonates in infant deaths has increased by around 10% in middle and low-income countries since 1990 (The World Bank 2016).[6] Similarly, in Germany this share rose from 32% in 1890 to 52% in 1932 (Statistisches Bundesamt 1951).

Similarities are not only present with respect to infant morality rates but extend to the state of health care supply. In 1930, Germany's physician coverage ratio was approximately equal to 0.75 physicians per 1,000 population. This supply density is comparable to many developing countries today. Figure A4 displays physician coverage ratios for major world regions over the last 20 years. The lowest coverage ratio can be observed in Sub-Saharan Africa with less than 0.3 physicians per 1,000 population. Average coverage ratios in the Middle East & Northern Africa as well as South Asia are closer to our sample average of

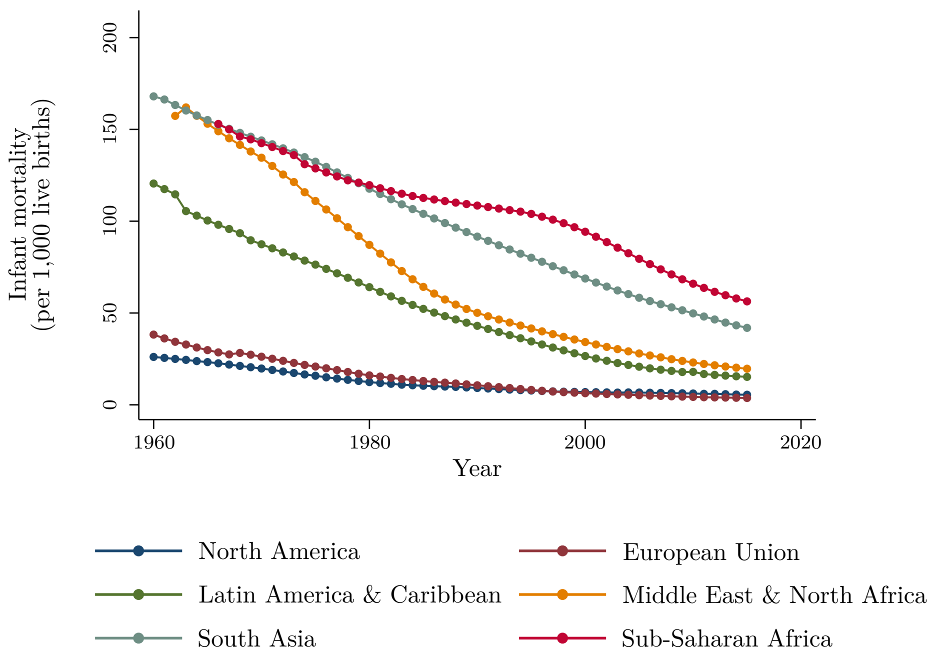
[6]Neonatal mortality refers to infants dying within the first 28 days of life (Andrews et al. 2008).

Figure A3: Infant mortality: Historical comparison

(a) Germany, 1900–1955



(b) Selected world regions, 1960–2015



Notes: Panel (a) depicts historical infant mortality per 1,000 live births in Germany for the years 1900–1955, panel (b) shows infant mortality estimates per 1,000 live births for selected world regions over the time frame 1960–2015. Source: Statistisches Reichsamt (1900–1940), Statistisches Bundesamt (1945–1955) , World Development Indicators (2016).

one physician per 1,000 of population.[7]

Similar coverage ratios can be observed in many richer African or poorer Asian or Latin American countries, e.g. Bolivia, Pakistan or regions of India. However, these countries typically feature somewhat lower mortality rates. On the other hand, countries which have comparable mortality ratios today (e.g. those in Sub-Saharan Africa), typically also have lower physician supply ratios compared to Germany in 1930. This may reflect in part increases in medical technology which make treatments more effective.

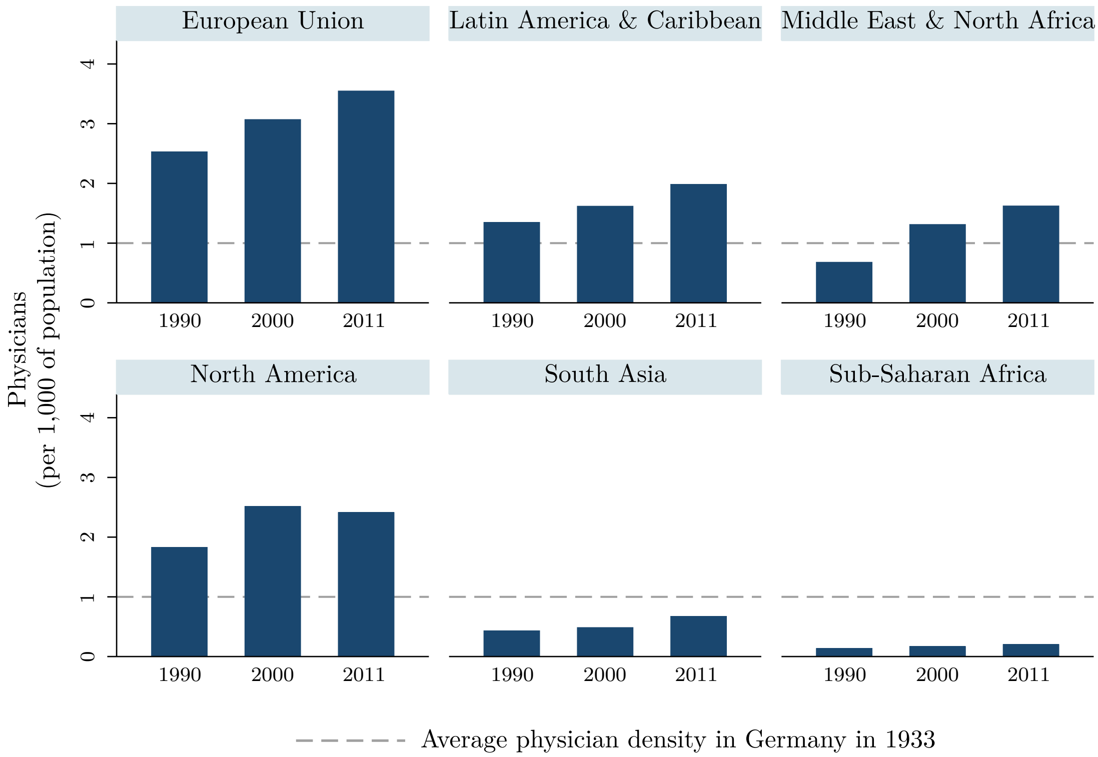
Important parallels also exist regarding the main causes of infant mortality. Neonatal deaths in developing countries today are usually connected to inadequate access to basic medical care at and immediately following birth; the leading causes of neonatal death being infections (36%) such as sepsis, pneumonia, tetanus and diarrhoea, complications surrounding birth (27%) and asphyxia (23%).[8] Low birth weight is often a contributing factor (Lawn et al. 2005, Andrews et al. 2008). Postneonatal mortality can be attributed to malnutrition, infectious diseases and home environment (Andrews et al. 2008). These causes are similar to those common in developed countries in the early 20th century. The majority of infant deaths then were attributable to deficient pre- and post-natal care, infections and water- and food borne diseases, most commonly respiratory infections and gastrointestinal illnesses (Reichsgesundheitsamt 1926–1945, Cutler et al. 2006). Much like today, these diseases disproportionately affected the poor due to their living and working conditions and insufficient nutrition (Frohman and Brook 2006).

While infant mortality rates, causes of death and physician supply ratios are comparable to a certain degree between developing regions today and developed countries a century ago, the quality of health care has changed. Medical technology has advanced substantially since 1930. Still, a limiting factor is that medical technology diffusion from developed to developing countries is limited. Figure A5 shows prevalence rates of selected medical technologies for which comparable data is available (magnetic resonance imaging and

[7]Our sample average is slightly higher than physician density in the whole of Germany as our sample is selected on more populous areas where physician coverage ratios are generally higher.

[8]Around two-thirds of all births in developing countries occur at home and skilled-care is only available in about half of all cases. Postpartum visitation for the newborn is uncommon (Moss et al. 2002).

Figure A4: Physician density across world regions



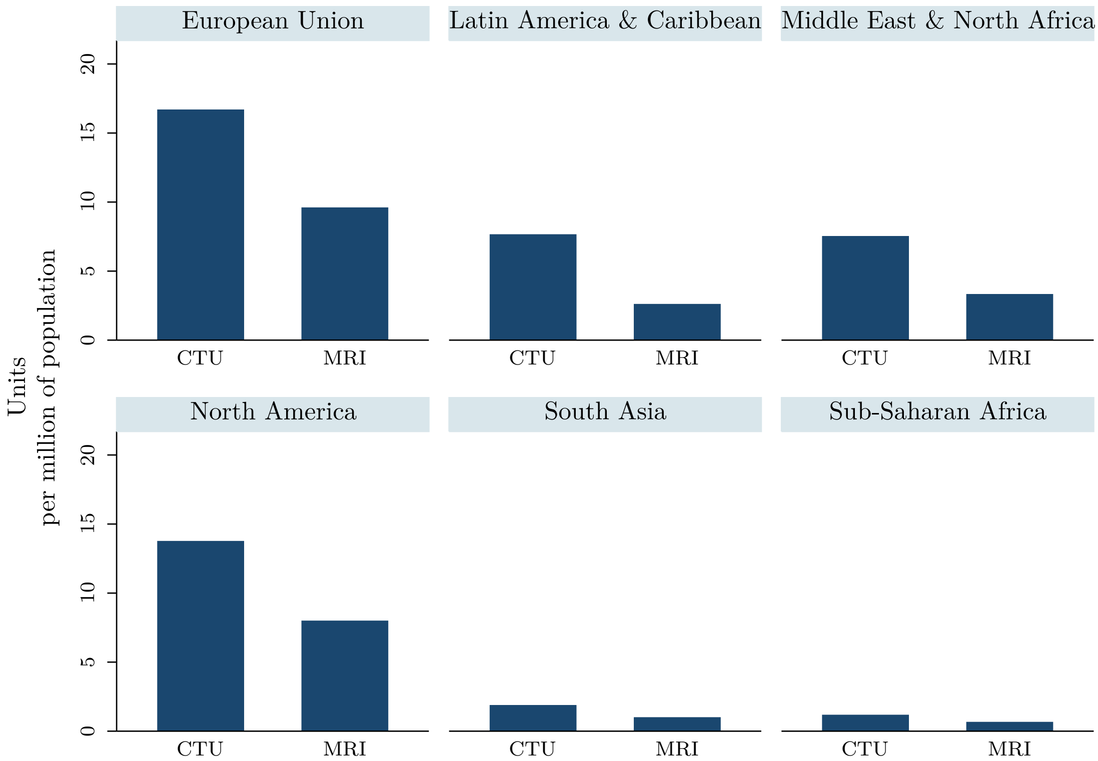
Notes: Bars depict the average physician density in a region for each available year. The horizontal line is the reference density in Germany in 1933. Source: World Development Indicators (2016).

computed tomography units) across world regions.[9] Especially in countries in South Asia and Sub-Saharan Africa, access to advanced medical technology is very limited. This lack of access is particularly pronounced for rural and poor population segments as medical technology in developing countries is usually concentrated in cities and private hospitals (Malkin 2007, Peters et al. 2008). Moreover, Perry and Malkin (2011) estimate that around 40% of health-care equipment in developing countries is out of service compared to less than 1% in developed countries.[10] Access problems also extend to drugs, many of which are not available in low and middle income countries, especially in the public health sector. Even if certain medicines are available in the private sector, their price often substantially exceeds the international reference price which renders them prohibitively costly to large parts of the population (Cameron et al. 2009).

[9] In the context of infant mortality, data on ultrasound and incubator prevalence would be preferable. However, such data is unavailable. The available evidence suggests that supply is equally poor (Lawn et al. 2010, Ruiz-Peláez et al. 2004).

[10] The lack of a reliable energy supply also constitutes a major hindrance to the employment of medical technology in developing countries. This is especially relevant for therapeutic devices such as neonatal incubators which need to be powered constantly. It is also of consequence for drugs and vaccines who need to be stored at low temperature to remain viable (Howitt et al. 2012).

Figure A5: Medical technology across world regions in 2013



Notes: Bars depict the average prevalence of computed tomography (CTU) and magnetic resonance imaging (MRI) units in a region. Source: World Health Organization, WHO (2016).

Another consideration is that modern medical technology is of limited relevance for addressing the main causes of infant mortality (e.g. Cutler et al. 2006, Andrews et al. 2008). Simple treatment measures are often sufficient to address many health issues. For example, gastrointestinal diseases are typically treated by ensuring sufficient rest, fluids, nutrition and possibly drugs like Acetaminophen to reduce fever and pain. In the case of birth asphyxia, resuscitation by tactile stimulation or the clearing of upper airway secretions using a covered finger or oral mucus trap is normally sufficient. A need for external ventilation is only given in exceptional circumstances (Moss et al. 2002).

Early diagnosis, disease prevention and health practices are relatively more important than treatment after a health problem has fully developed. Physicians play an important role in disease prevention by ensuring health and sanitary practices. Especially for perinatal deaths, pre- and post-natal care practices are a crucial factor (Cutler et al. 2006). For example, proper umbilical core care by physicians using antibacterial agents after birth has been shown to reduce infection of the cord and neonatal sepsis (Moss et al. 2002).

Late-onset sepsis can be prevented by ensuring a clean caregiving environment. Similarly, a frequent problem in developing countries is hypothermia which affects more than half of all newborns and is associated with increased risk of neonatal infections, acidosis, coagulation defects, respiratory distress syndrome and brain haemorrhage. Neonatal incubators are only needed in extreme cases and hypothermia can generally be prevented by simple measures such as ensuring a warm environment during delivery, early breastfeeding, proper bathing, drying/swaddling and skin-to-skin contact with the mother (Moss et al. 2002).

The importance of health practices is mirrored in modern development economics (e.g. Dupas 2011). A large body of research focuses on how physicians can influence their patients' well-being by encouraging health-related behavior and compliance with hygienic standards (cf. Stanton and Clemens 1987, Fewtrell et al. 2005, Terza et al. 2008). Similarly, propagation of sanitary practices was also the leading sentiment in health policy in the early 20th century (Frohman and Brook 2006). The germ theory of disease was already well established at the time and the importance of a sanitary environment for the prevention of diseases was understood. In fact, many practices recommended by development organizations today are similar to those of physician organizations in Germany during the early 20th century. Examples are the recommendation of exclusive and immediate breastfeeding, in particular for low birth weight newborns and the dissemination of sanitary procedures such as sterilization of water and milk, among others (Moss et al. 2002, Frohman and Brook 2006). In the Weimar Republic propagation of such practices was primarily carried out through infant welfare centers staffed with physicians who provided free medical examinations to both mothers and their infants. Similar to development research today, policy debates focused on how to establish physicians as a recognized authority and improve compliance with their recommendations (Frohman and Brook 2006).

One field where technological progress has been very influential are vaccinations. Although many vaccines were invented during the first half of the 20th century (e.g. Rabies, Plague, Diphteria, Pertussis, Tuberculosis, Tetanus and Yellow Fever) vaccination rates were still relatively low. Immunization rates increased sharply after 1960, and vaccinations for many important childhood diseases were invented afterwards (e.g. Polio, Measles, Mumps, Rubella and Hepatitis B). However, even though morbidity consequences from these diseases are high, historical data suggests that direct mortality due to them was rare even prior to the availability of vaccines (e.g. Cutler et al. 2006). The exception to this is Tuberculosis, which we exclude from the empirical analysis. To illustrate, the measles mortality rate in Germany in 1930 amounted to about 2 cases per 100,000 of population. This figure is comparatively low compared to mortality from Pneumonia (approx. 80 cases per 100,000 of population) or total infant mortality (7,500 per 100,000 live births). Furthermore, even though immunization rates have increased steadily in developing countries over the last 20 years, vaccination is still far from universal especially among poorer population segments (WHO 2016). While vaccines for measles and diphteria have become more widespread, for many of other diseases we consider in the analysis, this is not the case. Vaccines are either unavailable (gastrointestinal diseases, bronchitis, scarlet fever) or uncommon and do not offer long-term protection (influenza, pneumonia, typhoid fever). While vaccines matter for reductions in *child* mortality, they matter less so for *infant* mortality, where most deaths occur in the first month after birth.

Another important development was the discovery of antibiotics. Although penicillin was discovered in 1928, antibiotics were not commercially available for civilians before 1945. Similarly, the first sulfonamide drug Prontosil was first developed in 1935 but only gained widespread use during the 1940s. A limitation is that antibiotics and antimicrobial drugs are only effective against bacterial diseases, they are ineffective against viral diseases such as influenza or bronchitis. Other commonly prescribed drugs were already available and prescribed by physicians in 1930. Simple nonsteroidal anti-inflammatory drugs (e.g. acetylsalicylic acid, introduced as Aspirin by Bayer in 1899) and common pain and fever medications (e.g. Phenacetin, which metabolizes to Paracetamol (Acetaminophen), introduced in 1887) were readily sold in pharmacies (Jeffreys 2008).

In the paper, we analyze historical data from Germany, one of today's developed economies, and document the diminishing returns to health care provision during the later stages of the mortality transition. We do not want to argue that the historical context of our study provides a control case for present day developing countries. However, we provide this comparison to be clear about the dimensions in which the historical situation in today's developed countries compares to the situation in other developing countries, and those in which it does not. Undisputably, medical technology has advanced since the period of our study.