Essays in Empirical Health Economics

Peter Roman Redler

2025

Essays in Empirical Health Economics

Inaugural-Dissertation
zur Erlangung des Grades

Doctor oeconomiae publicae
(Dr. oec. publ.)

an der
Ludwig-Maximilians-Universität München

2025
vorgelegt von

Peter Roman Redler

Referent:	Prof. Dr. Joachim Winter
Korreferentin:	Prof. Dr. Amelie Wuppermann
Promotionsabschlussberatung:	29. Januar 2025

Für Carla, Philipp und Jonas

Acknowledgements

As I reflect on the journey that led to this dissertation, I am thankful for those who have supported and guided me.

I am profoundly grateful to my doctoral advisor, Joachim Winter, for his invaluable expertise, steadfast support, and insightful guidance. His trust in allowing me to freely explore research ideas, combined with the resources provided, has been crucial throughout my doctoral journey.

My gratitude also goes to Amelie Wuppermann, who ignited my interest in health economics during my master's studies and has been a steady source of support throughout this journey. I am also thankful to Florian Englmaier for serving as the third member of my dissertation committee.

To my co-authors, I extend my sincere gratitude. I am especially appreciative of Friederike, whose brilliance and creativity have been truly inspiring, and Corinna, whose professionalism and approach to research set a standard to which I have aspired throughout my PhD.

I have been fortunate to do research and teach alongside wonderful colleagues, many of whom have become dear friends. Sharing lunch and coffee breaks with Silvia, Lorenz, Eleonora, Fabian, Lena, Sebastian, Robin, Bernhard, Christina, Sebi, Maria, and Kevin added moments of joy and connection to my experience.

I am immensely thankful to my parents, who have nurtured my passion for discovery and learning and supported me every step of the way throughout my academic endeavours. I would also like to thank my sons Philipp and Jonas, who were both born during this PhD journey: Despite the new challenges you brought, the joy and meaning you have added to my life is beyond measure. Finally, I am deeply grateful to my wife, Carla, for her faith in me, and her unwavering support and love.

Peter Roman Redler

Munich

September 2024

Contents

Chapter 1

Introduction

The core predicament of medicine - the thing that makes being a patient so wrenching, being a doctor so difficult, and being a part of society that pays the bills they run up so vexing - is uncertainty.
Complications: A surgeon's notes on an imperfect science
Atul Gawande (2002)

Uncertainty pervades every aspect of healthcare, from individual patient outcomes to medical decision-making and the broader impacts of health policies. This uncertainty is compounded by the immense complexity of human biology, individual decisions, healthcare systems, and other societal factors that influence health. Health economics provides tools and frameworks to analyze and interpret this complex and uncertain landscape.

Arrow (1963)'s paper *Uncertainty and the Welfare Economics of Medical Care* laid the theoretical groundwork for the field of health economics. This work emphasized that the unique characteristics of healthcare markets, particularly information asymmetry and uncertainty, necessitate a distinct economic approach. Since then, health economics has evolved to address a wide array of issues, from individual health behaviors to global health policies.

The evolution of health economics is evident in the diverse topics and methodologies that researchers have explored over the decades. The following studies illustrate how the field has tackled complex issues, from individual behaviors to societal impacts. Grossman and Jacobowitz (1981) analyzed the effects of public policies on infant mortality rates using multiple regression analysis, demonstrating how economic evaluations can guide population-level health interventions. At the level of physician decision-making, Kessler and McClellan (1996) examined defensive medicine using regional variations in malpractice laws as a natural experiment, revealing how legal and financial incentives influence medical care. Gruber and Frakes (2006) investigated the relationship between smoking cessation and obesity with a two-way fixed effects model, showcasing the field's ability to uncover unexpected links between

health behaviors. Expanding to a broader societal scope, Currie et al. (2015) explored the impact of environmental health risks on housing values using a difference-in-differences approach. More recently, Miller et al. (2023) assessed the economic consequences of being denied abortion access with an event study model, highlighting the role of health economics in informing contentious policy debates.

Building on this rich tradition of addressing complex issues related to health through an economic lens, this dissertation contributes to the diverse landscape of health economics, presenting four papers that explore different facets of the field. These studies investigate peer influences on preventive healthcare behaviors, analyze the effects of health policies on hospitals' strategic behavior, and examine geographic inequalities in mortality within and across nations.

Chapter 2, co-authored with Friederike Reichel, titled **When Do Peers Influence Preventive Health Behavior? Evidence from Breast Cancer Screening**, is the first of these papers. It examines the interaction between peer effects and relative timing in preventive healthcare decisions. We analyze mammogram participation within the German breast cancer screening program, which offers standardized invitations and free biennial screenings for women aged 50-69.

Our study contributes to the literature on preventive healthcare and peer effects in several key ways. First, we provide causal evidence that peer effects in preventive healthcare are sensitive to relative timing. Women are more likely to attend screenings when their peers' appointments are scheduled shortly before their own. Specifically, a one standard deviation increase in the share of peers with appointments in the preceding seven days increases the likelihood of participation by 1.7 percentage points. Second, through a natural field experiment, we show that scheduling peers' appointments on the same day does not affect participation, highlighting the nuanced nature of peer influence and challenging simplistic notions of social influences on

health decisions. Third, we demonstrate how using administrative data can approximate social networks, offering a cost-effective method for studying peer effects in healthcare. By leveraging address and birth date data, we identify peers based on spatial proximity and age similarity, focusing on specific timing criteria, such as appointments within the same day or week, to measure peer influence.

Our analysis reveals that peer influence requires time to manifest, suggesting that information sharing or behavioral observation may drive the effect. We also document spatial autocorrelation in participation within villages and find positive associations between screening uptake and both socioeconomic status and regional measures of social capital, indicating broader community and regional influences.

These findings have important implications for the design of screening programs and other preventive health initiatives. They suggest that careful scheduling of appointments, accounting for peer networks and the temporal aspects of peer influence, could increase participation rates without more costly interventions. These implications extend beyond individual timing, suggesting broader considerations for enhancing preventive health program participation through strategic social choice architecture design.

Chapter 3, co-authored with Corinna Hartung, is titled **The Effect of Minimum Volume Requirements on Strategic Behavior of Hospitals**. This study examines how Minimum Volume Requirements (`MVR`) influence strategic hospital behavior and patient outcomes. We analyze data from comprehensive billing records and quality reports spanning all German hospitals from 2008 to 2017, focusing on total knee replacement surgeries and two types of complex cancer surgeries.

Our research makes novel contributions to the literature on the strategic behavior of hospitals. First, we demonstrate that `MVR` increase the number of surgeries in hospitals near the threshold, but this effect is mainly seen in

knee replacements and not in cancer surgeries, which involve a more complex interdisciplinary decision-making process. This finding reveals how hospitals strategically respond to policy incentives.

Second, we provide evidence of misreporting in hospital quality reports near the minimum thresholds. Specifically, hospitals report more procedures than they have billed, raising concerns about the reliability of these public documents, which can influence patient decision-making and public perception of hospital performance.

Third, our instrumental variables analysis reveals no significant changes in patient age or health outcomes for the marginal patients treated due to `MVR` pressure. However, we observe that hospitals draw patients from further away when they are at risk of missing minimum thresholds. This suggests that hospitals do not change clinical eligibility criteria or compromise care quality, but instead draw patients from a wider geographical area to meet the threshold.

Despite identifying some adverse effects, their magnitude appears to be small. This suggests that minimum volume requirements may still be a viable policy tool for centralizing surgeries in higher-volume hospitals, potentially improving patient outcomes.

Our findings contribute to the literature on strategic behavior of healthcare providers, focusing on hospitals' responses to volume-based quality measures. We provide evidence of modest increases in procedure volumes, misreporting in quality reports, and shifts in patient geographical distribution. These insights reveal how hospitals strategically adapt to minimum volume requirements.

Chapter 4, co-authored with Amelie Wuppermann, Joachim Winter, Hannes Schwandt, and Janet Currie, is titled **Geographic Inequality in Income and Mortality in Germany** and was published in *Fiscal Studies* (Redler et al. 2021). This study examines the evolution of socio-economic inequality in mortality rates in Germany from 1990 to 2015 across different age groups and

genders. Building on the methodology introduced by Currie and Schwandt (2016), we utilize comprehensive data from the German Federal Statistical Office, contributing valuable descriptive insights to the literature on health disparities in post-reunification Germany.

First, we document a significant reduction in mortality inequality across German districts, particularly for infants, children, and the elderly. Second, we demonstrate that this reduction was primarily driven by improvements in the regions of former East Germany following reunification. This result highlights the profound impact of political and economic changes on health outcomes.

Our most striking finding is that while income gaps between East and West Germany persist, the East has made remarkable progress in reducing mortality rates, even surpassing the West in some demographics when controlling for income. This observation challenges simplistic notions about the relationship between income and health outcomes, suggesting that other factors, possibly related to healthcare system changes or social policies, played a crucial role in mortality reductions.

Our analysis of Germany's unique historical context provides insights into the complex interplay between socioeconomic factors, policy changes, and health outcomes. This result underscores that substantial population-level health improvements are possible, even when economic disparities persist.

Chapter 5, titled **Inequality in mortality between Black and White Americans by age, place, and cause and in comparison to Europe, 1990 to 2018**, was conducted with a large team of international collaborators, led by Hannes Schwandt and Janet Currie. It was published in *PNAS* (Schwandt et al. 2021). In this comprehensive study, we compare mortality inequality trends within the United States across racial groups and between the United States and six European countries.

First, the study reveals that while the gap in life expectancy between

Black and White Americans has narrowed significantly (by 48.9%) between 1990 and 2018, inequalities in life expectancy remain more pronounced in the United States than in Europe. This finding underscores the persistent nature of health disparities in the US.

Second, it demonstrates that improvements in lower-income areas had the greatest impact on reducing the racial life expectancy gap in the US. This result highlights the importance of targeting interventions and resources to economically disadvantaged areas to address health inequalities.

The study also provides important insights into the comparative trends between the US and Europe. In 1990, Black Americans had much higher mortality rates compared to European countries, while White American mortality rates were similar to those in Europe. However, diverging trends emerged over time. While Black Americans made some progress in reducing the mortality gap with Europeans, White Americans have increasingly fallen behind European mortality rates.

Notably, by 2018, mortality gradients across various European countries fell into a narrow band, despite different starting points and economic conditions. This European convergence stands in stark contrast to the US situation, where a general health disadvantage became apparent. By 2018, European mortality rates were lower than US White American rates across all income areas, highlighting a widening gap between the US and Europe in terms of population health outcomes.

These findings have significant implications for public health policy and research. They suggest that while progress has been made in reducing racial health disparities within the US, there is still substantial room for improvement, particularly when compared to European benchmarks. The study underscores the need for continued efforts to address health inequalities, especially in lower-income areas.

References

Arrow, K. J. (1963). "Uncertainty and the Welfare Economics of Medical Care". *American Economic Review* 53.5: 941–973.

Currie, J. and Schwandt, H. (2016). "Inequality in mortality decreased among the young while increasing for older adults, 1990–2010". *Science* 352.6286: 708–712.

Currie, J., Davis, L., Greenstone, M., and Walker, R. (2015). "Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings". *American Economic Review* 105.2: 678–709.

Gawande, A. (2002). *Complications: A surgeon's notes on an imperfect science.* Metropolitan Books.

Grossman, M. and Jacobowitz, S. (1981). "Variations in Infant Mortality Rates Among Counties of the United States: The Roles of Public Policies and Programs". *Demography* 18.4: 695–713.

Gruber, J. and Frakes, M. (2006). "Does falling smoking lead to rising obesity?" *Journal of Health Economics* 25.2: 183–197.

Kessler, D. and McClellan, M. (1996). "Do doctors practice defensive medicine?" *The Quarterly Journal of Economics* 111.2: 353–390.

Miller, S., Wherry, L. R., and Foster, D. G. (2023). "The economic consequences of being denied an abortion". *American Economic Journal: Economic Policy* 15.1: 394–437.

Redler, P., Wuppermann, A., Winter, J., Schwandt, H., and Currie, J. (2021). "Geographic Inequality in Income and Mortality in Germany". *Fiscal Studies* 42.1: 147–170.

Schwandt, H., Currie, J., Bär, M., Banks, J., Bertoli, P., Bütikofer, A., Cattan, S., Chao, B. Z.-Y., Costa, C., González, L., Grembi, V., Huttunen, K., Karadakic, R., Kraftman, L., Krutikova, S., Lombardi, S., Redler, P., Riumallo-Herl, C., Rodríguez-González, A., Salvanes, K. G., Santana, P.,

Thuilliez, J., Van Doorslaer, E., Van Ourti, T., Winter, J. K., Wouterse, B., and Wuppermann, A. (2021). "Inequality in mortality between Black and White Americans by age, place, and cause and in comparison to Europe, 1990 to 2018". *Proceedings of the National Academy of Sciences* 118.40: e2104684118.

Chapter 2

When Do Peers Influence Preventive Health Behavior? Evidence from Breast Cancer Screening

Abstract

We analyze the potential for social choice architecture to increase take-up rates of breast cancer check-ups in a large sample of women in Germany. We provide causal evidence that the relative timing of check-up appointments among peers matters for participation: A woman is more likely to participate in breast cancer screening when her peers' appointments are scheduled shortly before her own. A simple intervention, however, shows that scheduling peers' appointments on the same day does not affect participation. We discuss possible mechanisms underlying the observed pattern of peer effects and highlight policy implications.[1]

[1]This chapter is co-authored with Friederike Reichel. We thank Valeria Burdea, Silvia Castro, Timm Opitz, Davide Pace, Klaus Schmidt, Peter Schwardmann, Joachim Winter, Amelie Wuppermann, and seminar audiences at LMU Munich, TU Munich, and the Annual Conference of the German Health Economics Association 2023 for helpful comments. Miriam Müller provided excellent research assistance. We used the generative AI tools ChatGPT and Claude in order to edit the text for clarity, form, and grammar. We reviewed and edited this content as needed and take full responsibility for the content of the publication. Financial support from the German Science Foundation (through Collaborative Research Center TRR 190, project number 280092119 and the GRK 1928), the Bavarian Academy of Sciences and Humanities, the Joachim Herz Foundation, and the Elite Network of Bavaria (through the Evidence-Based Economics program) is gratefully acknowledged. IRB approval was granted by the ethics committee of the economics department at LMU Munich. The study was pre-registered at AsPredicted (#102076).

2.1 Introduction

Preventable diseases are a common cause of death in high-income countries (Mokdad et al. 2004; World Health Organisation 2019). While preventive health care is commonly viewed as highly cost-effective, only a small share of individuals takes up preventive services (Borsky et al. 2018). The abundant supply of often-free preventive health care in high-income countries suggests that low take-up rates are primarily a demand-side issue.[2]

Numerous impediments to the demand for preventive health care have been discussed in the literature. Among them are a lack of or an inadequate delivery of information (Alsan et al. 2019; Torres et al. 2021; Alsan and Eichmeyer 2023), pecuniary and non-pecuniary costs associated with participation (Banerjee et al. 2010; Campos-Mercade et al. 2021), flawed information processing (Loewenstein et al. 2013; Handel and Kolstad 2015; Einav et al. 2020), motivated avoidance of (health) information (Oster et al. 2013; Golman et al. 2017; Golman et al. 2022) and a lack of attention or awareness (Milkman et al. 2021; Dai et al. 2021). Some of these barriers are already addressed in preventive health care programs in high-income countries.[3] Others are difficult to address due to institutional constraints or welfare concerns.[4]

In search of new strategies to increase the take-up of preventive health care, we turn to the literature on peer effects and social dynamics in preventive health care behavior (Bouckaert et al. 2020; Francetic et al. 2022; Karing

[2]For example, in the United States, health insurance groups are obligated to provide coverage for any service recommended by the U.S. Preventive Services Task Force, without any cost to the patient, regardless of the associated expenses.

[3]In our setting, for example, information is abundant and reliable, costs of participation are low, individual benefits are considered to be high (Katalinic et al. 2020), and measures are in place to minimize the psychological costs of receiving a positive diagnosis.

[4]Correcting the motivated avoidance of health information, for instance, is not unambiguously welfare-improving (Brunnermeier and Parker 2005; Schwardmann 2019). Often an increase in awareness is achieved by information campaigns and reminders targeted at the individual, for example, through text messages (Milkman et al. 2021; Dai et al. 2021). Targeting individuals may, however, not always be feasible due to privacy concerns or institutional constraints.

2024). We investigate the potential of social choice architecture, specifically the timing of check-up invitations and appointments, to activate these peer effects. We focus on breast cancer screening in Germany to answer the following two related research questions. Does inviting peers simultaneously to their breast cancer screening check-ups increase participation rates? When do peers influence an individual's decision to take up breast cancer screening?

The German breast cancer screening (BCS) program has many advantages to study these questions. First, BCS programs are an important instance of organized prevention efforts. Among women, breast cancer is the most prevalent cancer and responsible for a higher loss in disability-adjusted life years than any other cancer (Wild et al. 2020).[5]. In an effort to decrease the morbidity and mortality of breast cancer, many high-income countries have established nationwide BCS programs. Due to EU-wide guidelines, the German BCS is highly comparable to other European BCS programs.[6] This extends the policy relevance of our results beyond the context of our study. Even though most of the aforementioned reasons for low take-up rates are addressed by the German BCS program, participation rates are stagnating at 50% (Deutsches Mammographie-Screening-Programm 2022). As a consequence, there is a need to consider so far neglected aspects, such as the social choice architecture under which the individual makes her participation decision. Importantly, the German BCS program records participation decisions of all eligible women and demographic data that allows us to identify peers.[7] In Germany, all women aged 50-69 are offered free biennial check-ups.[8] Eligible women receive

[5]According to the Zentrum für Krebsregisterdaten (2021), 17% of all cancer-caused deaths among women in Germany were attributed to breast cancer in 2020. About one in eight women in Germany will develop breast cancer in the course of their lives (Erdmann et al. 2021)

[6]Following the European Union's recommendations, as of 2020, 25 EU Member States have implemented population-based screening programs for breast cancer (European Commission 2022).

[7]We will use address data and birthday data to construct peer networks.

[8]Women in this age group have a heightened risk of developing breast cancer, yet are still projected to derive gains from early detection and therapy.

a standardized letter inviting them to an appointment in roughly six weeks from the date they receive the letter.[9] We have access to data on 20,500 individuals residing in 19 villages in Germany.

Does inviting peers simultaneously to their breast cancer screening check-ups increase participation rates? We try to answer our first research question with a natural field experiment set within the German BCS program. Note that under the status quo, invitation letters, and appointments, are issued by birth date, so within a village, residents will receive their invitation letters at different times. By synchronizing invitation letters and appointments at the village level, our intervention seeks to turn the individual decision to participate in BCS into a more collective one. As a consequence, our intervention may increase take-up rates through several channels that have been documented in the literature on peer effects. In particular, our intervention may increase the awareness of BCS (Milkman et al. 2021; Dai et al. 2021) as well as the visibility of individual behavior (Bursztyn and Jensen 2017), thereby activating signaling (Karing 2024), influence (Esguerra et al. 2023) and conformity motives (Bernheim 1994; Funk 2010).Our experiment prevents us from distinguishing between these mechanisms, but it was designed to accommodate their potential involvement. For example, a woman may become more aware of Breast Cancer Screening when an acquaintance mentions her own invitation or upcoming appointment. Our intervention ensures that this heightened awareness coincides with the point in time when the woman needs to make up her mind about BCS.

Specifically, our randomized controlled trial assigns 9 villages to the treatment group for which invitation letters are sent out simultaneously and appointments are concentrated into the smallest possible time period. The remaining 10 villages are assigned to the control group, for which appointments are scattered across multiple weeks.

We document a precise null result with a minimal detectable effect size of

[9]The proposed appointment does not have to be confirmed but can be rescheduled (see Section 2.2).

2.5 percentage points. Contrary to our hypothesis, synchronizing invitation letters and appointments does not increase participation rates. It then remains an open question whether there even are peer effects in our setting that could be leveraged.

When do peers influence an individual's decision to take up breast cancer screening? We try to answer our second research question with administrative data. The relative timing of peers' invitations and appointments may matter for an individual's likelihood to attend.

As argued above, peers who are invited at the same time as the individual may increase her awareness of breast cancer prevention. An individual could also be more likely to participate, if more of her peers have their appointment on the same day as she does, for reasons such as being able to coordinate transport. Yet another possibility is that peers exert influence on the individual by showing behavior that the individual then wants to conform to (Bernheim 1994; Funk 2010): The share of peers who have their appointment prior to the individual's could positively influence the individual's likelihood of attending. Investigating these possibilities thus sheds light on the nature of peer effects in our setting. We identify peers based on spatial proximity and similarity in age using address data and birth dates. We then define several relative timing criteria (i.e. appointments on the same day, appointments within a week, etc.) and calculate the share of peers who fulfill the respective criterion. We can estimate the causal effect of the relative timing of peers' invitations and appointments on individual participation because of a property of the invitation algorithm that leads to exogenous variation in these peer shares.

We find that the probability that a woman will participate increases as the share of her peer group that has an appointment in the 7 days leading up to her appointment increases. This result is robust to several pre-registered definitions of peers. For example, an increase by one standard deviation in the share of peers residing within 500m of an individual who have their

appointment scheduled in the 7 days leading up to her appointment increases her likelihood of participating by 1.7 percentage points. We do not find evidence for the effect of other relative timing criteria, such as being invited on the same day.

The results of our two empirical approaches are consistent. The share of peers that are invited on the same day does not increase an individual's likelihood of attending. Note that this is precisely the peer share that was targeted by our intervention. Rather, our peer share analysis suggests that some time needs to pass between peers' appointments for peers to be influential.

In addition, we document spatial autocorrelation in participation within villages. A proxy for socioeconomic status is positively associated with participation. One regional measure of social capital, the share of registered Christians, is also positively correlated with participation.

Until now, interventions to boost participation rates in BCS have primarily focused on modifying the invitation letter, yielding mixed outcomes (Goldzahl et al. 2018; Bertoni et al. 2020). Changing a letter is an extensively studied behavioral intervention (Robitaille et al. 2021; Allcott 2011; Allcott and Rogers 2014). It may, however, as in our setting, be infeasible.[10] In contrast, our equally cheap nudge changes the social choice architecture for the individual's decision (Benartzi et al. 2013).

Our analysis of peer effects makes several contributions. First, we take low-cost methods for approximating social networks that have been documented in other contexts (Drago et al. 2020; Beaman et al. 2021), and use them to deliver new insights into preventive health behavior. Second, we build on previous studies focusing on the peer influence exerted by spouses, co-workers, and family members (Pruckner et al. 2020; Castro and Mang 2024; Goldberg et al. 2023), and show that neighbors can also affect preventive health behavior.

[10] In Germany, national guidelines dictate the letter's exact wording.

Bouckaert et al. (2020) and Francetic et al. (2022) exploit discontinuities in the eligibility status of peers to identify peer effects. In their settings, peer effects are partly due to the preventive health care offer becoming more salient. We show that even conditional on all peers being eligible, peer effects exist - as long as the timing is right.

Recently, several studies on the welfare consequences of screening point to the detrimental effects of overdiagnosis (Einav et al. 2020; Kowalski 2023). Our data preclude us from evaluating the welfare implications of BCS programs. More specifically, we are interested in identifying conditions that are conducive to high take-up rates of preventive health care offers.

In the next Section 2.2, we describe the German BCS program, the setting of our study. In Section 2.3, we describe the data that we have access to. Section 2.4 contains the experimental design and the results of our intervention. In Section 2.5, we describe the empirical approach to our peer effect analysis and its results. Section 2.6 concludes.

2.2 Background: Breast Cancer Screening in Germany

The population-wide German breast cancer screening (BCS) program, based on European Union guidelines, was established in 2005 and covers women aged 50 to 69 (Biesheuvel et al. 2011). Women in this age range who reside in Germany are invited every two years for a mammogram. This diagnostic procedure includes two X-rays of each breast from different angles. The images are independently assessed by two physicians, and any abnormal findings are referred for assessment by a specialist. If there is an abnormal finding and further testing is required, this is communicated within two weeks.

Invitation letters are sent out via mail and the content of the letter is standardized (see Appendix Figure 2.6). A decision aid booklet which

contains information on the procedure, on breast cancer in general, on possible outcomes, and on advantages and disadvantages of participation is enclosed with the invitation for first-time invitees.[11]

The letter contains a proposed appointment location, date, and time. This appointment slot is reserved for the woman, and confirming the appointment is optional. Invitation letters are typically sent out four to six weeks ahead of the appointment. The recipient of the letter can request to reschedule the appointment. If a woman does not attend the screening, she receives one reminder with a proposed alternative appointment.

The mammograms are performed in screening units. While urban and suburban areas are typically served by outpatient locations, e.g. in hospitals or specialist practices, rural areas are routinely served by mobile mammography units (MMUs). These use equipment of similar quality to outpatient locations, but they only remain at each location for a few weeks. During this time, they serve as the screening location for all local women. The duration of a stay is determined by the capacity of the unit and the number of expected appointments for local women. Any area served by an MMU is typically visited every 24 months.

Goldzahl et al. (2018) and Carrieri and Wübker (2016) show that the screening programs in Europe have increased mammography rates in the relevant age groups. However, national participation rates vary widely (Wübker 2014). In Germany, the overall uptake of screening within the program is around 50%, which is low in comparison to other EU countries, and has recently decreased (Deutsches Mammographie-Screening-Programm 2022). Age-standardized breast cancer incidence and breast cancer mortality rates in Germany are slightly above the EU-28 country average (Dafni et al. 2019).

Mammograms are also performed outside of the BCS program, typically

[11] Available at `https://www.mammography-screening.de/download/downloads/broschueren/2019-08-13_G-BA_Entscheidungshilfe_Mammographie_EN_RZ_Web_2_2.pdf`

if symptoms such as pain or breast lumps are present, or for preventive reasons, for example, given a family history of breast cancer. Therefore, the counterfactual to screening participation within the BCS program is not necessarily non-screening. Screening outside the BCS program is called opportunistic screening. The scope of opportunistic screening in Germany is not precisely known. The largest public German health insurance conglomerate reports that 8 to 12% of women in the target age range undergo a mammography exam outside the BCS program (Tillmanns et al. 2021). Nonetheless, screening within the BCS program might still be preferable to opportunistic screening due to higher diagnostic quality standards, i.e. better equipment and more experienced physicians.

2.3 Data

We use data from multiple sources. Our main dataset is administrative data from the German breast cancer screening (BCS) program. These data contain individual-level information on screening invitations, screening participation, past participation, age, and address.

The sample area is a predominantly rural area in Germany (19 ZIP code areas in one federal state, approx. 200,000 inhabitants). Our sample comprises all women aged 50 to 69 years in this area (n = 20,500). The program's regional structure corresponds to the administrative structure of Germany's ZIP codes: Women who reside in the same ZIP code are assigned to the same screening location. Every woman in the sample is invited to one mammogram during our study period. We use the initially proposed appointment date for all analyses.

For our natural field experiment (Section 2.4), the ZIP code areas correspond to treatment clusters. These areas are entirely non-urban: All towns in our sample have fewer than 50,000 inhabitants. Our data was collected

over a period of several months (2022-2023) from one mobile mammography unit (MMU) that served 19 ZIP codes from four separate MMU locations. In the rest of this paper, we will refer to the 19 ZIP code areas as villages and to the four different MMU locations as sites 1-4.[12] Participation in BCS is our main outcome variable. We use previous participation and age as our main control variables. In an additional specification, we also control for the geographical distance between the individual's address and the MMU location and for weather, a school breaks dummy, and the current COVID-19 incidence.

For our *Peer Shares* approach (Section 2.5), we use individuals' birth dates and precise information on current addresses to create proxies for peer relationships. We combine these with the exact timing of the first proposed appointment to construct our explanatory variables.[13]

In addition, we ran two accompanying surveys to better understand the take-up decision process. Respondents were predominantly women who participated in the screening. We supplement these data with regional-level characteristics for heterogeneity analyses. We evaluate the role of social capital and social economic status, which we proxy with administrative data that include voting data, data on religious affiliation, and local unemployment rates.

2.4 Intervention: Synchronizing Invitations

The purpose of our natural field experiment is to explore the potential of a simple, cheap, and scalable change to the invitation strategy to increase participation in breast cancer screening (BCS). In this section, we outline the design and implementation of our intervention and present the main results and complementary survey evidence.

[12]There is one town in our sample that has three ZIP code areas which we will treat as separate entities. Each one of the other 16 ZIP code areas corresponds to a geographically separate small area (a German *Gemeinde*).

[13]For more details, see Section 2.5.1.

2.4.1 Design

We used random assignment to allocate villages to either the treatment or the control group. Each treated village was assigned the shortest possible time slot to accommodate all appointments of its residents. Appointments for women residing in untreated villages were scattered over a larger time interval, as determined by the status quo invitation algorithm. Since invitation letters are usually sent six weeks prior to the appointment, our intervention synchronizes the receipt of the invitation letters and the dates of appointments for the treatment group.[14]

This alteration to the status quo invitation strategy intends to leverage peer effects. While we cannot discriminate among them, several channels through which our intervention could increase BCS participation seem plausible. First, our treatment may make individuals more aware of BCS. As an acquaintance mentions her own invitation or upcoming appointment, a woman may become more aware of her own invitation or upcoming appointment.[15] If, for instance, a woman was initially undecided or somewhat inattentive to the letter, such a conversation may nudge her toward participation. Second, our treatment may increase the visibility of individual behavior, which has been shown to encourage various target behaviors (Bursztyn and Jensen 2017; Karing 2024). If our intervention succeeds in making BCS a local topic of conversation, individuals' (intended) behavior should become more visible to others. An individual may be asked more often about her intention to participate. She

[14]One might be concerned about the lower level of privacy our treatment introduces. Women may not want to run into acquaintances or to be recognized by anyone at the screening. To evaluate this concern, we conducted a pre-intervention online survey with a representative sample of participants ($n = 170$). A vast majority of respondents (77%) indicated that they would not mind seeing an acquaintance at the check-up site. A smaller proportion of 12% indicated that such an encounter would indeed bother them, and 11% mentioned that they would enjoy meeting someone they know.

[15]BCS is a topic of conversation among invited women. In a representative online survey, 57% of women report that they talk with others about their decision to participate. Data from a second survey show that 88% of respondents believe that others talk openly about their participation in BCS.

may also realize through conversations that she is likely to be seen by peers at the check-up site. Third, simply observing more peers participating in BCS within a shorter time span of her own appointment may induce a woman to conform to this observed behavior (Bernheim 1994; Funk 2010). Alternatively, participating peers may serve as a timely reminder of a woman's own upcoming appointment. Lastly, our treatment may reduce participation costs by making neighborhood carpooling to the check-up site more feasible. Thus, we hypothesize that a woman is more likely to participate in BCS if more of her peers are invited at the same time as she is and have their appointments close in time to hers.

Negative selection of screening participants on risk (Einav et al. 2020) makes it important to draw in previous non-participants. Investigating heterogeneous treatment effects by previous participation allows us to identify the effect of our intervention on this particularly policy-relevant subgroup of previous non-participants. The official invitation letter presents public health information in a formal manner, encouraging a woman to consider her personal costs, risks, and benefits associated with the check-up. Clearly, this invitation strategy fails to convince previous non-participants to take up BCS. These women may be more receptive to the informal communication channels that our intervention seeks to activate. Since our intervention only complements the invitation strategy currently in place and does not substitute any of its features, we do not expect any adverse effects on previous participants. The strong correlation between past participation and present participation suggests that any observed treatment effect on previous non-participants could also persist long-term.

Our sample comprises 20,500 women in 19 villages. During our intervention, the mobile mammography unit (MMU) served the local population at four different sites. These sites are typically chosen to be central and easily accessible, such as the parking lot of a shopping mall or a village square.

Table 2.1: Experimental Design

Status quo - Appointments

Week 1					**Week 2**					**Week 3**					**Week 4**					**Week 5**				
M	Tu	W	Th	F	M	Tu	W	Th	F	M	Tu	W	Th	F	M	Tu	W	Th	F	M	Tu	W	Th	F
All villages																								

Intervention - Appointments

Week 1					**Week 2**					**Week 3**					**Week 4**					**Week 5**				
M	Tu	W	Th	F	M	Tu	W	Th	F	M	Tu	W	Th	F	M	Tu	W	Th	F	M	Tu	W	Th	F
Control villages					Treated vill. 1					Control villages					Treated vill. 2					Control villages				

Notes: The status quo algorithm only considers dates of birth and the date of the last appointment. At a given site, the appointments of women from all surrounding villages are scheduled within a couple of weeks. The intervention reserves slots of a few days for treated villages (two in the example above). The capacity constraints of the MMU and the expected number of women attending their appointments determine the length of these slots. The remaining slots are filled with appointments of women residing in control villages according to the status quo algorithm.

The impact of our treatment is limited by two factors. First, by law, all appointments must be offered within 22 to 26 months of a woman's previous appointment.[16] Second, given that an MMU only remains at each site for a couple of weeks or months at most, the appointments of women in our control group are also relatively close to each other.

Table 2.1 illustrates the scheduling of appointments under the intervention. Appendix Figure 2.7 shows that this part of the manipulation, i.e. the concentration of appointments in treated villages, was successfully implemented across all four screening sites. The standard deviation of the distribution of appointments within a village measures the degree to which appointments are dispersed over time. These are systematically smaller in treated villages than in untreated villages.[17] In both the treatment and the control group, around 15% of women rescheduled their appointments. Appendix Figure 2.8 shows the final distribution of appointments after rescheduling. Again, the standard deviations of final appointment dates are systematically smaller in treated

[16] If a woman is invited for the first time, this range applies to her 50^{th} birthday.

[17] The average standard deviation of initially proposed appointment dates weighted by village size within the treated (untreated) villages is 5.24 days (14.71 days).

villages than in untreated villages.[18]

Usually, letters are sent out six weeks prior to the proposed appointment. Under the current practice, appointments are scattered over time, and as a result, the arrival of appointment letters is similarly dispersed. Our intention was to measure the combined effect of both receiving the letters *and* having the appointments close in time. As a result of an unintended deviation in the implementation process, all invitation letters for appointments at sites 3 and 4 (which correspond to the second half of the trial) were sent out simultaneously, regardless of treatment status.[19] While this was not intended and reduces our chance of precisely estimating our main treatment effect, it allows us to separately estimate (i) the effect of sending out the letters simultaneously, and (ii) the effect of sending out the letters simultaneously and synchronizing the appointments.

2.4.2 Estimation

As pre-specified, we estimate the average treatment effect by the following equation

$$Y_{iv} = \beta_0 + \beta_1 T_v + \gamma prevpart_i + \delta age_i + \epsilon_{iv}. \tag{2.1}$$

Y_{iv} represents a binary participation variable, and $prevpart_i$ is a dummy that equals 1 if a woman participated in the screening previously (at least once during 2018-2021). We also control for age. Previous participation is highly predictive of our outcome and thus increases the power of our design.

We cluster our standard errors at the village level, our level of randomization. Because there are at most 19 clusters in our setting, we use the wild-cluster bootstrap proposed by Cameron et al. (2008) and

[18]The average standard deviation of final appointment dates in the treatment (control) group is 15.13 days (21.01 days).

[19]Appendix Table 2.5 details how the intervention differs between the first and the second half of the trial.

use Rademacher-weights, as suggested by Canay et al. (2021). We report bootstrapped p-values in addition to our estimates.[20]

The following estimation interacts previous participation with the treatment dummy to detect heterogeneous treatment effects by previous participation.

$$Y_{iv} = \beta_0 + \beta_1 T_v + \beta_2 T_v \times prevpart_i + \gamma prevpart_i + \delta age_i + \epsilon_{iv}. \quad (2.2)$$

2.4.3 Results

On average, women who previously participated and women who did not previously participate in the BCS are unaffected by our intervention. Figure 2.1 shows participation rates by treatment status and previous participation.

Table 2.2 presents the experimental results. Throughout all specifications, previous participation is strongly associated with participation. We separately estimate (i) the effect of simultaneous appointments (columns (1) - (3)), (ii) the effect of simultaneous letters (columns (4) - (6)), and (iii) the combined effect of simultaneous appointments and letters (columns (7) - (9)) on participation. Within each of the three groups of columns, the leftmost column presents results from Estimation 2.1. The middle column includes a dummy for previous participation as in Estimation 2.2. In the rightmost column, control variables are added.[21]

Simultaneous appointments (Table 2.2 columns (1) - (3)): There is no overall effect of bunching appointments on participation. The participation rate of residents of villages who are invited within as narrow a time frame as possible is not different from villages whose residents' appointments are

[20] We follow the approach of Alan et al. (2023) who face a similar econometric setting.

[21] Reassuringly, treatment assignment is balanced across all control variables, except for precipitation. Appendix Table 2.6 shows that control villages have a larger population and are located closer to their screening site, but these differences are not statistically significant.

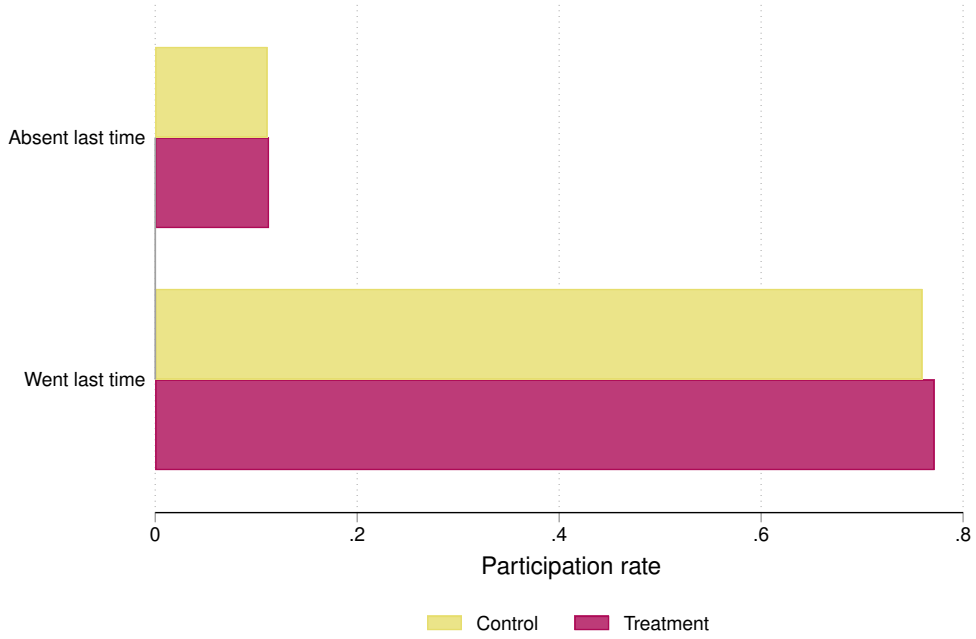


Figure 2.1: Participation Rate by Previous Attendance

Notes: Previous attendance, *Absent last time* v. *Went last time*, refers to participation in 2018-2021 (i.e. a woman did or did not attend during this period). Previous attendance has been known to be a strong predictor of participation (Deutsches Mammographie-Screening-Programm 2022).

scattered over time. Throughout all columns, the point estimates for the interaction effects are positive, but not significantly distinguishable from zero. We thus do not find evidence for a heterogeneous treatment effect by previous attendance.

Simultaneous letters (Table 2.2 columns (4) - (6)): Due to the accidental treatment arm discussed in Section 2.4.1, we now isolate the effect of receiving the invitation letters simultaneously within a village. To this end, we are comparing the control villages of the first two sites to the control villages of the last two sites. Appointments were not bunched for any of these villages. The control villages of the last two sites, however, received invitations at the same time. Thus, they act as treated villages for this analysis. Sending out letters simultaneously without bunching appointments naturally generates variability

Table 2.2: Main Results - Dependent Variable: Participation

	Simultaneous appointments			Simultaneous letters			Sim. Appointments + letters		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treated	0.006	0.000	-0.002	-0.007	-0.007	-0.019	0.001	-0.006	-0.012
	(0.010)	(0.006)	(0.010)	(0.007)	(0.016)	(0.013)	(0.015)	(0.013)	(0.019)
Went last time * Treated		0.011	0.007		0.000	0.007		0.012	0.012
		(0.022)	(0.021)		(0.029)	(0.026)		(0.027)	(0.027)
Lead time				0.000	0.000	-0.000	0.000	0.000	-0.000
				(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Went last time	0.658***	0.654***	0.659***	0.653***	0.653***	0.655***	0.660***	0.653***	0.654***
	(0.012)	(0.015)	(0.014)	(0.015)	(0.021)	(0.022)	(0.014)	(0.021)	(0.022)
Age	-0.002**	-0.002**	-0.002**	-0.002**	-0.002**	-0.002*	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Precipitation, in cm			0.006			0.003			0.004
			(0.006)			(0.009)			(0.005)
Avg. temperature (C)			0.000			-0.001			-0.001
			(0.000)			(0.002)			(0.001)
Distance to MMU (km)			0.001			0.001			0.001
			(0.001)			(0.001)			(0.001)
School break			-0.011*			-0.013*			-0.013
			(0.006)			(0.006)			(0.012)
Covid 7d inc.			-0.007**			-0.008**			-0.006*
			(0.003)			(0.003)			(0.003)
Constant	0.223***	0.225***	0.247***	0.215***	0.216***	0.282**	0.180**	0.184**	0.241***
	(0.043)	(0.044)	(0.049)	(0.050)	(0.043)	(0.102)	(0.063)	(0.064)	(0.049)
Observations	18336	18336	18095	11042	11042	10816	11362	11362	11284
R^2	0.429	0.429	0.432	0.422	0.422	0.426	0.432	0.432	0.433
Clusters	19	19	19	10	10	10	15	15	15
WC Bootstrap p-value	0.541	0.825	0.950	0.332	0.715	0.475	0.962	0.907	0.880

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. Wild-Cluster Bootstrap p-values for (joint) significance of treatment effects are reported. Columns (1) – (3) compare all control villages (*Treated*= 0) with all treatment villages (*Treated*= 1). Columns (4) - (6) compare control villages of the first half of the trial (*Treated*= 0) to control villages of the second half of the trial (*Treated*= 1). Columns (7) - (9) compare control villages of the first half of the trial (*Treated*= 0) with all treatment villages (*Treated*= 1). See Appendix Table 2.5 for a breakdown of how the intervention differs between the first and the second half of the trial. Results for first invites are reported separately in the Appendix.

in the time interval between the receipt of the letter and the appointment date. The variable *lead time* captures this gap in time. As one might expect, the relationship between lead time and participation is negative. We do not find evidence that sending out the invitation letters at the same time increases participation.

Simultaneous letters and appointments (Table 2.2 columns (7) - (9)): We document no effect of our originally planned intervention on participation. Again, this effect does not mask heterogeneous treatment effects by previous attendance.

Appendix Table 2.7 restricts the analysis to first-time invitees. We find significant and negative treatment effects for this new cohort (6.3% difference, p-value = 0.05). We argue for caution in interpreting these results for several reasons. First, because we cannot control for prior participation, statistical power becomes a more prominent concern, leading to a presumably higher intra-cluster correlation (ICC). Second, as we observe only one cohort per village, the sample size within some villages is very small (n<50 for 6 villages). This might further elevate the ICC, given that participation rates are correlated within cohorts. Due to these concerns and in line with our pre-registration, we exclude first-time invitees from the primary analysis.

The null result of the intervention needs to be interpreted in the context of two limitations. First, there is a statistical power constraint, as the number of clusters is small, and the size of the clusters varies significantly. The confidence interval of the main treatment effect estimates indicates that we can only rule out effect sizes larger than approximately 2.5 percentage points. Since the implementation costs of synchronizing letters and appointments within villages are negligible, even a true effect size of 1 or 2 percentage points would be relevant and might warrant changes to the invitation strategy.[22]

The second main limitation is the specific setting of our study, namely

[22]Synchronizing appointments requires a small change in the appointments-scheduling algorithm.

the rural area served by mobile mammography units. Collaborating with MMUs was partly driven by logistical considerations. There are at least two reasons why our intervention is likely to yield stronger results in a less rural setting. First, an initial pilot study carried out prior to the national roll-out of breast cancer screening reports a substantially higher participation rate at mobile screening units than at non-mobile screening units (Kolip and Wurche 2005).[23] Assuming that it becomes increasingly harder to draw in marginal non-participants at higher participation rates, further increasing participation rates at MMUs is more difficult than at stationary units. Second, as Figures 2.7 and 2.8 document, appointments at MMUs fall into relatively short time spans even absent our intervention. Consequently, our intervention can be interpreted as light touch, further reinforcing conditions that we expected to be conducive to high take-up rates.

2.4.4 Survey evidence

We ran two surveys alongside the intervention to complement our understanding of the decision environment and to explore mechanisms such as awareness, visibility, communication, and coordination. Detailed procedures and results are reported in the Appendix.

Our survey results suggest that our intervention increased the perceived knowledge of peers' participation behavior and the perceived openness towards conversations around BCS. It also suggests that there is scope for unambiguously welfare-improving

[23] Unfortunately, the program's yearly assessment reports do not break down participation rates by mobile units and stationary units, but anecdotally, participation rates are still thought to be higher at mobile units.

2.5 Peer Effects on Participation

Our pre-registered *peer shares* setup aims to evaluate whether and when peer influence matters for an individual's decision to participate in BCS. We first describe our setup in detail and then present the results.

2.5.1 Setup

In our setup, we consider all women in one village as potential peers to each other. We use two dimensions as proxies for peer relationships to classify peers: Geographic proximity and age proximity. We then test whether the relative timing of initially proposed invitation dates between those peers affects participation.

Relative timing might matter because the dates of appointments are closely linked to the dates of letter receipt, as the letters are usually sent a fixed period before the appointments. This creates opportunities for women to influence each other's decisions through conversations when they receive the letter, just before the appointment, or at any point in between. The decision to participate remains flexible, as appointments neither need to be confirmed nor actively canceled. This results in ongoing potential for peer effects.

We employ two alternative approximations of peer relationships that generate unweighted and undirected peer networks. Our first approximation is based on address data. The underlying assumption is that geographic proximity is related to the likelihood of peer interactions. In a similar setting to ours, rural Austria, Drago et al. (2020) document a high communication intensity among neighbors that declines monotonically with geographical distance. Marmaros and Sacerdote (2006) find that geographic proximity is a more important determinant of peer interactions than interests or family background, further substantiating our assumption. Whenever two women live

Table 2.3: Criteria to Define Peers

Peer criterion	**Dimensions**
Geographical distance d	$d < \bar{d}$ meters, $\bar{d} \in \{50, 200, 500, 1000\}$
Time between birth dates k	$k < \bar{k}$ years, $\bar{k} \in \{0.5, 1, 2\}$
d and k	Combinations of criteria above (4 x 3 = 12)

Notes: The peers of a woman are defined as all women who live fewer than $\bar{d}$ meters away, have an age difference of less than $\bar{k}$ years, or a combination of both criteria (i.e. women who live less than $\bar{d}$ meters away and have an age difference of less than $\bar{k}$ years). These criteria were pre-registered.

within a given cutoff distance from each other, we consider them as linked.[24] We vary the cutoff distance.

Our second approximation relies on the relative age of peers. The underlying rationale is a homophily argument (McPherson et al. 2001), which suggests that women of similar ages are more likely to interact with each other, possibly due to shared age-based experiences such as attending school together. As a narrower third approach, we combine the two previous dimensions and define peers as women who live within a specified distance and have birth dates that are within a specified range of months. We end up with several approximations of peer networks, all of which were pre-registered. Combining these with pre-registered relative timing criteria, we can analyze along which dimensions peer effects matter for the decision to get a mammogram. Having several approximations of peer networks allows us to evaluate the robustness of any effect.

For each invited individual i, we calculate the number of peers n who fulfill each peer criterion. Table 2.3 lists the specific cutoffs used as peer criteria for our peer effects models.

Given a set of n peers, we then analyze the role of the relative timing of the proposed screening dates. We do this based on the assumption that the relative timing of the proposed date is related to peer interactions. Intuitively,

[24]We calculate the haversine distance, i.e. the shortest distance over the earth's surface between two coordinates.

Table 2.4: Relative Timing Criteria

Relative timing	**Intuition**
Same day	Similar timing of letter, exact same date of appointment
Within two days	Similar timing of letter and appointment
Before	Peer had appointment before i (excludes same day)
1 to 7 days before	Peer had appointment closely before i
More than 7 days before	Gap between appointments, peer's appointment earlier (Placebo)
More than 7 days after	Gap between appointments, peer's appointment later (Placebo)

Notes: To construct a peer share for a woman i, we calculate the share of a set of peers who fulfill a binary relative timing criterion c. The sets of peers vary by specification as defined by Table 2.3. The relative timing refers to the peer's proposed initial appointment relative to the proposed initial appointment of woman i. The intervals for the criteria *Before*, *More than 7 days before* and *More than 7 days after* are only bounded on one side.

if two women receive the letter on the same day and have an identical proposed screening date, this could increase the likelihood of an interaction among them that relates to breast cancer screening. Table 2.4 shows the relative timing criteria used.

We then construct peer shares for each individual i as the share of n women who fulfill a relative timing criterion. Figure 2.2 illustrates how a peer share using peers as determined by geographical proximity is determined. All women living within distance $\bar{d}$ of woman i are counted, here $n = 10$. Then, the number of peers who fulfill the relative timing criterion is counted, here $x = 4$. This results in a peer share $share_i = 0.4$. If the relative timing criterion was *Same day*, this would imply that 40% of individual i's peers have their appointment on the exact same day as individual i.

Peer shares are a function of the timing of the proposed screening dates. They vary based on the geographical location of the woman, the age of the woman and the proposed appointment. Appendix Table 2.8 shows how the calculated peer shares vary. For smaller sets of defined peers (e.g. $d < 50m$), the values mechanically diverge more.

For exogeneity of the set of our key explanatory variable $share_i$, the

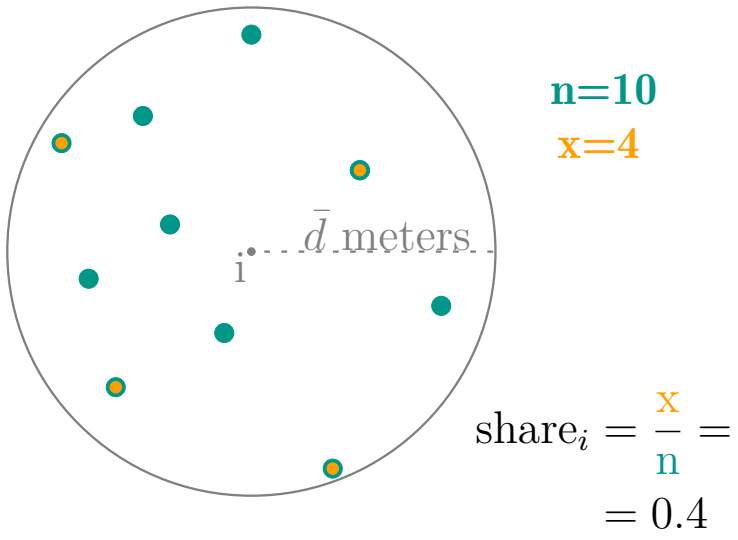


Figure 2.2: Geographical Peer Shares Calculation

Notes: All n women residing within (haversine) distance of $\bar{d}$ meters of woman i are considered her peers. Among these peers, all x women whose appointments fulfill relative timing criterion c with respect to woman i's appointment are counted. This count, x, is divided by the number of peers, n, in order to calculate woman i's peer share, $share_i \in [0, 1]$.

proposed appointment time needs to be random conditional on previous participation and distance to the screening location. The source of this randomness is the invitation algorithm. The algorithm proposes appointment dates. It uses information on previous participation of the invited women so that utilization levels of the screening site do not deviate strongly from planned utilization levels. This is accomplished by splitting invitees into three groups: First invites, those who participated last time, and those who did not participate last time. A mix of invitees from these groups is then assigned to each time slot. Crucially, within our groups, the algorithm sorts primarily by birth date or by the date of previous invitation. The order of appointments for first invites within a village is thus determined by the order of the birth dates. Because participation is plausibly unrelated to the birth order within a village, appointment timing is exogenous within villages for first invites. Since the previously proposed screening date again matters for the appointment in subsequent years, the source of exogeneity is passed on within cohorts

across screening years. Since appointment timing directly affects whether a relative timing criterion is fulfilled (see Table 2.4), the peer shares are, in turn, determined orthogonally to potential participation.

Due to an inadvertent change in the invitation algorithm because of our experimental intervention (see Section 2.4), this exogeneity argument does not hold for treated villages. In order to schedule appointments for the treated group, the original invitation algorithm, which sorts primarily by age, was replaced with an algorithm that primarily sorts by address. The resulting peer shares are thus not orthogonal to potential participation.[25] We therefore exclude all treated villages from the peer share analysis.

Given our setup, we estimate the following linear probability model for each version of the constructed peer shares. Y_{iv} represents screening participation that equals 1 if a woman participates in the screening at any point during our study period.

$$Y_{iv} = \beta_0 + \beta_1 share_i + \gamma prevpart_i + \pi_v + X\beta + \epsilon_i. \tag{2.3}$$

Our coefficient of interest is β_1, which can be interpreted as the change in the participation rate in % given an increase in $share_i$ of 1%. We control for previous participation as well as age, distance from residence to the screening location, and the absolute number of peers n. Due to village fixed effects, any estimates can be understood as effects relative to the average woman in a municipality.

2.5.2 Results

We first present results for estimating Equation 2.3. We flexibly vary the construction of $share_i$ as previously discussed. The variable varies across the peer dimensions geographic distance $\bar{d}$ and age difference $\bar{k}$ and the relative

[25] Sorting by address likely results in endogeneity of the share variable. We show in Section 2.6 how a proxy for SES is related to participation.

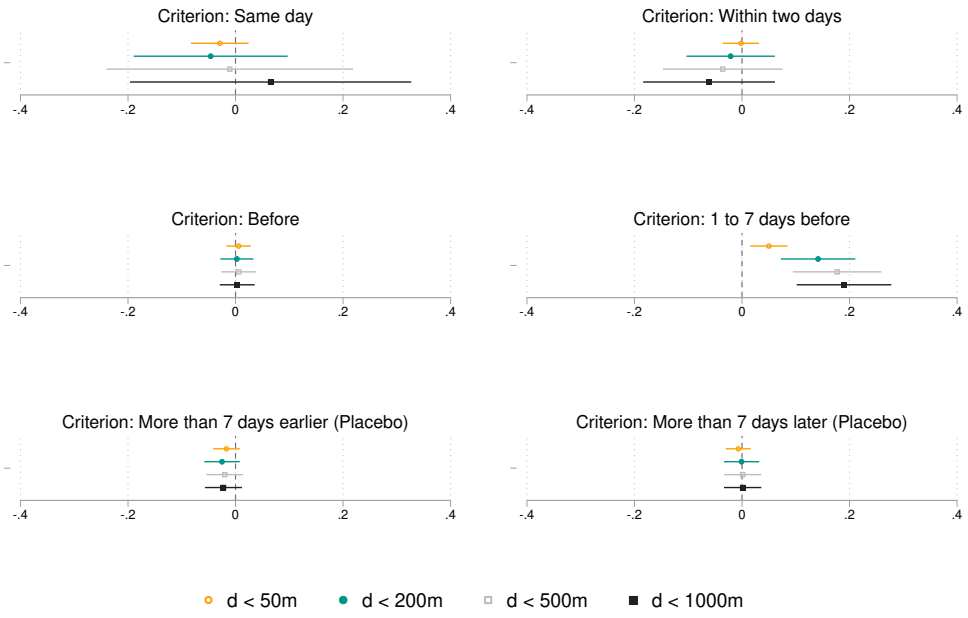


Figure 2.3: Peer Shares by Geographic Distance

Notes: This figure presents resulting β_1 from estimations of Equation 2.3. Each bar and its standard errors represent one regression that includes village fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 2.4). Each bar per panel represents a different peer criterion, here by distance in meters (see Table 2.3). The full regression results are presented in Appendix Tables 2.9 - 2.14.

timing of the invitations (see Section 2.5.1).

Figure 2.3 shows the estimated β_1 coefficients for different versions of Equation 2.3 where the peer criterion is geographic distance. Each set of a point estimate and its 95% confidence interval represents a separate regression. Detailed regression results are presented in Appendix Tables 2.9 to 2.14.

Most coefficients are statistically indistinguishable from zero. Notably, the results for the criterion *Same day* suggest that identical timing of appointments does not play a role. This is in line with our survey result that carpooling is uncommon (Appendix Table 2.27).

The panel for *1 to 7 days before* shows significant positive estimates of the coefficient of interest. This reflects that when more peers of an individual

within a specified geographic distance are invited to the screening the week before the individual, her own participation probability increases. These effects are relevant in size. For example, the point estimate for $d < 500m$ suggests that an increase in this peer share by one standard deviation (0.093, see Appendix Table 2.8) leads to an estimated increased participation probability of 1.7 percentage points. Coefficients for $d < 50m$ are closest to zero. This is unsurprising because the number of peers within such a close distance is often small. Given the construction of the explanatory variable $share_i$, extreme values such as 0 or 1 are more common. These extreme values, in combination with a smaller number of peers, likely bias estimates towards zero.[26]

To explore heterogeneity within the results for the criterion *1 to 7 days before*, we split our sample by weekday of the proposed appointments in Figure 2.4. While confidence bands increase, the relevant β_1 point estimates are lower when the initial appointment is earlier in the week (Monday or Tuesday). The results suggest that appointments later in the week are more susceptible to peer influence. A mechanism consistent with smaller peer effects on Monday and Tuesday is the lack of opportunity to rearrange one's week. For example, if a woman is influenced by a peer on the weekend - when typically more leisure is available which might result in peer interactions - she might only be able to make time to attend her appointment if it is later in the week.

Figure 2.5 is analogous to Figure 2.3, but the strategy to proxy peers differs. Now, women of similar age are considered her peers. Specifically, the distance between the birth dates of woman i and her potential peers is considered. As above, the share of peers that fulfill a relative timing criterion with respect to their screening appointment determines the values of $share_i$.

Here, a similar overall picture emerges. Again, the estimates for the timing criteria *1 to 7 days before* are positive. While the estimates are smaller in absolute terms, they are still significantly different from zero. Defining the

[26] Another aspect that biases these estimates toward zero is that we control for the number of peers linearly.

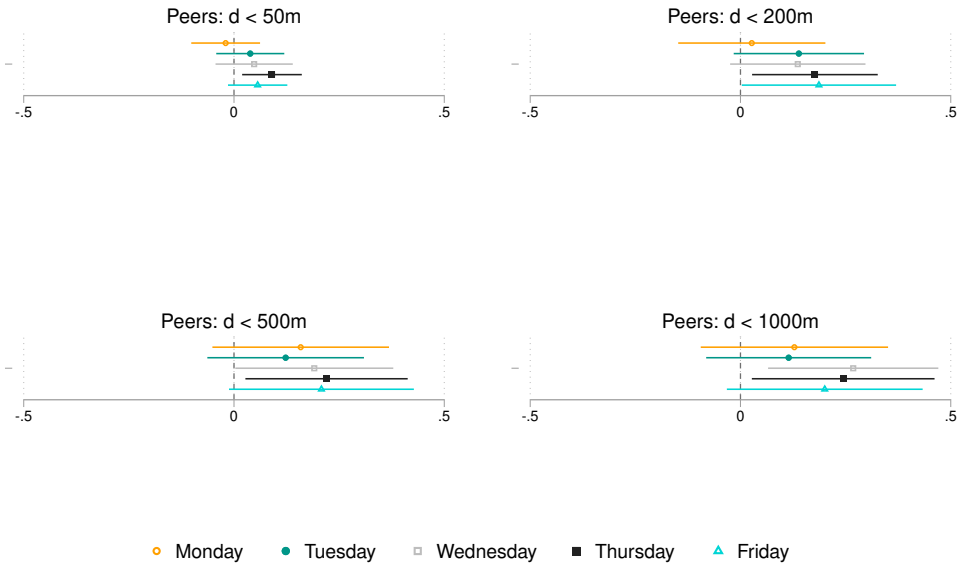


Figure 2.4: Coefficients for Relative Timing Criterion *1 to 7 days before* by Day of the Week

Notes: This figure presents β_1 estimates of Equation 2.3 separately by day of the week of the invited woman's initial appointment. Each bar and its standard errors represent one regression that includes village fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel states the peer criterion (see Table 2.3). Each bar per panel represents a different day of the week. The relative timing criterion used here is *1 to 7 days before*.

age difference cutoff differently does not play a large role for the estimates of any relative timing criterion.

In Appendix Figures 2.9 to 2.11, we combine the peer criteria geographical distance and age distance. Here, only individuals who live both within distance $\bar{d}$ and are close in age are considered peers. The relative timing criteria remain unchanged. This results in less statistical power. However, the point estimates do not change much compared to the previous results: Again, we observe positive estimates for the relative timing criterion *1 to 7 days before* that are mostly statistically significantly different from zero.

Overall, these results support the hypothesis that peers serve as role models

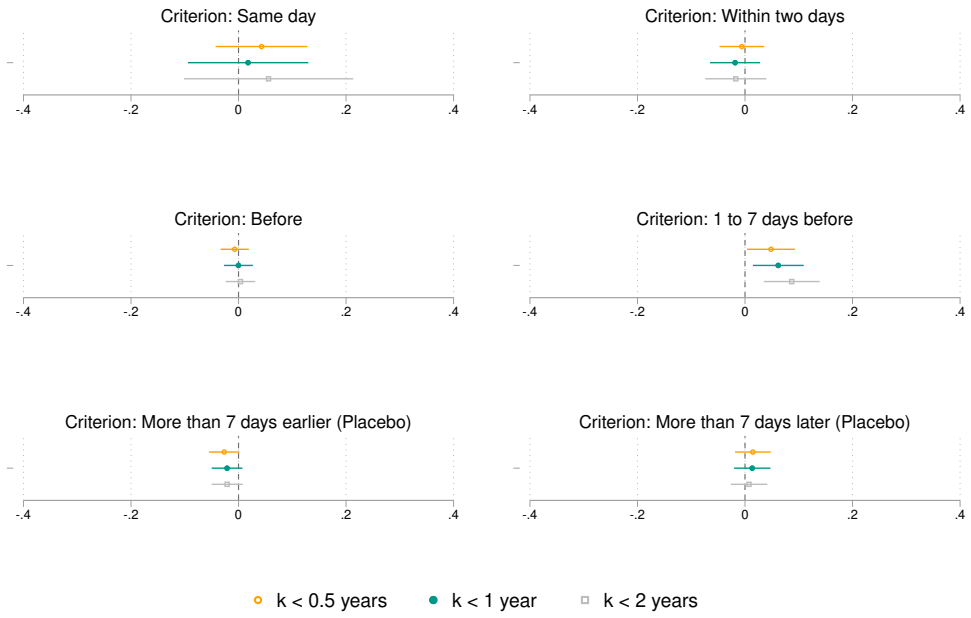


Figure 2.5: Peer Shares by Age Distance

Notes: This figure presents β_1 from estimations of Equation 2.3 which can be interpreted as the change in participation in % given an increase in $share_i$ by 1%. Each point estimate and its 95% confidence bands represent one regression that includes village-fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 2.4). Each bar per panel represents a different peer criterion, here by difference in age (see Table 2.3). The full regression results are presented in Appendix Tables 2.15 - 2.20.

and influence an individual's decision to attend when their appointments are just before the individual's appointment. Salience might also play a role: If a woman observes that a peer participates in the screening, it might serve as a reminder for her own appointment. Because the estimates for the peer criterion *Before* are close to 0, it does not seem to matter only whether a peer's appointment was before the individual's own appointment: Proximity matters.

2.6 Discussion

In this paper, we provided causal evidence that the relative timing of peers' preventive check-ups matters for individual participation: A higher share of an individual's peers who have an appointment in the days leading up to her own appointment increases her likelihood of participating in BCS. The results from our intervention indicate that synchronizing invitation letters and appointments does not increase participation rates. These results do not contradict each other. Our intervention significantly and sizeably increases the expected share of peers who have their appointment on the same day (for an illustration, see Figure 2.8). As our analysis of peer shares shows, a higher share of an individual's peers who have their appointment on the same day does not increase her likelihood of attending.

The sequence of social signals that a woman may be exposed to from the point in time that she is invited until her appointment can further reconcile our results. If there are conversations about BCS, receiving the letter at the same time as the entire village exposes an individual to mixed signals: With participation rates at 50%, in expectation half of the conversation partners will share an intention to participate and the other half will share an intention not to participate. The situation differs starkly once the individual's appointment is imminent. Since the act of participation is more salient than the act of non-participation, the individual will be mainly exposed to positive signals on the value of BCS.[27] Evidence from behavioral economics suggests that people underreact to empty signals - in our setting, the act of non-participation (Tversky and Kahneman 1973; Enke 2020; Jin et al. 2021). Thus, close to her appointment, a woman may overestimate participation in BCS and be inclined to conform to the perceived norm of participating (Bernheim 1994; Funk 2010).

[27]Arguably, a peer is more likely to share that she just went to or is going to her appointment rather than the fact that she could have gone to an appointment that she was invited to six weeks ago and decided against.

Participating peers may also remind her of her own upcoming appointment. In principle, this dynamic could have unfolded as a result of our intervention. Our peer share analysis, however, suggests that some time needs to pass between peers' appointments and the individual's appointment for peers to be influential. This longer time spell may be required for conversations to take place and for individuals to be able to react to peer influence.

We have identified the timing of check-ups as an unexplored dimension that can be manipulated even within a large-scale, tightly regulated preventive health care program. While we still consider the optimal design of screening invitation campaigns to be an open question, we hope that our identification of the scheduling as an important and mutable feature of these programs may prove useful to future research and policy design.

References

Alan, S., Corekcioglu, G., and Sutter, M. (2023). "Improving workplace climate in large corporations: A clustered randomized intervention". *Quarterly Journal of Economics* 138.1: 151–203.

Allcott, H. (2011). "Social norms and energy conservation". *Journal of Public Economics*. Special Issue: The Role of Firms in Tax Systems 95.9: 1082–1095.

Allcott, H. and Rogers, T. (2014). "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation". *American Economic Review* 104.10: 3003–3037.

Alsan, M. and Eichmeyer, S. (2023). "Experimental Evidence on the Effectiveness of Non-Experts for Improving Vaccine Demand". *American Economic Journal: Economic Policy*.

Alsan, M., Garrick, O., and Graziani, G. (2019). "Does Diversity Matter for Health? Experimental Evidence from Oakland". *American Economic*

Review 109.12: 4071–4111.

Banerjee, A. V., Duflo, E., Glennerster, R., and Kothari, D. (2010). "Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives". *BMJ* 340: c2220.

Beaman, L., BenYishay, A., Magruder, J., and Mobarak, A. M. (2021). "Can Network Theory-Based Targeting Increase Technology Adoption?" *American Economic Review* 111.6: 1918–1943.

Benartzi, S., Peleg, E., and Thaler, R. H. (2013). "Choice Architecture and Retirement Saving Plans". *The Behavioral Foundations of Public Policy*. Princeton University Press: 245–263.

Bernheim, B. D. (1994). "A Theory of Conformity". *Journal of Political Economy* 102.5: 841–877.

Bertoni, M., Corazzini, L., and Robone, S. (2020). "The Good Outcome of Bad News: A Field Experiment on Formatting Breast Cancer Screening Invitation Letters". *American Journal of Health Economics* 6.3: 372–409.

Biesheuvel, C., Weige, S., and Heindel, W. (2011). "Mammography Screening: Evidence, History and Current Practice in Germany and Other European Countries". *Breast Care* 6.2: 104–109.

Borsky, A., Zhan, C., Miller, T., Ngo-Metzger, Q., Bierman, A. S., and Meyers, D. (2018). "Few Americans receive all high-priority, appropriate clinical preventive services". *Health Affairs* 37.6: 925–928.

Bouckaert, N., Gielen, A. C., and Van Ourti, T. (2020). "It runs in the family – Influenza vaccination and spillover effects". *Journal of Health Economics* 74: 102386.

Brunnermeier, M. K. and Parker, J. A. (2005). "Optimal Expectations". *American Economic Review* 95.4: 1092–1118.

Bursztyn, L. and Jensen, R. (2017). "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure".

Annual Review of Economics 9.1: 131–153.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors". *The Review of Economics and Statistics* 90.3: 414–427.

Campos-Mercade, P., Meier, A. N., Schneider, F. H., Meier, S., Pope, D., and Wengström, E. (2021). "Monetary incentives increase COVID-19 vaccinations". *Science* 374.6569: 879–882.

Canay, I. A., Santos, A., and Shaikh, A. M. (2021). "The Wild Bootstrap with a "Small" Number of "Large" Clusters". *Review of Economics and Statistics* 103.2: 346–363.

Carrieri, V. and Wübker, A. (2016). "Quasi-Experimental Evidence on the Effects of Health Information on Preventive Behaviour in Europe". *Oxford Bulletin of Economics and Statistics* 78.6: 765–791.

Castro, S. and Mang, C. (2024). "Breaking the silence – Group discussions and the adoption of menstrual health technologies". *Journal of Development Economics* 169: 103264.

Dafni, U., Tsourti, Z., and Alatsathianos, I. (2019). "Breast Cancer Statistics in the European Union: Incidence and Survival across European Countries". *Breast Care* 14.6: 344–353.

Dai, H., Saccardo, S., Han, M. A., Roh, L., Raja, N., Vangala, S., Modi, H., Pandya, S., Sloyan, M., and Croymans, D. M. (2021). "Behavioural nudges increase COVID-19 vaccinations". *Nature* 597.7876: 404–409.

Deutsches Mammographie-Screening-Programm (2022). "Jahresbericht Evaluation 2020". Technical Report. Kooperationsgemeinschaft Mammographie.

Drago, F., Mengel, F., and Traxler, C. (2020). "Compliance Behavior in Networks: Evidence from a Field Experiment". *American Economic Journal: Applied Economics* 12.2: 96–133.

Einav, L., Finkelstein, A., Oostrom, T., Ostriker, A., and Williams, H. (2020).

"Screening and Selection: The Case of Mammograms". *American Economic Review* 110.12: 3836–3870.

Enke, B. (2020). "What You See Is All There Is". *Quarterly Journal of Economics* 135.3: 1363–1398.

Erdmann, F., Spix, C., Katalinic, A., Christ, M., Folkerts, J., Hansmann, J., Kranzhöfer, K., Kunz, B., Manegold, K., Penzkofer, A., Treml, K., Vollmer, G., Weg-Remers, S., Barnes, B., Buttmann-Schweiger, N., Dahm, S., Fiebig, J., Franke, M., Gurung-Schönfeld, I., Haberland, J., Imhoff, M., Kraywinkel, K., Starker, A., Berenberg-Gossler, P. von, and Wienecke, A. (2021). "Krebs in Deutschland für 2017/2018". Technical Report. Robert Koch-Institut 13.

Esguerra, E., Vollmer, L., and Wimmer, J. (2023). "Influence Motives in Social Signaling: Evidence from COVID-19 Vaccinations in Germany". *American Economic Review: Insights* 5.2: 275–291.

European Commission (2022). *Cancer screening in the European Union*. Publications Office of the European Union.

Francetic, I., Meacock, R., and Sutton, M. (2022). "Understanding Concordance in Health Behaviours among Couples: Evidence from the Bowel Cancer Screening Programme in England". *Journal of Economic Behavior & Organization* 201: 310–345.

Funk, P. (2010). "Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System". *Journal of the European Economic Association* 8.5: 1077–1103.

Goldberg, J., Macis, M., and Chintagunta, P. (2023). "Incentivized Peer Referrals for Tuberculosis Screening: Evidence from India". *American Economic Journal: Applied Economics* 15.1: 259–291.

Goldzahl, L., Hollard, G., and Jusot, F. (2018). "Increasing breast-cancer screening uptake: A randomized controlled experiment". *Journal of Health Economics* 58: 228–252.

Golman, R., Hagmann, D., and Loewenstein, G. (2017). "Information Avoidance". *Journal of Economic Literature* 55.1: 96–135.

Golman, R., Loewenstein, G., Molnar, A., and Saccardo, S. (2022). "The Demand for, and Avoidance of, Information". *Management Science* 68.9: 6454–6476.

Handel, B. R. and Kolstad, J. T. (2015). "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare". *American Economic Review* 105.8: 2449–2500.

Jin, G. Z., Luca, M., and Martin, D. (2021). "Is No News (Perceived As) Bad News? An Experimental Investigation of Information Disclosure". *American Economic Journal: Microeconomics* 13.2: 141–173.

Karing, A. (2024). "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone". *Quarterly Journal of Economics* Forthcoming.

Katalinic, A., Eisemann, N., Kraywinkel, K., Noftz, M. R., and Hübner, J. (2020). "Breast cancer incidence and mortality before and after implementation of the German mammography screening program". *International Journal of Cancer* 147.3: 709–718.

Kolip, P. and Wurche, K.-D. (2005). "Mammografie Screening, Was MultiplikatorInnen vor Ort wissen sollten, Erfahrungen-Informationen-Tipps". Technical Report. Institut für Public Health und Pflegeforschung der Universität Bremen.

Kowalski, A. E. (2023). "Behaviour within a Clinical Trial and Implications for Mammography Guidelines". *Review of Economic Studies* 90.1: 432–462.

Loewenstein, G., Friedman, J. Y., McGill, B., Ahmad, S., Linck, S., Sinkula, S., Beshears, J., Choi, J. J., Kolstad, J., Laibson, D., Madrian, B. C., List, J. A., and Volpp, K. G. (2013). "Consumers' misunderstanding of health insurance". *Journal of Health Economics* 32.5: 850–862.

Marmaros, D. and Sacerdote, B. (2006). "How do friendships form?" *Quarterly*

Journal of Economics 121.1: 79–119.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). "Birds of a Feather: Homophily in Social Networks". *Annual Review of Sociology* 27.1: 415–444.

Milkman, K. L. et al. (2021). "A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment". *Proceedings of the National Academy of Sciences* 118.20: e2101165118.

Mokdad, A. H., Marks, J. S., Stroup, D. F., and Gerberding, J. L. (2004). "Actual causes of death in the United States, 2000". *Jama* 291.10: 1238–1245.

Oster, E., Shoulson, I., and Dorsey, E. R. (2013). "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease". *American Economic Review* 103.2: 804–830.

Pruckner, G. J., Schober, T., and Zocher, K. (2020). "The company you keep: health behavior among work peers". *European Journal of Health Economics* 21.2: 251–259.

Robitaille, N., House, J., and Mazar, N. (2021). "Effectiveness of Planning Prompts on Organizations' Likelihood to File Their Overdue Taxes: A Multi-Wave Field Experiment". *Management Science* 67.7: 4327–4340.

Schwardmann, P. (2019). "Motivated health risk denial and preventative health care investments". *Journal of Health Economics* 65: 78–92.

Tillmanns, H., Schillinger, G., and Dräther, H. (2021). "Inanspruchnahme von Früherkennungsleistungen der gesetzlichen Krankenversicherung durch AOK-Versicherte im Erwachsenenalter. 2009-2020". Technical Report. Wissenschaftliches Institut der AOK.

Torres, C., Ogbu-Nwobodo, L., Alsan, M., Stanford, F. C., Banerjee, A., Breza, E., Chandrasekhar, A. G., Eichmeyer, S., Karnani, M., Loisel, T., Goldsmith-Pinkham, P., Olken, B. A., Vautrey, P.-L., Warner, E., Duflo, E., and COVID-19 Working Group (2021). "Effect of Physician-Delivered COVID-19 Public Health Messages and Messages Acknowledging Racial

Inequity on Black and White Adults' Knowledge, Beliefs, and Practices Related to COVID-19: A Randomized Clinical Trial". *JAMA Network Open* 4.7: e2117115.

Tversky, A. and Kahneman, D. (1973). "Availability: A heuristic for judging frequency and probability". *Cognitive Psychology* 5.2: 207–232.

Wild, C. P., Weiderpass, E., and Bernard W. Stewart (2020). "World Cancer Report: Cancer Research for Cancer Prevention". Technical Report. International Agency for Research on Cancer. Lyon, France.

World Health Organisation (2019). "WHO's Global Health Estimates (GHE)". Technical Report.

Wübker, A. (2014). "Explaining variations in breast cancer screening across European countries". *The European Journal of Health Economics* 15.5: 497–514.

Zentrum für Krebsregisterdaten (2021). "Altersstandardisierte Krebssterblichkeit ging auch 2020 weiter zurück". Technical Report.

Appendix to Chapter 2

Tables

Table 2.5: Differences in Intervention Over Time

Trial half	Number of villages	Control villages	Treatment villages
1	11	Appointments + letters scattered	Appointments + letters bunched
2	8	Appointments scattered + letters bunched	Appointments + letters bunched

Notes: The initially planned design was implemented for the first half of the trial, 11 clusters and $n = 7,365$. A weaker version where invitation letters are received simultaneously also in the control group was implemented for the second half of the trial, 8 clusters and $n = 13,354$.

Table 2.6: Balance Table - Intervention

Variable	(1) Control N	(1) Control Mean/SE	(2) Treatment N	(2) Treatment Mean/SE	T-test Difference (1)-(2)
Went last time	10	0.564	9	0.542	0.022
		(0.015)		(0.008)	
Village population	10	8703.500	9	6829.556	1873.944
		(3028.442)		(1431.012)	
Number of invited women in village	10	1254.800	9	908.889	345.911
		(432.040)		(213.979)	
Age	10	59.376	9	59.519	-0.143
		(0.056)		(0.089)	
Distance to MMU (km)	10	6.704	9	8.653	-1.948
		(1.312)		(1.540)	
Covid 7d inc.	10	3.444	9	3.643	-0.199
		(0.306)		(0.829)	
Precipitation, in cm	10	0.418	9	0.108	0.310***
		(0.032)		(0.029)	
Avg. temperature (C)	10	13.810	9	10.556	3.254
		(2.334)		(3.140)	
Lead time	10	42.883	9	58.449	-15.566
		(5.932)		(13.826)	
School summer break	10	0.266	9	0.134	0.132
		(0.116)		(0.110)	

Notes: The value displayed for t-tests are the differences in the means across the groups, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.7: Results for First Invites - Dependent variable: Participation

	Simultaneous Appointments		Simultaneous Letters		Sim. Appointments + Letters	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.066**	-0.077***	-0.042	-0.125***	-0.088**	-0.130***
	(0.028)	(0.024)	(0.027)	(0.024)	(0.036)	(0.038)
Lead time			-0.001	-0.001	-0.000	0.000
			(0.001)	(0.001)	(0.001)	(0.001)
Age	-0.018***	-0.016***	-0.017**	-0.011	-0.022***	-0.021***
	(0.004)	(0.004)	(0.006)	(0.007)	(0.004)	(0.004)
Precipitation, in cm		0.003		-0.012		-0.000
		(0.019)		(0.020)		(0.018)
Avg. temperature (C)		0.004*		-0.002		0.006
		(0.002)		(0.003)		(0.004)
Distance to MMU (km)		0.004		-0.003		0.002
		(0.003)		(0.003)		(0.002)
School break		-0.089*		-0.101**		-0.170***
		(0.046)		(0.032)		(0.037)
Covid 7d inc.		-0.012**		-0.011**		-0.021***
		(0.005)		(0.005)		(0.005)
Constant	1.372***	1.261***	1.400***	1.272***	1.598***	1.603***
	(0.222)	(0.244)	(0.348)	(0.374)	(0.200)	(0.295)
Observations	2274	2230	1434	1396	1341	1323
R^2	0.022	0.028	0.017	0.021	0.031	0.041
Clusters	19	19	10	10	15	15

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. Columns (1) – (3) compare all control villages (*Treated*= 0) with all treatment villages (*Treated*= 1). Columns (4) - (6) compare control villages of the first half of the trial (*Treated*= 0) to control villages of the second half of the trial (*Treated*= 1). Columns (7) - (9) compare control villages of the first half of the trial (*Treated*= 0) with all treatment villages (*Treated*= 1).

Table 2.8: Peer Shares: Descriptive Statistics

Relative timing criterion	Peers	N	Mean	SD	p25	p75
Same day	d <50m	10110	0.054	0.137	0.000	0.000
	d <200m	10750	0.051	0.049	0.020	0.071
	d <500m	10793	0.050	0.032	0.032	0.063
	d <1000m	10813	0.050	0.029	0.035	0.059
Within 2 days	d <50m	10110	0.159	0.219	0.000	0.250
	d <200m	10750	0.158	0.089	0.100	0.203
	d <500m	10793	0.158	0.067	0.116	0.191
	d <1000m	10813	0.157	0.062	0.120	0.188
Up to today	d <50m	10110	0.513	0.367	0.000	0.800
	d <200m	10750	0.514	0.290	0.205	0.713
	d <500m	10793	0.512	0.284	0.209	0.705
	d <1000m	10813	0.512	0.282	0.208	0.702
The week before	d <50m	10110	0.205	0.243	0.000	0.250
	d <200m	10750	0.202	0.113	0.069	0.218
	d <500m	10793	0.201	0.093	0.083	0.210
	d <1000m	10813	0.200	0.088	0.085	0.208
7+ days earlier (Placebo)	d <50m	10110	0.308	0.347	0.000	0.545
	d <200m	10750	0.311	0.281	0.000	0.550
	d <500m	10793	0.311	0.275	0.023	0.564
	d <1000m	10813	0.312	0.273	0.029	0.566
7+ days later (Placebo)	d <50m	10110	0.329	0.339	0.000	0.571
	d <200m	10750	0.331	0.264	0.077	0.538
	d <500m	10793	0.332	0.257	0.090	0.538
	d <1000m	10813	0.332	0.256	0.095	0.534
Same day	k <0.5 years	11043	0.113	0.105	0.041	0.153
	k <1 year	11043	0.101	0.082	0.043	0.137
	k <2 years	11043	0.086	0.060	0.046	0.109
Within 2 days	k <0.5 years	11043	0.289	0.204	0.135	0.400
	k <1 year	11043	0.278	0.185	0.137	0.383
	k <2 years	11043	0.254	0.146	0.145	0.349
Up to today	k <0.5 years	11043	0.564	0.283	0.200	0.695
	k <1 year	11043	0.560	0.278	0.222	0.703
	k <2 years	11043	0.554	0.271	0.238	0.699
The week before	k <0.5 years	11043	0.321	0.203	0.067	0.328
	k <1 year	11043	0.312	0.189	0.077	0.325
	k <2 years	11043	0.297	0.168	0.087	0.315
7+ days earlier (Placebo)	k <0.5 years	11043	0.243	0.278	0.000	0.357
	k <1 year	11043	0.248	0.277	0.000	0.361
	k <2 years	11043	0.257	0.273	0.019	0.364
7+ days later (Placebo)	k <0.5 years	11043	0.229	0.226	0.049	0.337
	k <1 year	11043	0.231	0.223	0.055	0.334
	k <2 years	11043	0.236	0.222	0.060	0.344

Notes: This table presents descriptive characteristics of the *peer shares* values from section 2.5. The shares are constructed according to the set of peers and the share of them that fulfill a relative timing criterion (for an illustrative example, see Figure 2.2). The number of women, the mean of the peer share, the standard deviation, the 25th percentile, and the 75th percentile are reported. Deviations in N result from women with 0 peers according to the peer criterion.

Table 2.9: Peer Shares Estimation - Criterion: Same Day

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	-0.029	-0.046	-0.011	0.065
	(0.027)	(0.073)	(0.12)	(0.13)
Went last time	0.65***	0.66***	0.65***	0.66***
	(0.0074)	(0.0071)	(0.0071)	(0.0071)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000025
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.27***	0.26***	0.27***	0.25***
	(0.048)	(0.048)	(0.048)	(0.048)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table and the subsequent tables 2.10 to 2.20 present results for estimations of Equation 2.3. The peer criterion is given in the respective column header and the relative timing criterion is presented in the Table title.

Table 2.10: Peer Shares Estimation - Criterion: Within 2 Days

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	-0.0018	-0.021	-0.036	-0.061
	(0.017)	(0.042)	(0.057)	(0.063)
Went last time	0.65***	0.66***	0.65***	0.66***
	(0.0074)	(0.0071)	(0.0071)	(0.0071)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000024
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.27***	0.26***	0.28***	0.26***
	(0.048)	(0.048)	(0.048)	(0.048)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.11: Peer Shares Estimation - Criterion: Up to Today

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	0.0059	0.0025	0.0058	0.0031
	(0.012)	(0.016)	(0.016)	(0.017)
Went last time	0.65***	0.66***	0.65***	0.66***
	(0.0074)	(0.0072)	(0.0072)	(0.0072)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000025
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.26***	0.26***	0.26***	0.24***
	(0.055)	(0.061)	(0.062)	(0.062)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.12: Peer Shares Estimation - Criterion: 1 to 7 Days Before

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	0.050***	0.14***	0.18***	0.19***
	(0.018)	(0.035)	(0.042)	(0.045)
Went last time	0.65***	0.66***	0.66***	0.66***
	(0.0073)	(0.0071)	(0.0071)	(0.0071)
n within distance	-0.0042***	-0.00040**	-0.00013***	-0.000025
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.26***	0.23***	0.23***	0.21***
	(0.049)	(0.049)	(0.049)	(0.049)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.13: Peer Shares Estimation - Criterion: 7+ Days Earlier (Placebo)

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	-0.017	-0.025	-0.020	-0.022
	(0.013)	(0.017)	(0.017)	(0.018)
Went last time	0.65***	0.66***	0.66***	0.66***
	(0.0074)	(0.0072)	(0.0072)	(0.0072)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000024
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.30***	0.31***	0.31***	0.30***
	(0.055)	(0.060)	(0.061)	(0.061)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.14: Peer Shares Estimation - Criterion: 7+ Days Later (Placebo)

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	-0.0070	-0.00091	0.0013	0.0012
	(0.012)	(0.017)	(0.018)	(0.018)
Went last time	0.65***	0.66***	0.66***	0.66***
	(0.0074)	(0.0072)	(0.0072)	(0.0072)
n within distance	-0.0041***	-0.00040**	-0.00013***	-0.000025
	(0.0011)	(0.00017)	(0.000045)	(0.000018)
Constant	0.26***	0.26***	0.27***	0.25***
	(0.050)	(0.052)	(0.053)	(0.053)
Village FE + Controls	Yes	Yes	Yes	Yes
Observations	10110	10750	10793	10813

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.15: Peer Shares Estimation - Criterion: Same Day

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	0.043	0.018	0.056
	(0.044)	(0.057)	(0.080)
Went last time	0.66***	0.66***	0.66***
	(0.0074)	(0.0074)	(0.0073)
n within age diff.	0.000062	-0.0000030	0.000021
	(0.00019)	(0.00010)	(0.000048)
Constant	0.18*	0.22**	0.18*
	(0.10)	(0.11)	(0.10)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.16: Peer Shares Estimation - Criterion: Within 2 Days

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	-0.0059	-0.018	-0.017
	(0.021)	(0.024)	(0.029)
Went last time	0.66***	0.65***	0.65***
	(0.0073)	(0.0074)	(0.0074)
n within age diff.	0.00000086	-0.000032	0.0000059
	(0.00019)	(0.00010)	(0.000049)
Constant	0.24**	0.28***	0.23**
	(0.100)	(0.11)	(0.10)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.17: Peer Shares Estimation - Criterion: Up to Today

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	-0.0071	-0.00012	0.0036
	(0.013)	(0.014)	(0.014)
Went last time	0.66***	0.66***	0.66***
	(0.0072)	(0.0072)	(0.0072)
n within age diff.	0.000015	-0.000011	0.000013
	(0.00018)	(0.000097)	(0.000048)
Constant	0.22**	0.24**	0.20**
	(0.089)	(0.094)	(0.095)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.18: Peer Shares Estimation - Criterion: 1 to 7 Days Before

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	0.048**	0.062**	0.087***
	(0.023)	(0.024)	(0.026)
Went last time	0.66***	0.66***	0.66***
	(0.0071)	(0.0071)	(0.0071)
n within age diff.	0.000036	0.000011	0.000035
	(0.00018)	(0.000097)	(0.000048)
Constant	0.21**	0.21**	0.15
	(0.090)	(0.095)	(0.096)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.19: Peer Shares Estimation - Criterion: 7+ Days Earlier (Placebo)

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	-0.027*	-0.021	-0.021
	(0.015)	(0.015)	(0.015)
Went last time	0.66***	0.66***	0.66***
	(0.0073)	(0.0073)	(0.0073)
n within age diff.	0.000034	-0.0000031	0.000015
	(0.00018)	(0.000097)	(0.000048)
Constant	0.22**	0.24**	0.21**
	(0.089)	(0.094)	(0.095)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Table 2.20: Peer Shares Estimation - Criterion: 7+ Days Later (Placebo)

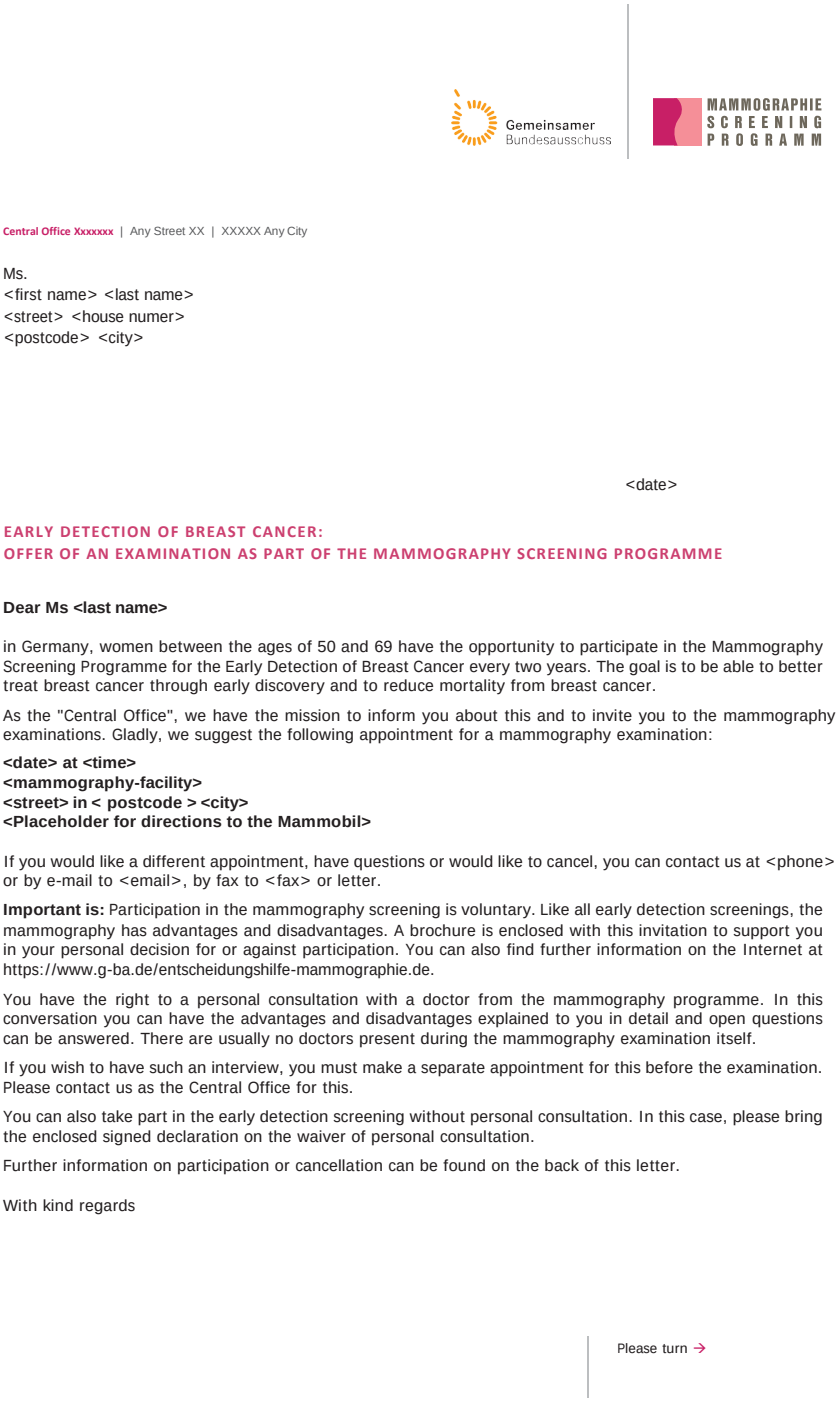
	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	0.015	0.013	0.0075
	(0.017)	(0.017)	(0.017)
Went last time	0.66***	0.66***	0.66***
	(0.0071)	(0.0071)	(0.0071)
n within age diff.	-0.0000044	-0.000021	0.0000098
	(0.00019)	(0.000098)	(0.000048)
Constant	0.23***	0.25***	0.21**
	(0.090)	(0.095)	(0.095)
Village FE + Controls	Yes	Yes	Yes
Observations	10817	10817	10817

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: See Table 2.9 Notes.

Figures



Gemeinsamer Bundesausschuss

MAMMOGRAPHIE SCREENING PROGRAMM

Central Office Xxxxxxx | Any Street XX | XXXXX Any City

Ms.
<first name> <last name>
<street> <house numer>
<postcode> <city>

<date>

EARLY DETECTION OF BREAST CANCER:
OFFER OF AN EXAMINATION AS PART OF THE MAMMOGRAPHY SCREENING PROGRAMME

Dear Ms <last name>

in Germany, women between the ages of 50 and 69 have the opportunity to participate in the Mammography Screening Programme for the Early Detection of Breast Cancer every two years. The goal is to be able to better treat breast cancer through early discovery and to reduce mortality from breast cancer.

As the "Central Office", we have the mission to inform you about this and to invite you to the mammography examinations. Gladly, we suggest the following appointment for a mammography examination:

<date> at <time>
<mammography-facility>
<street> in < postcode > <city>
<Placeholder for directions to the Mammobil>

If you would like a different appointment, have questions or would like to cancel, you can contact us at <phone> or by e-mail to <email>, by fax to <fax> or letter.

Important is: Participation in the mammography screening is voluntary. Like all early detection screenings, the mammography has advantages and disadvantages. A brochure is enclosed with this invitation to support you in your personal decision for or against participation. You can also find further information on the Internet at https://www.g-ba.de/entscheidungshilfe-mammographie.de.

You have the right to a personal consultation with a doctor from the mammography programme. In this conversation you can have the advantages and disadvantages explained to you in detail and open questions can be answered. There are usually no doctors present during the mammography examination itself.

If you wish to have such an interview, you must make a separate appointment for this before the examination. Please contact us as the Central Office for this.

You can also take part in the early detection screening without personal consultation. In this case, please bring the enclosed signed declaration on the waiver of personal consultation.

Further information on participation or cancellation can be found on the back of this letter.

With kind regards

Please turn →

MAMMOGRAPHIE SCREENING PROGRAMM

IF YOU DO NOT WISH TO PARTICIPATE

You are entitled to this offer every two years. If you do not wish to participate this time, we will contact you again in two years.

If you do not wish to receive any further invitations, please inform us in text form by fax to <fax>, by letter to <address> <postcode> <city> or by e-mail to <email>. If you change your decision later, please inform us. We will then send you a new invitation.

If you do not participate, you will not suffer any disadvantages in terms of health insurance and care. Even if you should develop breast cancer at some point, your health insurance will of course cover the treatment costs.

INFORMATION ON PARTICIPATION - PLEASE NOTE IN ADVANCE

The costs of the examination are covered by your statutory health insurance. A referral is not necessary. If you are privately insured, please clarify the cost absorption with your insurance company in advance.

- The mammography screening is for women who have no signs of breast disease.
- If you have already had a mammography screening examination within the last 22 months or have had a mammography within the last 12 months for other reasons (e.g. after breast cancer), please let us know in advance.
- If you need help or are dependent on a wheelchair, please contact us in advance, as the central office.

ON THE DAY OF THE EXAMINATION - PLEASE NOTE

Please bring your insurance card, this invitation letter and the completed questionnaire. If you do not want a personal consultation, also the signed waiver.

Please do not use powder, deodorant or cream on the chest and underarm area on the day of the examination as this may interfere with the X-ray images.

THE RESULT OF THE MAMMOGRAM

Mammography is used to look for abnormalities that indicate breast cancer. You will usually be informed by the mammography unit within seven working days whether such abnormalities have been found or not. If abnormalities are found, this does not mean that it has to be breast cancer. In most cases, the suspicion can be disproved. However, further examinations are necessary. You will then receive another invitation.

PRIVACY

Your address was provided to us by your municipality in accordance with the legal requirements for data protection. The protection of your data is ensured at all times. Your examination results are only available at the mammography facility and are subject to medical confidentiality. Further information on the use of your data can be found in the enclosed brochure.

DECLARATION ON THE WAIVER OF THE PERSONAL CONSULTATION

I have been informed of the main advantages and disadvantages of the mammography screening programme by the enclosed documents and waive my right to an additional personal consultation with a doctor of the programme before the examination.

<first name> <last name>, born on <date of birth>

Date | Signature

Figure 2.6: Invitation Letter (translated), Page 2

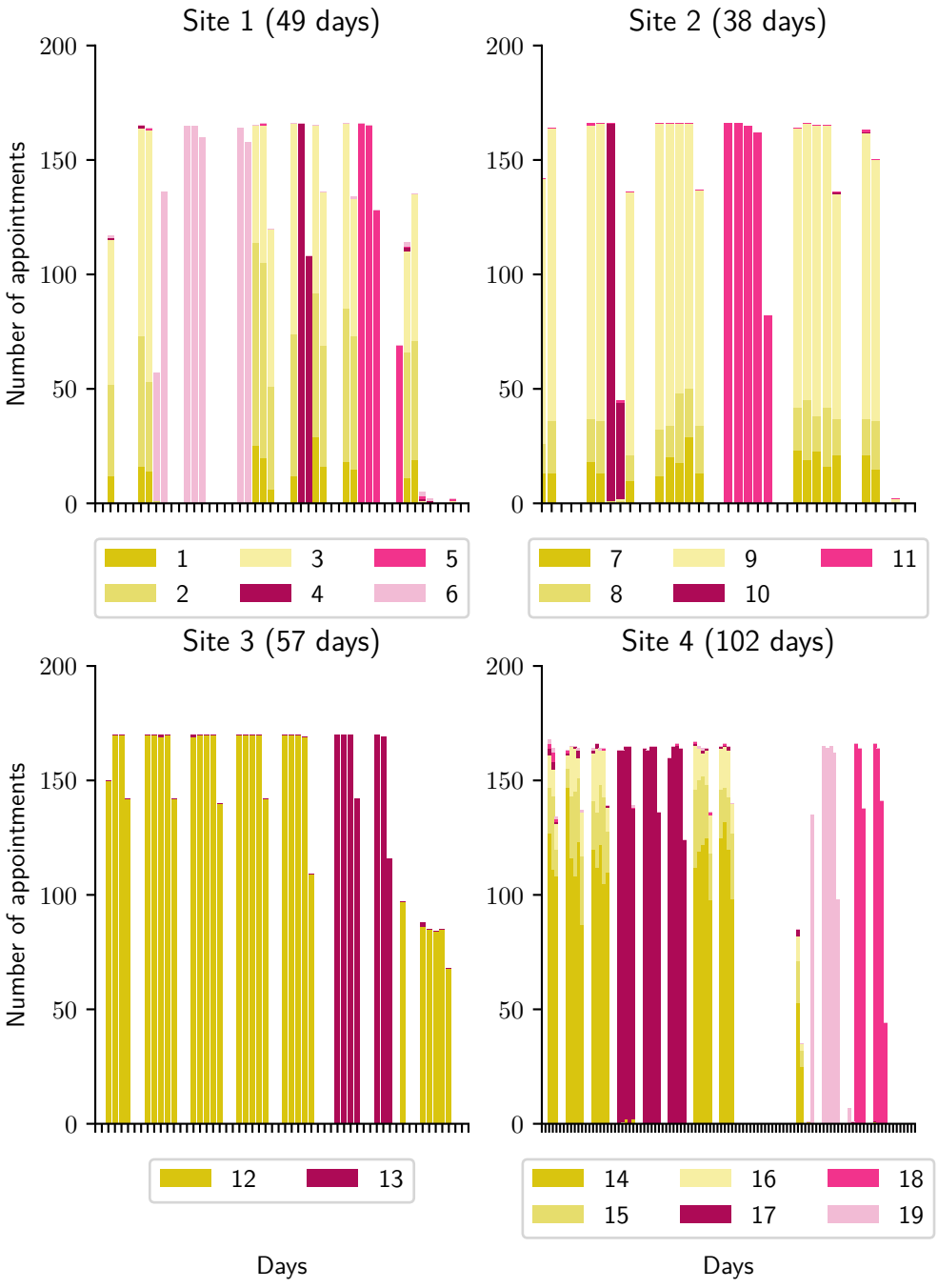


Figure 2.7: Manipulation of Initial Appointments

Notes: Control villages are depicted in green shades and treatment villages are depicted in red shades. The average standard deviation of initial appointment dates weighted by village size in the treatment (control) group is 5.24 (14.71).

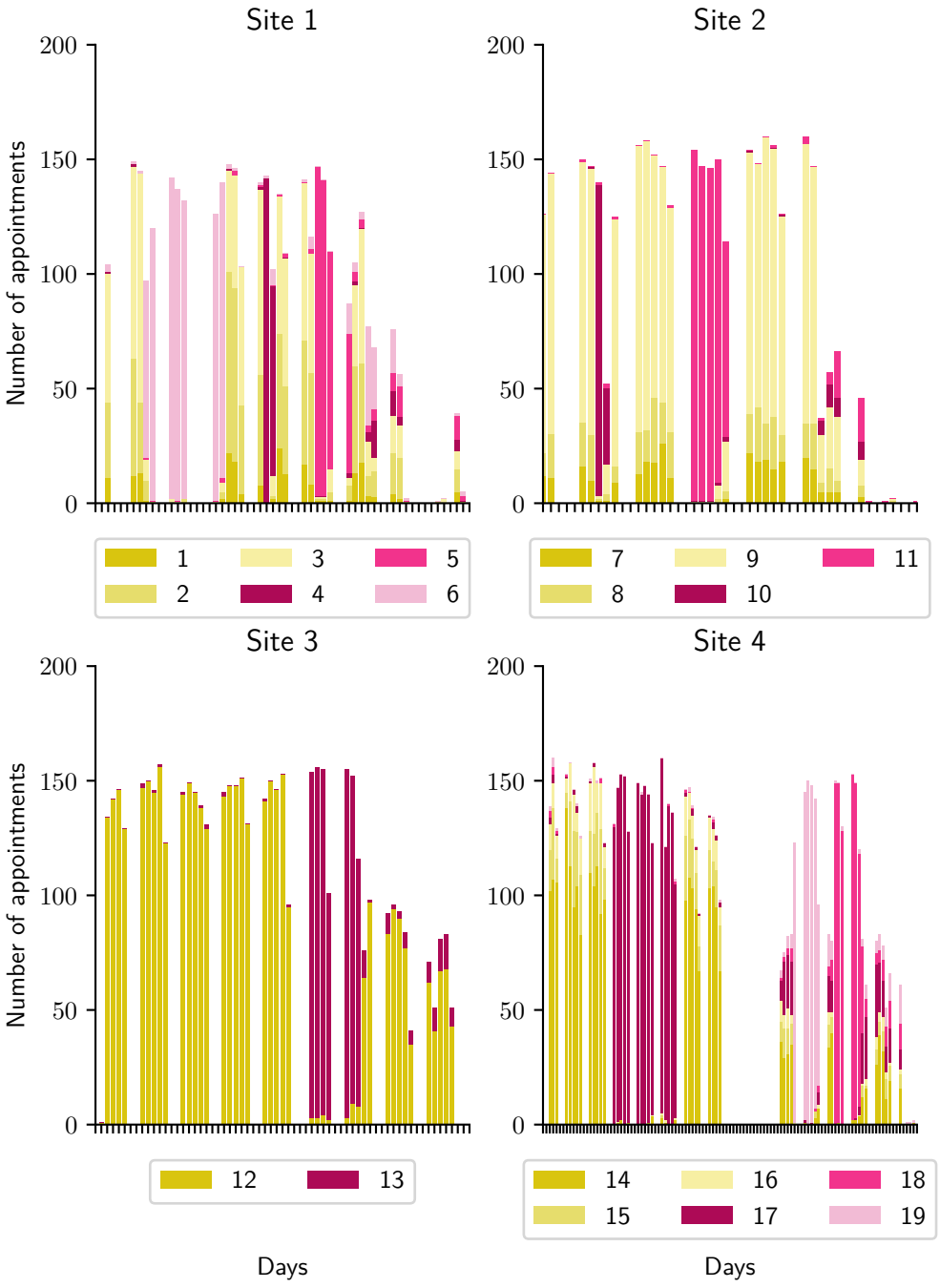


Figure 2.8: Manipulation of Final Appointments

Notes: Control villages are depicted in green shades, and treatment villages are depicted in red shades. The average standard deviation of final appointment dates weighted by village size in the treatment (control) group is 15.13 (21.01).

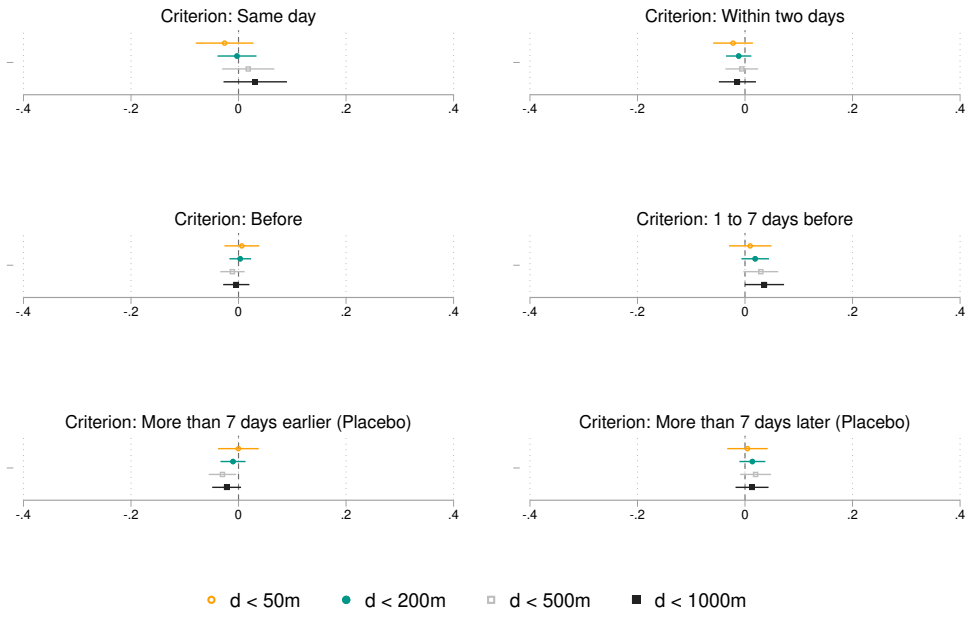


Figure 2.9: Peer Shares by Geographic Distance, Women with <0.5 Years Age Difference

Notes: This figure presents resulting β_1 from estimations of Equation 2.3 which can be interpreted as the change in participation in % given an increase in $share_i$ by 1%. Each point estimate and its 95% confidence bands represent one regression that includes village-fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 2.4). Each bar per panel represents a different peer criterion, here by distance in meters (see Table 2.3) and age difference, as only women with <0.5 years age difference are considered.

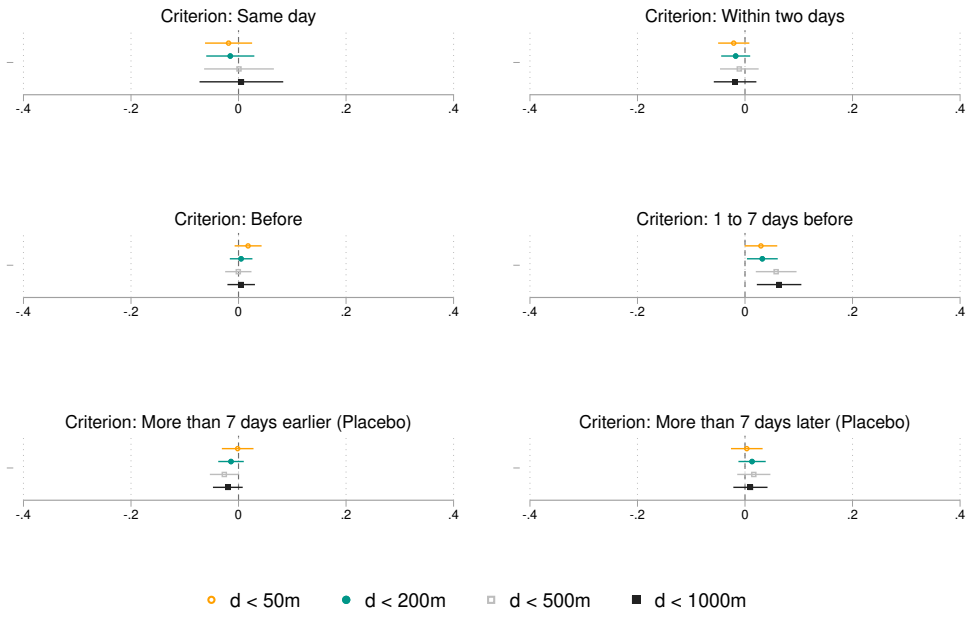


Figure 2.10: Peer Shares by Geographic Distance, Women with <1 Year Age Difference

Notes: This figure presents resulting β_1 from estimations of Equation 2.3 which can be interpreted as the change in participation in % given an increase in $share_i$ by 1%. Each point estimate and its 95% confidence bands represent one regression that includes village-fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 2.4). Each bar per panel represents a different peer criterion, here by distance in meters (see Table 2.3) and age difference, as only women with <1 year age difference are considered.

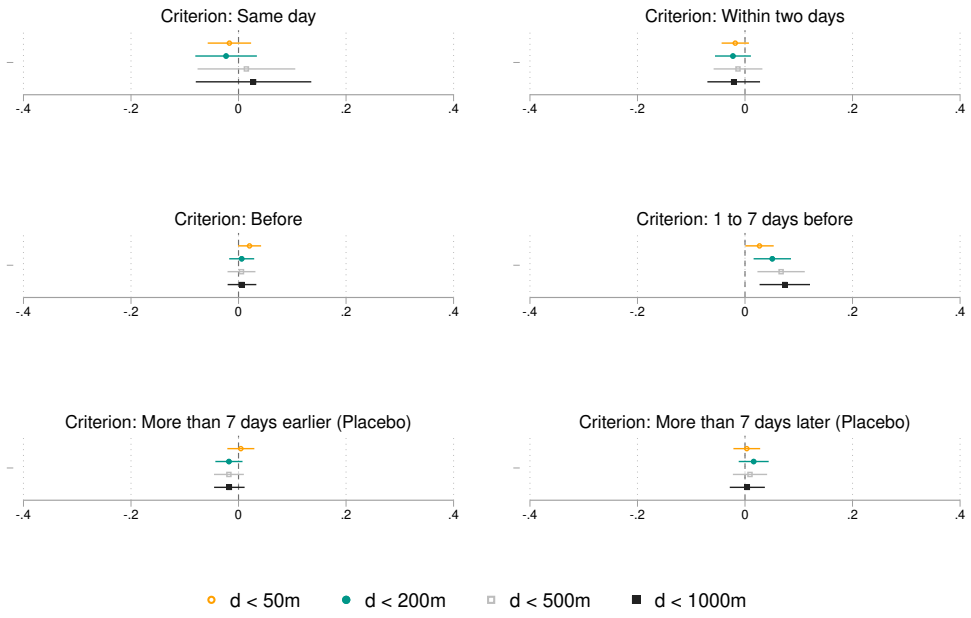


Figure 2.11: Peer Shares by Geographic Distance, Women with <2 Year Age Difference

Notes: This figure presents resulting β_1 from estimations of Equation 2.3 which can be interpreted as the change in participation in % given an increase in $share_i$ by 1%. Each point estimate and its 95% confidence bands represent one regression that includes village-fixed effects and controls for previous participation, the size of the peer group, distance to the mammography unit, and age. The header of each panel indicates the relative timing criterion (see Table 2.4). Each bar per panel represents a different peer criterion, here by distance in meters (see Table 2.3) and age difference, as only women with <2 years of age difference are considered.

Spatial Correlation and Social Determinants

Spatial correlation

This section presents additional findings that we did not pre-register. First, we implement a variation of the *peer shares* approach from Section 2.5. We estimate the association between *participation rates* of peers and own participation conditional on a set of covariates. Again, geographical distance and time between birth dates are the criteria that we use flexibly to proxy relevant peers (see Table 2.3). The criterion used to calculate the *participation peer share* is now different: It is simply the share of peers that participated in the screening during our study period and thus not related to the relative timing of invitation dates.

This exercise should not be interpreted as causal as there is no plausibly exogenous variation in the participation rates of peers. It also represents a case of the reflection problem (Manski 1993). We estimate equations of the following linear probability model

$$Y_{iv} = \beta_0 + \beta_1 share_i + \eta_i + \pi_v + X\beta + \epsilon_i. \tag{2.4}$$

η_i represents invitation-type effects as our sample now also includes individuals who were invited for the first time. We further control for village fixed effects, distance to the site, age, and the number of peers. Our coefficient of interest, β_1 coefficients, can be interpreted as follows: An increase in the participation rate of the peers by 1% is associated with a change in their own participation probability of β_1%. Due to the village fixed effects, our estimates are always relative to the average woman in the village.

The estimates reported in Table 2.21 are positive and significant for all distances other than 1000m (for which the point estimate is also positive). This represents a significant positive spatial autocorrelation within villages.

Table 2.21: Participation Peer Shares - Peers by Location

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
Share	0.047***	0.11***	0.20***	0.074
	(0.0095)	(0.025)	(0.042)	(0.058)
Age	-0.0033***	-0.0032***	-0.0032***	-0.0031***
	(0.00058)	(0.00056)	(0.00056)	(0.00056)
n within distance	-0.0041***	-0.00020	-0.000083**	-0.000032**
	(0.00088)	(0.00014)	(0.000035)	(0.000015)
Constant	0.56***	0.51***	0.49***	0.54***
	(0.034)	(0.035)	(0.039)	(0.044)
Village + Inv. Type FE	Yes	Yes	Yes	Yes
Observations	18970	20315	20388	20421

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents results from estimating Equation 2.4 with variations of the peer criterion distance (in meters).

In a similar vein, we use address data to construct spatial grids of all villages. We construct square grids with sides 100 meters in length and calculate the participation rates in a given grid cell. For each grid cell, we define all adjacent grid cells to the north, south, west, and east as neighboring cells. We then calculate the Moran's I statistic that tests for spatial autocorrelation between participation rates against the null hypothesis of random spatial autocorrelation.[28]

Table 2.22 shows the resulting p-values. These p-values can be interpreted as the probability of observing a spatial pattern as extreme as the one found in the data if the null hypothesis is true. We also impose different conditions on the minimum number of women within a grid cell to be regarded in the calculation of Moran's I. The null hypothesis of no spatial autocorrelation at a very close geographical level can thus be rejected.

Next, we test for age-based autocorrelation in participation. We estimate versions of Equation 2.4 where the peer group is now defined as all individuals

[28]We cannot present the heat maps of participation rates for data protection reasons.

Table 2.22: Spatial Autocorrelation in Participation: Moran's I

Minimum n	Grid cells	Moran's I	p-value
1	5547	0.005	0.389
2	4084	0.039	0.032
3	3031	0.052	0.022
4	2220	0.069	0.018
5	1595	0.056	0.085
6	1131	0.105	0.022
7	807	0.074	0.128
8	560	0.057	0.248
9	391	0.041	0.351
10	280	0.046	0.358

Notes: This table presents the results for the calculation of Moran's I for the participation rate across adjacent pairs of grid cells. The grid cells are 100 meters x 100 meters. In each row, all grid cells with less than the minimum n of women per grid cell are disregarded, thus gradually reducing the number of grid cells and, accordingly the number of pairs. The reported p-values are presented against an assumed random distribution of participation rates.

whose age differs by a maximum of $\bar{k}$ years. The coefficients can be interpreted as the predicted difference of the individual's participation probability compared to the average woman in a village conditional on controls.

Results are reported in Table 2.23. The point estimates are indistinguishable from zero, indicating no age-based autocorrelation within villages. They also do not change much depending on the specified maximum age distance.

Finally, we combine the criteria geographical distance $\bar{d}$ and age distance k. Now, the relevant peers in the denominator of the $share_i$ variable from Equation 2.4 are individuals who live within distance $\bar{d}$ and whose birth date is less than a specified number of $\bar{k}$ months away from the individual's birth date. Table 2.24 presents the β_1 estimates for different combinations of $\bar{d}$ and $\bar{k}$. The estimates are consistently positive and significant, stressing the

Table 2.23: Participation Peer Shares - Peers by Age

	(1) 0.5 years	(2) 1 year	(3) 2 years
Share	-0.013	-0.014	-0.030
	(0.042)	(0.059)	(0.084)
Age	-0.0035***	-0.0038***	-0.0039***
	(0.00067)	(0.00070)	(0.00073)
n within age relation	-0.00017	-0.00015*	-0.000089**
	(0.00016)	(0.000081)	(0.000039)
Constant	0.61***	0.65***	0.67***
	(0.062)	(0.063)	(0.061)
Village + Inv. Type FE	Yes	Yes	Yes
Observations	20432	20432	20432

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents results from estimating Equation 2.4 with variations of the peer criterion age difference (in years).

strength of the spatial correlation. The effect sizes are comparable to those in Table 2.21.

Socioeconomic status and social capital

Finally, we evaluate associations of participation rates with additional variables that relate to socioeconomic status and social capital. First, we include voting data. Specifically, we use the turnout in the previous national election as a measure of political participation. Second, we use the local unemployment rate as a proxy for regional socioeconomic deprivation. We also calculate the share of the adult population that is registered as catholic or protestant Christians, which we interpret as a proxy for social capital (Strømsnes 2008; Traunmüller 2011).

At the individual level, we also use the number of other invited women who live within 50 meters as a measure of the population density of the individual's immediate neighborhood. As larger lot sizes are negatively associated with

Table 2.24: Participation Peer Shares - Peers by Distance and Age

	(1) 50m	(2) 200m	(3) 500m	(4) 1000m
0.5 years	0.0260**	0.023***	0.027**	0.023
	(0.013)	(0.008)	(0.013)	(0.017)
1 year	0.020**	0.026***	0.072***	0.083***
	(0.010)	(0.010)	(0.017)	(0.023)
2 years	0.025***	0.032***	0.091***	0.096***
	(0.009)	(0.013)	(0.023)	(0.031)

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents results from estimating Equation 2.4 with variations of the peer criterion across two dimensions, i.e. by both geographical distance (column) and age difference (row). Only β_1 estimates are reported. Control variables and fixed effects are the same as in Tables 2.21 and 2.23.

this proxy - especially in a non-urban setting - this serves as a proxy for socioeconomic status. Other research has used lot sizes and property values as a proxy for SES (Juhn et al. 2011; Ware 2019). Lower SES women are also presumably more likely to live in the same apartment building or block as others, while higher SES women tend to live in larger properties.[29]

Table 2.25 shows that women who live in more densely populated areas are less likely to participate, suggesting a positive relationship between socioeconomic status and participation at the individual level. For the regional variables, only the share of registered Christians is positively associated with participation. This is in line with Salmon et al. (2022) reporting a positive association between church attendance and screening in the United States. Women in regions with more of this type of social capital are seemingly more likely to participate in the screening.

[29]A SES gradient in mammography participation has been documented in other settings, e.g. by Khan et al. (2021) in Australia or Lemke et al. (2015) in Germany. Smith et al. (2019) provide an overview.

Table 2.25: Heterogeneity in Participation Rates

	(1)	(2)	(3)	(4)
SES proxy (density)	-0.390***		-0.340**	-0.387**
	(0.094)		(0.102)	(0.103)
Unemployment (per cent)		0.010	0.010	0.002
		(0.014)	(0.014)	(0.015)
Election turnout		0.001	-0.013	0.054
		(0.108)	(0.110)	(0.094)
Share of registered Christians		0.181*	0.122	0.176*
		(0.065)	(0.075)	(0.066)
Constant	0.410***	0.239	0.309*	0.277*
	(0.018)	(0.115)	(0.121)	(0.112)
Invitation-type FE	Yes	Yes	Yes	Yes
Additional controls	No	No	No	Yes
Observations	20719	20719	20719	20432
R-squared	0.381	0.381	0.381	0.384

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents results from regressing individual participation on a set of individual and village-level variables. The SES proxy (density) is defined as the number of other invited women who live within 50 meters. Unemployment, election turnout, and the share of registered Christians is available at the village level. Additional individual-level control variables include weather, distance to the site, school breaks, and COVID-19 7-day incidence.

Surveys

During the first half of the trial, we distributed an on-site paper-based survey among BCS participants (Survey 1). During the second half of the trial, we attached a QR code for an online survey (Survey 2) to the invitation letter.[30] In both cases, the treatment status of respondents is inferred by the self-reported residence village. Table 2.26 details sample sizes and response rates.

Table 2.26: Surveys - Overview

			Response rates				
	N	Sample	Overall	Treatment	Control	Mode	BCS participation
Survey 1	946	BCS participants	0.27	0.4	0.18	on-site	1
Survey 2	629	BCS invitees	0.1	0.1	0.11	online	0.9

Notes: For Survey 2, BCS participation is self-reported (intended) participation. Prior to the intervention, we ran an exploratory online survey ($n = 170$) with a separate and representative sample. All survey materials are in the subsequent Appendix section.

At least two assumptions have to hold to be able to attribute any differences in a survey item by treatment status to our intervention. First, there are no baseline differences by treatment status with respect to the survey item. Second, there is no differential selection by treatment in taking the survey. Both assumptions are likely to be violated in our setting.[31] Thus, the following results are to be interpreted as consistent with, but not necessarily confirming, our approach.

[30]Since the medical on-site personnel reported that administering the on-site survey (Survey 1) interrupted their workflow, we were unable to extend the paper-based survey to the second half of the trial and employed the online survey (Survey 2) instead.

[31]Due to the low number of clusters (i.e. villages) and strong variation of survey participation rates over time (see Figure 2.12), both assumptions are likely to be violated despite randomized treatment assignment.

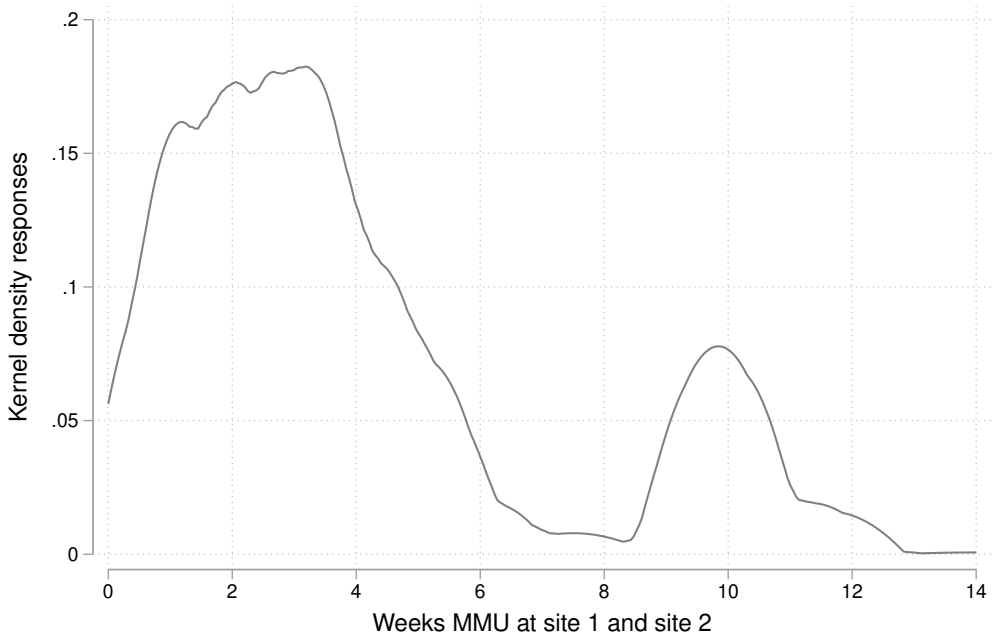


Figure 2.12: Survey 1 Responses Over Time

Notes: The 946 survey 1 responses were all collected at sites 1 and 2. Most of the answers were collected in the first five weeks.

Survey 1

Survey 1 elicited an individual's previous participation in breast cancer check-ups - within and outside of the BCS program, migration history, education, perceptions about acquaintances' participation in BCS, mode of transport to the appointment, and whether or not she would like to attend her appointments more regularly or not.

Table 2.27 documents that only migration status differs by treatment status, which likely reflects baseline differences. The positive signs of the coefficients in columns (1) and (2) indicate that, as intended, our intervention somewhat increased knowledge about how many acquaintances are participating in the BCS. Column (1) shows that in the treatment group, 3 percentage points more women report knowing what share of their acquaintances are participating in the BCS (p-value = 0.18). Interestingly, respondents perceive that roughly

Table 2.27: Survey 1

	(1) Perceived participation (binary)	(2) Perceived participation	(3) Car pooling	(4) Education	(5) Migration status
Treatment	0.038	0.059	0.000	0.010	-0.070**
	(0.027)	(0.091)	(0.010)	(0.046)	(0.024)
Constant	0.626***	3.281***	0.033***	2.992***	2.016***
	(0.015)	(0.036)	(0.009)	(0.039)	(0.016)
Observations	849	550	908	848	895
R^2	0.002	0.001	0.000	0.000	0.002
Clusters	11	11	11	11	11

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. *Perceived participation (binary)* takes on the value of 0 if and only if the respondent states that she has no idea how many of her acquaintances are participating in the program and 1 otherwise. *Perceived participation* is defined only for women for who *perceived participation (binary)* = 1. It takes on values from 0 to 4, where 0 means that a respondent believes almost no one of her acquaintances to participate, and with step sizes of 1/4 of acquaintances, the maximum value of 4 means that she believes almost all of her acquaintances to participate. *Car pooling* is a binary variable taking on the value of 1 if a respondent was carpooling in order to reach the check-up site. *Education* is constructed from an 8-point Likert scale with higher values corresponding to higher levels of education. The variable *migration status* takes on the value of 1 if the respondent was born in the village she currently resides in and a maximum value of 4 if she was born outside of Germany.

75% of their acquaintances participate in the BCS compared to an overall participation rate of 47%. Since all Survey 1 respondents are BCS participants, this suggests a strong correlation in peers' participation behavior, which is consistent with previous results.

Our intervention may have lowered the costs of reaching the check-up site by making carpooling more feasible. We document, however, that baseline rates of carpooling are very low at 3% and are not affected by our treatment. There is a significant and negative correlation (p-value < 0.01) between having participated in a BCS check-up and a check-up outside of the BCS (i.e. opportunistic screening), suggesting that BCS reaches women who otherwise would not proactively engage in breast cancer prevention.

Survey 2

The vast majority of Survey 2 responses (83%) were collected prior to the appointment and almost 90% of respondents were either planning to go to their check-up or reported having gone already. In addition to all questions included in Survey 1, we asked respondents how they perceived others' willingness to talk about breast cancer prevention check-ups.

First, we confirm two findings from Survey 1. Only 0.4% of survey takers are carpooling to reach the MMU.[32] We again document strong substitutability between BCS and opportunistic screening (p-value $< .01$). Notably, 22% of Survey 2 respondents say that they would like to attend the biennial check-ups more regularly. This implies that the policy goal of increasing participation rates is aligned with individual interests and that an intervention to target these individuals could be welfare-increasing.

Since most respondents take the survey right after receiving the letter, i.e. before they could be affected by our intervention, we now differentiate between pre-appointment and post-appointment responses. The results are

[32] A vast majority of 70% reports going alone by car.

Table 2.28: Survey 2

	(1) Car pooling	(2) Education	(3) Migration status
Treatment	0.008	-0.494	-0.171
	(0.006)	(0.367)	(0.131)
Constant	0.000**	3.862***	2.074***
	(0.000)	(0.358)	(0.106)
Observations	627	627	627
R^2	0.005	0.011	0.010
Clusters	8	8	8

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. *Car pooling* is a binary variable taking on the value of 1 if a respondent was carpooling in order to reach the check-up site. *Education* is constructed from an 8-point Likert scale with higher values corresponding to higher levels of education. The variable *migration status* takes on the value of 1 if the respondent was born in the village she currently resides in and a maximum value of 4 if she was born outside of Germany.

displayed in Table 2.28 - 2.30.

In line with our exploratory survey, breast cancer prevention is not a stigmatized topic: 88% of respondents believe that others talk openly about their participation in BCS. This belief is held more often in the treatment group (2.8% difference, p-value = 0.01) and is more pronounced, although not significant when restricting the sample to post-appointment responses (see Table 2.29). We interpret this increased openness towards conversations about breast cancer prevention check-ups as a sign that, in our setting, peers can be used to disseminate public health information.

Consistent with both the exploratory survey and Survey 1, only 38% of respondents state that they have no idea how many of their peers are participating in the BCS.

When restricting the analysis to post-appointment responses, we find that our intervention significantly increased the share of women reporting knowledge of their acquaintances' participation behavior (p-value = 0.06, see

Table 2.29: Survey 2 - Conversations

	(1) Belief in talking openly	(2) Pre belief in talking openly	(3) Post belief in talking openly
Treatment	0.028**	0.023	0.048
	(0.008)	(0.014)	(0.067)
Constant	0.867***	0.871***	0.848***
	(0.006)	(0.004)	(0.024)
Observations	627	519	108
R^2	0.002	0.001	0.004
Clusters	8	8	8

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. *Belief in talking openly* is a binary variable taking on the value of 1 if and only if the respondent states that she believes women are talking openly about their participation in the program and 0 otherwise. Column (2) restricts the analysis of *belief in talking openly* to responses that were collected prior to the respondent's appointment. Column (3) restricts the analysis of *belief in talking openly* to responses that were collected after the respondent's appointment.

Table 2.30). This suggests that our intervention has given participants a better understanding of their peers' behavior.

Table 2.30: Survey 2 - Perceptions

	(1) Perceived participation (binary)	(2) Pre perceived participation (binary)	(3) Post perceived participation (binary)	(4) Perceived participation
Treatment	0.076	0.056	0.180*	-0.169
	(0.048)	(0.053)	(0.080)	(0.092)
Constant	0.595***	0.608***	0.544***	3.289***
	(0.044)	(0.049)	(0.061)	(0.040)
Observations	627	519	108	391
R^2	0.006	0.003	0.026	0.007
Clusters	8	8	8	8

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the village level. *Perceived participation (binary)* takes on the value of 0 if and only if the respondent states that she has no idea how many of her acquaintances are participating in the program and 1 otherwise. Column (2) restricts the analysis of *perceived participation (binary)* to responses that were collected prior to the respondent's appointment. Column (3) restricts the analysis of *perceived participation (binary)* to responses that were collected after the respondent's appointment. *Perceived participation* is defined only for women for who *perceived participation (binary)* = 1. It takes on values from 0 to 4, where 0 means that a respondent believes almost no one of her acquaintances to participate, and with step sizes of 1/4 of acquaintances, the maximum value of 4 means that she believes almost all of her acquaintances to participate.

Survey Material

Paper based survey (Survey 1)

I Preliminaries

We are scientists at the Ludwig-Maximilians-University Munich and do research on the topic of breast cancer screening. We would be very pleased if you answered the questions below.
Please be assured that your data will be treated confidentially and will not be traced back to you. By consenting, you confirm that you are of legal age and that you know that filling out the questionnaire is voluntary.

II Residence

Question 1: Which village do you reside in?

Question 2: Where were you born?
Reply options:

- *In my current place of residence.*
- *In [federal state], but not in my current place of residence.*
- *In Germany but outside of [federal state].*
- *Outside of Germany.*

III Mammography

Question 1: Have you ever gone for breast cancer screening?
Reply options (multiple answers possible):

- *Yes, like today, as part of the mammography screening program.*
- *Yes, in a gynecology or radiology office.*
- *No.*

Question 2: Would you like to go for breast cancer screening *more often, as often, or less often* than you have in the past?
Reply options: *more often, as often, less often*

Question 3: Do you have a rough idea of how many women in your circle of acquaintances participate in the mammography screening program?
Reply options: Relative shares *1, 0.75, 0.5, 0.25, 0, don't know*

Question 4: How did you get to the examination today?
Reply options:

- *Alone by car.*
- *In a carpool.*
- *By public transport.*
- *On foot or by bike.*

IV Demographics

Question 1: In which year were you born?

Question 2: What is your highest high school or college degree?
Reply options:

- *Without a general school leaving certificate.*
- *Without a general school leaving certificate.*
- *Secondary school diploma (Realschulabschluss).*
- *High school diploma or equivalent (Abitur).*
- *Apprenticeship.*
- *Study without a degree.*
- *Preliminary diploma (Vordiplom).*
- *Diploma (Diplom).*
- *Ph.D.*

V Closing Questions

Prompt 1: Please enter today's date (day, month).

Prompt 2: Space for general comments.

Online survey (Survey 2)

I Preliminaries

Welcome to this survey. We are scientists from the University of Munich and do research on the topic of breast cancer prevention. We kindly ask you to answer the following questions. This shouldn't take more than 15 minutes of your time.

Please be assured that your data will be treated confidentially and cannot be traced back to you. By agreeing, you confirm that you are of legal age and understand that participation in the survey is voluntary.

II Context

You have been invited to breast cancer screening as part of the mammography screening program. (The invitation included the link to this survey.)

Question: Are you planning to go for breast cancer screening this year as part of the mammography screening program? Which is most applicable to you?
Reply options:

- *I plan to attend.*
- *I plan not to attend.*
- *I'm still undecided.*
- *I have already participated.*
- *I have not already participated.*

III Social Questions

Question 1: Have you ever gone for breast cancer screening?
Reply options (multiple answers possible):

- *Yes, as part of the mammography screening program.*
- *Yes, at the family doctor or in a gynecology or radiology practice.*
- *No.*

Question 2: Did you go for breast cancer screening in 2020 as part of the mammography screening program?
Reply options: *yes, no, don't remember*

Question 3: Would you say that women talk openly about participating in the mammography screening program?
Reply options: *yes, rather yes, rather no, no*

Question 4: Do you have a rough idea of how many women in your circle of acquaintances participate in the mammography screening program?
Reply options: Relative shares *1, 0.75, 0.5, 0.25, 0, don't know*

Question 5: Would you like to go to breast cancer screening *more often, as often or less often* than you have done in the past?
Reply options: *more often, as often, less often*

IV Questions on Participation

A-E is chosen depending on the answer in II.

IV.A Attendance Planned

Question 1: Why do you want to participate in the mammography screening program this year?

Question 2: If you are participating in the mammography screening program this year, how do you plan to get to the screening?

Reply options:

- *Alone by car.*
- *In a carpool.*
- *By public transport.*
- *On foot or by bike.*
- *Don't know.*

IV.B Attendance Not Planned

Question Why don't you want to participate in the mammography screening program this year?

IV.C Undecided

Question 1: Why are you still undecided whether or not to participate in the mammography screening program this year?

Question 2: If you are participating in the mammography screening program this year, how do you plan to get to the screening?
Reply options: See IV.A

IV.D Participated

Question 1: Why did you participate in the mammography screening program this year?

Question 2: How did you get to your examination within the mammography screening program?
Reply options: See IV.A

IV.E Didn't Participate

Question: Why didn't you participate in the mammography screening program this year?

V Social Potency

We now ask you to fill out a short personality questionnaire. If you don't have time for this, you can skip this part.

For each statement, please indicate to what extent it applies to you.
Reply options: *true, false*

- **Statement 1:** I am quite effective at talking people into things.
- **Statement 2:** I am very good at influencing people.
- **Statement 3:** I do not like to be the center of attention on social occasions.
- **Statement 4:** I do not enjoy trying to convince people of something.
- **Statement 5:** In most social situations, I like to have someone else take the lead.
- **Statement 6:** In social situations, I usually allow others to dominate the conversation.
- **Statement 7:** When it is time to make decisions, others usually turn to me.

VI Demographics

Question 1: Which village do you reside in?

Question 2: Where were you born?
Reply options:

- *In my current place of residence.*
- *In [federal state], but not in my current place of residence.*
- *In Germany but outside of [federal state].*
- *Outside of Germany.*

Question 3: Is German your mother tongue?

Reply options: *yes, no*

Question 4: What is your family status?

Reply options: *married, partnership, widowed, divorced, single*

Question 5: Do you have children?

Reply options: *yes, no*

Question 6: What is your highest school or college degree?

Reply options:

- *Without a general school leaving certificate.*
- *Without a general school leaving certificate.*
- *Secondary school diploma (Realschulabschluss).*
- *High school diploma or equivalent (Abitur).*
- *Apprenticeship.*
- *Study without a degree.*
- *Preliminary diploma (Vordiplom).*
- *Diploma (Diplom).*
- *Ph.D.*

Question 7: What is your husband's or partner's highest educational qualification?

Reply options: see the previous question

Question 8: Are you currently employed?

Reply options: *full time, part-time, no*

Question 9: Are you involved in your community? For example in a club or in church?

Reply options: *yes, no*

Question 10: Please enter your date of birth (DD/MM/YYYY).

Chapter 3

The Effect of Minimum Volume Requirements on Strategic Behavior of Hospitals

Abstract

Minimum volume requirements (MVR) are increasingly used to centralize surgical procedures, aiming to improve patient outcomes. This study examines the effects of MVR on German hospitals' behavior and patient outcomes. We find that MVR increase the number of knee replacements in hospitals that are at risk of missing the minimum threshold, but not the number of pancreatic and esophageal surgeries. We find evidence of misreporting in quality reports: hospitals below the minimum threshold are more likely to over-report their procedure volumes. We find no effect on patient characteristics or length of stay for marginal patients. Overall, the observed adverse effects of MVR are small, indicating their suitability as a policy tool for centralizing surgeries.[1]

[1]This chapter is co-authored with Corinna Hartung. We are grateful to Mathias Bühler, Sarah Eichmeyer, Kevin Kloiber, Sarah Miller, Andrew Proctor, Simon Reif, Lise Rochaix, Andreas Steinmayr, Gregory Veramendi, Joachim Winter, and Amelie Wuppermann and to seminar participants at LMU Munich and the Annual Conference of the German Health Economics Association for valuable comments. We thank Jurek Siebert for excellent research assistance. We used the generative AI tools Claude and ChatGPT for coding and in order to edit the text for clarity, form, and grammar. We reviewed and edited this content as needed and take full responsibility for the content of the publication. Financial support from the DFG (German Research Foundation) via GRK 1928 and the Evidence-Based Economics Programme within the Elite Network of Bavaria is gratefully acknowledged.

3.1 Introduction

The volume-outcome relationship in healthcare is well-established: hospitals performing higher volumes of specific procedures consistently achieve better patient outcomes (Pieper et al. 2013). This phenomenon, observed across a wide range of surgical interventions, has profound implications for healthcare policy (Levaillant et al. 2021). In response, many OECD countries have implemented policies to centralize these procedures (Vonlanthen et al. 2021). Minimum Volume Requirements (MVR) have emerged as a key policy tool in this effort. These policies set thresholds for the number of procedures a hospital must perform annually. While they aim to improve patient outcomes by ensuring hospitals have sufficient experience with specific procedures, MVR may also influence treatment choices and patient outcomes through altered incentives, as demonstrated in other contexts (Batty and Ippolito 2017). Consequently, the implementation of MVR raises important questions about potential unintended consequences, particularly regarding strategic hospital behavior and its impact on patient care and outcomes.

This study examines the effects of minimum volume requirements (MVR) on German hospitals operating near the threshold—those performing surgeries at volumes close to the minimum requirement. We focus on the causal effect of MVR on procedure volumes, accuracy in quality reporting, overtreatment (i.e., how marginal additional patients differ from average patients), and health outcomes.

Hospitals must meet these minimum procedure thresholds within a calendar year, with potential loss of reimbursement for non-compliance. However, lax enforcement creates a complex incentive structure. We analyze comprehensive billing data and publicly available quality reports from 2008 to 2017[2]. We

[2]Quality reports are mandatory, publicly available documents that hospitals must submit annually, detailing their performance on various metrics, including procedure volumes. They are commonly used by policy makers and patients to evaluate hospital quality.

focus on total knee replacements, complex pancreatic surgeries, and complex esophageal surgeries because these are the procedures covered by MVR that are frequent, and clinically and economically important.

Our study employs multiple identification strategies to assess the impact of MVR on strategic hospital behavior and patient outcomes. First, we examine the effect of MVR on procedure volumes. We leverage two policy settings: the requirement to fulfill MVR within a calendar year, which might lead to pressure to reach the threshold towards the end of the year, and a four-year suspension of the policy for knee replacements, allowing us to compare years when the requirements were in effect versus suspended. To quantify the magnitude of hospitals' responses, we adapt methods from the literature on behavioral responses to policy thresholds. Specifically, we estimate hypothetical counterfactual procedure distributions in the absence of MVR.

Second, we employ a regression discontinuity approach to examine potential misreporting in quality reports near the minimum thresholds. This analysis examines the accuracy of quality reports, which are publicly available documents that may influence patient decision-making and public opinion, by comparing reported procedure counts with billing data.

Third, we utilize an instrumental variables approach to assess the impact of MVR-induced increases in procedure counts on patient characteristics and health outcomes. We construct a measure of *pressure* based on the deviation of observed procedure counts from predicted counts in a scenario without MVR. This pressure serves as an instrument for weekly procedure counts in a two-stage least squares estimation at the patient level, allowing us to examine the effects on marginal patient characteristics and health outcomes. This approach provides detailed insights into the specific changes that occur when hospitals face pressure from MVR.

Leveraging these identification strategies, we find evidence that minimum volume requirements moderately increase the number of surgical procedures

in hospitals at risk of missing the threshold, but only for knee replacements. This effect is not observed for complex pancreatic and esophageal surgeries, possibly due to the more complex interdisciplinary decision-making processes involved and the life-threatening nature of these procedures. We estimate the policy-induced increase in knee surgeries to be around 140 procedures nationally per year (+0.1%), a small effect at a societal level. We further rule out the simple explanation that surgeries are shifted from subsequent years.

Using a regression discontinuity approach, we find evidence of misreporting in quality reports for all three procedures. The proportion of hospitals reporting more procedures than they have billed increases markedly just below the minimum threshold. This pattern raises concerns about the reliability of quality reports and their potential to bias patients' future treatment decisions.

Our instrumental variables analysis reveals changes in patient composition due to `MVR` pressure related to the geographical distribution of patients. We observe a statistically significant decrease in the proportion of patients residing in the same district as the hospital, indicating that hospitals may be expanding their catchment areas to meet volume requirements. However, we do not find evidence for changes in other patient characteristics such as age. Furthermore, we detect no effects on health outcomes as measured by surgery complications or length of hospital stay. These findings suggest that while hospitals may be drawing patients from a wider geographical area, they are not substantially altering their patient selection criteria to meet `MVR` targets.

Overall, our analysis reveals some effects of the `MVR` policy, particularly in terms of strategic behavior by hospitals near the threshold for knee replacements and misreporting in quality reports for all procedures. However, these effects are relatively small in magnitude, and the policy's primary goal of centralizing complex surgeries is largely achieved. While minimum volume requirements have unintended consequences, they may still be a viable policy tool for promoting surgical centralization.

Our study provides novel evidence on hospitals' strategic responses to minimum volume thresholds, contributing to the literature on healthcare providers' behavior under policy changes and financial incentives. Previous research has shown that healthcare providers often respond creatively to new regulations and financial incentives, sometimes in unexpected ways. Alexander (2020) found that doctors responded to cost-reduction incentives that were in effect at some hospitals by strategically admitting healthier patients to participating hospitals, thus collecting more bonus payments. Eliason et al. (2018) revealed how long-term care hospitals strategically timed patient discharges to maximize Medicare payments. Dafny (2005) demonstrated how hospitals responded to Medicare price changes by upcoding patient diagnoses to increase reimbursements. Batty and Ippolito (2017) showed that hospitals could target care based on financial considerations, adjusting the amount of care delivered to uninsured patients when state laws reduced payments, without affecting quality metrics. These findings collectively underscore the ability of healthcare providers to adapt to and potentially exploit incentive schemes.

We also add to the literature on Minimum Volume Requirements (`MVR`) in healthcare. Several studies have examined their implementation and effects. Hentschker and Mennicken (2018) investigated the volume-outcome relationship for pancreatic surgery in Germany, finding that higher hospital volumes were associated with lower mortality rates. De Cruppé et al. (2020) analyzed the impact of suspending `MVR` for knee replacements in Germany, showing that the policy change led to an increase in the number of low-volume providers. Červený (2023) found that the positive relationship between volume and outcomes disappeared when accounting for selective referral and differences in provider quality, thus questioning the effectiveness of `MVR`[3].

Our study builds on this work by providing causal evidence on how hospitals

[3]A forthcoming systematic review by Scharfe et al. (2023) aims to comprehensively assess effects of `MVR` in various contexts.

strategically respond to MVR across multiple procedures, offering insights into the effectiveness and potential unintended consequences of such policies. Furthermore, we take a novel within-year perspective, analyzing how hospitals dynamically react to the pressure of meeting the MVR threshold within a calendar year.

The remainder of this paper is structured as follows. In Section 3.2, we provide background on the medical details, the options for policy makers, and the German setting. We describe the data sources in Section 3.3 and present descriptive statistics in Section 3.4. Sections 3.5, 3.6, and 3.7 present our main results on the effects of MVR on procedure volumes, misreporting in quality reports, and overtreatment and health, respectively. Finally, we discuss the implications of our findings in Section 3.8.

3.2 Background

The centralization of surgeries, specifically the concentration of procedures at high-volume hospitals, is a common policy goal in healthcare systems worldwide. In this section, we outline the rationale behind the push for surgical centralization, explore the options available to policymakers in pursuing this goal, and describe the minimum volume requirements (MVR) policy that is the focus of this paper. Additionally, context on the specific characteristics of the German setting is provided.

3.2.1 The Volume-Outcome Relationship

The academic literature consistently demonstrates a positive association between surgical volumes and patient outcomes. Luft et al. (1979) documented this volume-outcome relationship that has been replicated across various settings and procedures (e.g. Birkmeyer and Lucas (2002), Krautz et al. (2018), Ratnayake et al. (2022).

Two competing hypotheses explain the causal pathway underlying this relationship: the *Practice Makes Perfect* hypothesis (PMPH) suggests that increased provider experience leads to enhanced performance (Hentschker and Mennicken 2018; Hamilton and Ho 1998), while the *Selective Referral* hypothesis posits that higher-quality providers attract more patients, resulting in higher volumes (Dudley et al. 2000; Červený 2023).

Despite ongoing debates about causality, policymakers have implemented measures to improve patient outcomes based on the robust correlation between volume and quality. These interventions, such as the introduction of minimum volume requirements (MVR), aim to concentrate surgical procedures in high-volume centers. However, such policies may have unintended consequences on hospital behavior and patient outcomes due to financial incentives not previously considered in the literature.

While the volume-outcome relationship provides important context, our paper does not aim to evaluate this relationship or its underlying mechanisms. Instead, we focus on assessing one specific policy response: the implementation of minimum volume requirements. This approach allows us to examine the practical implications of centralization efforts in healthcare.

3.2.2 Centralization: Policy Options and International Comparison

OECD countries have implemented various policies to incentivize surgical centralization, aiming to improve patient outcomes (Vonlanthen et al. 2018). Three approaches are prominent: certification, central planning, and minimum volume requirements (MVR).

Certification involves recognizing providers based on excellence criteria which could include outcome rates, staffing ratios, or case volume. While not directly linked to reimbursement schemes, this approach aims to

influence patient and insurer decisions through reputational effects and recommendations. In some instances, these certificates might be a requirement for hospitals to perform the surgery.

Central planning represents a more direct approach, where regulators designate specific providers to perform sets of surgeries. These can be tied to predefined geographic areas. This method limits patient choice but allows for a more targeted allocation of resources and expertise.

Minimum volume requirements (MVR) have become a frequently implemented policy across many countries. This approach sets thresholds for the number of procedures a hospital or surgeon must perform to be eligible for reimbursement or to continue offering the service. They vary across countries in four dimensions: covered medical procedures, required minimum volume, relevant time frame (usually annual), and the entity to which the minimum applies (e.g., hospitals, physicians). The implementation and effectiveness of MVR policies continue to be subjects of ongoing research and debate in the healthcare policy arena.

Vonlanthen et al. (2018) provided an overview of country-specific centralization strategies regarding gastrointestinal surgery for twenty European countries, Canada, and the United States. In Europe, 65% of countries had implemented minimum volume requirements that apply to entire hospitals, while one country required minimum caseloads per surgeon. Other countries either lacked centralization policies entirely or had varying approaches implemented at sub-national levels.

Morche et al. (2018) compared minimum volume standards in Europe and found that the thresholds and the defined procedures vary strongly across countries. Another aspect that differs between countries is the level of enforcement and consequences for non-compliance.

In the United States, centralization is driven by various large healthcare organizations such as the Veterans Health Administration and large insurance

companies. The Leapfrog Group, a voluntary program representing many of these organizations, recommends minimum volume thresholds for 10 surgical procedures and attempts to monitor the quality of surgical appropriateness. Additionally, the *Take the Volume Pledge* campaign, introduced in 2015, recommends minimum volume thresholds. However, these pledges and minimum volume standards have stirred controversy among U.S. surgeons (Urbach 2015). For implants, Medicare has specified minimum volume thresholds. In Canada, healthcare is mostly the responsibility of each of the ten provinces, and policies to centralize surgeries vary by province.

Finally, the effectiveness of these centralization policies varies between acute and elective surgeries. Acute conditions, requiring immediate intervention typically at the nearest hospital, limit the applicability of centralization, especially in rural areas. Conversely, elective procedures allow greater flexibility in provider selection due to the extended timeframe between diagnosis and intervention. This flexibility enables more effective implementation of centralization policies for elective procedures, potentially optimizing resource allocation and patient outcomes by directing patients to high-volume hospitals.

3.2.3 Minimum Volume Requirements in Germany

Minimum Volume Requirements (`MVR`) were first introduced in Germany in 2006 by the *Federal Joint Committee*, a self-governing body consisting of representatives from hospitals, insurers, and physicians. Overseen by the Ministry of Health, it regulates health care in the social health insurance and hospital context.

The Federal Joint Committee has established `MVR` for several procedures, initially introduced alongside the reform towards diagnosis-related groups (DRG) in the early 2000s. These minimum volume requirements include liver transplantation (including partial liver living donation) with a minimum of 20 cases, kidney transplantation (including living donation) with 25 cases,

complex interventions on the esophagus and pancreas with 10 cases each, stem cell transplantation with 25 cases, and total knee replacement with 50 cases. Additionally, care for premature and newborn infants with a birth weight of less than 1250 grams requires a minimum of 25 cases.

The implementation of these requirements has not been without controversy. Notably, the minimum volume requirement for knee replacements was temporarily suspended from 2011 to 2014 due to a legal dispute between hospitals that were refused payment and the Federal Joint Committee (Ettelt 2017). This suspension made headlines and was well-known to hospitals and the public. More recently, the change in the minimum volume for care of premature infants (previously set at 14 cases) has led to the closure of some hospital birth units, garnering regional media attention.

An important aspect of `MVR` in Germany is their weak enforcement throughout our study period. By law, `MVR` affect reimbursement, with hospitals not meeting the requirements ineligible for compensation. However, enforcement has been limited in practice, and hospitals often successfully justify exceptions (De Cruppé and Geraedts 2018). Typically, reimbursement negotiations occur privately between hospital administrators and insurance representatives. The legal dispute over knee replacements in the early 2010s, when multiple hospitals were denied reimbursement for failing to meet minimum thresholds, was a rare instance of strict enforcement becoming public (Ettelt 2017)[4].

[4]After our study period, the Federal Joint Committee implemented a regulations to enhance enforcement. Hospitals must now report their case numbers for treatments subject to `MVR` to state health insurers, who review and verify these numbers. Recently, some thresholds have increased, among them pancreatic (from 10 to 15 in 2024 and to 20 in 2025) and esophageal surgeries (from 10 to 26 in 2023). New minimum volume requirements for lung cancer surgeries (40 in 2024, increasing to 75 in 2025) and breast cancer surgeries (50 in 2024, increasing to 100 in 2025) have also been introduced.

3.2.4 Relevant Surgical Procedures

This study focuses on three surgical procedures subject to `MVR`: total knee replacements, complex pancreatic surgery, and complex esophageal surgery. These elective procedures are performed at a large number of hospitals across Germany (Knee: >1000 hospitals, Pancreatic surgery: >600, Esophageal surgery: >400). Unlike rarer surgeries such as transplants, which are typically confined to large city hospitals, these operations are also commonly conducted in rural areas. This geographic distribution highlights the trade-off between hospital proximity and size. The volume-outcome relationship for these procedures is well-documented in the literature (Kugler et al. 2022; Krautz et al. 2018; Nimptsch et al. 2017; Nimptsch et al. 2018).

We do not analyze other procedures which are also covered by `MVR`. Besides kidney and liver transplants, we also exclude care for neonates with a birth weight below 1250 grams due to differences in the process of provider choice. Specifically, in the area of obstetrics, some hospitals are classified as level 1 perinatal centers and only these are allowed to care for very premature births.

Total knee replacements are complex surgeries during which a deteriorated or damaged knee joint is replaced by a prosthesis. In Germany, around 190,000 total knee replacement surgeries are performed each year. The procedure is typically performed when conservative therapy has failed. Therefore, the typical counterfactual is *additional conservative therapy*. Main risks include infections and post-surgery pain (Lützner et al. 2018). Arias-de la Torre et al. (2019) documented that patients at hospitals with higher volume show better outcomes. In 2006, the minimum threshold of 50 knee replacements was introduced in Germany. It was temporarily suspended from 2011 to 2014 due to a legal dispute between a hospital and the *Federal Joint Committee*. Financially, a hospital is reimbursed around €8,000 for a typical hospital stay with an uncomplicated knee replacement surgery.

Complex pancreatic and complex esophageal surgeries have both been subject to MVR regulations since 2006. The minimum volume for each is 10 procedures per calendar year. Complex pancreatic surgeries include the removal of the pancreas and the partial removal of the pancreas in patients with pancreatic cancer. They are performed around 14,000 times per year in Germany. Complex esophageal surgeries include the (partial) removal of the esophagus and reconstruction of the esophagus, mostly in patients with esophageal cancer. They are performed around 4,000 times per year in Germany. For both complex pancreatic and complex esophageal surgeries, the typical alternative — and thus the counterfactual — is *palliative treatment*, as the cancer is considered inoperable.. Main risks include bleeding and death (McGuigan et al. 2018; Yang et al. 2020). Lidsky et al. (2017) showed that patients who travel to higher volume hospitals for pancreatic surgery have better outcomes. Schlottmann et al. (2018) show that esophageal surgery outcomes are improved after centralization. For the German context, Nimptsch and Mansky (2017) show that outcomes in complex esophageal surgery and complex pancreatic surgery are better at higher-volume hospitals. In our dataset, hospital stays involving complex esophageal surgeries result in an average reimbursement of €31,000, while stays involving complex pancreatic surgeries average €22,000. However, these amounts should not be interpreted as the direct reimbursement for the surgical procedure alone. Instead, they reflect the DRG reimbursement for the entire hospital stay, which may encompass various treatments, procedures, and the overall duration of hospitalization.

Knee replacements and complex cancer surgeries (pancreas and esophagus) differ significantly in terms of typical patients and decision-making processes. For knee replacements, the process typically begins with a referring orthopedist who recommends the surgery based on the patient's condition and failed conservative treatments (Wengler et al. 2014). The decision-making process

in the hospital is often simple, with a single specialist or a small team making the recommendation.

For pancreatic and esophageal cancer surgeries, the decision-making process is more complex. Interdisciplinary tumor boards, comprising specialists such as surgeons, oncologists, radiologists, and pathologists, collectively determine patient treatment plans (El Saghir et al. 2014; Pillay et al. 2016). This multidisciplinary approach may alter the impact of `MVR` on treatment decisions compared to less complex procedures. The involvement of multiple specialists, potentially unaware of the hospital's current surgical volume, could influence `MVR`'s effect on decision-making. Our analysis will empirically examine the actual impact of `MVR` on treatment choices in these complex cases.

3.3 Data

Our main data source is the universe of inpatient claims in German hospitals from 2008 to 2017[5]. For these *billing data*, the unit of observation is one patient episode at a hospital; it is not possible to identify re-admissions. The data include detailed billing information on diagnoses (ICD-10) and procedure codes. We classify procedures as relevant for the `MVR` based on the yearly guidelines by the Federal Joint Committee, which list all specific procedure codes for a given category.

The data also include information on the timing of the procedures and the entire hospital stay as well as claim size in EUR and basic demographic patient characteristics. A hospital identifier is also included, allowing us to create a hospital-level panel dataset that includes `MVR`-related surgery volume at a weekly and monthly level. We thus have a count per hospital of the *stock* and *flow* of surgeries per month and week for all procedure-years. The stock here refers to the number of performed surgeries at this hospital in the given

[5]Specifically, we use *DRG-Statistic* data by the RDC of the Federal Statistical Office and Statistical Offices of the Federal States (2017)

calendar year before a specific week/month started.

Our second data source is the universe of structured quality reports from 2008 to 2017.These *quality report data* are unavailable for 2009 and 2011 as they were only required biennially until 2012. The quality reports are mandatory for all hospitals and are available online as PDF files for each hospital-year[6]. Among a host of information about staffing, organizational structure, and patient counts, the reports include the total number of performed surgeries for all categories covered by the minimum volume requirements for a given year(Auras et al. 2012). We link this data at the hospital level to the billing data. Our analysis of potential misreporting in Section 3.6 relies on the comparison between these two data sources. The discrepancies in procedure counts between quality reports and billing data, particularly around the `MVR` thresholds, provide insights into potential strategic reporting behavior by hospitals.

Our data sources have limitations that are important to acknowledge. Key discrepancies arise from differing reporting timelines: billing data are generated shortly after patient discharge, while quality reports are compiled annually, often months later. This temporal gap can create inconsistencies. Additionally, billing data are managed by medical coding specialists focused on accurate coding for reimbursement, whereas quality reports are overseen by managers evaluating overall hospital performance. These distinct roles and objectives may lead to variations in data entry and interpretation. Inconsistent timing rules, such as recording the date of procedure versus admission/discharge dates, can also contribute to discrepancies between data sources. It is further important to note that routine billing data may not always provide consistent and accurate coding of diagnoses and procedures.

[6]Publicly available via https://qb-referenzdatenbank.g-ba.de/

Table 3.1: Patient Statistics by Surgery Type (All hospitals)

	Esophageal surgery		Pancreatic surgery		Knee replacement	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (years)	63.36	12.35	64.40	12.77	69.05	9.57
Length of stay (days)	30.42	26.79	28.74	22.83	12.24	6.12
Reimbursement (k EUR)	30.65	35.56	22.38	27.45	7.67	2.73
No. of diagnoses	14.36	10.03	14.09	9.34	5.81	3.84
No. of procedures	20.71	16.79	17.65	14.49	3.43	2.71
No. of cancer diagnoses	1.47	0.96	1.20	1.09	0.01	0.11
Shares:						
Female	0.24		0.44		0.64	
Lives close to hospital	0.40		0.48		0.53	
Surgery Complications	–		–		0.02	
Death	0.10		0.10		0.00	
N	40,214		109,803		1,505,785	

Notes: Data source: German hospital billing data, 2008–2017. "Lives close to hospital" indicates patient resides in the same district as the hospital. "Surgery Complications" only refer to knee replacement complications (ICD code T.85). Reimbursement is measured in thousands of euros.

3.4 Descriptive Statistics

Table 3.1 presents descriptive statistics for all patients who underwent a procedure subject to minimum volume requirements in one of the categories analyzed in this paper between 2008 and 2017.

On the right, we observe approximately 1.5 million patients who underwent total knee endoprosthetic surgery. These patients have a mean age of 69 years, with 53% residing in the same district as the hospital. In the middle and left columns, we present data on approximately 40,000 patients who underwent complex esophageal surgery and 110,000 who received complex pancreatic surgery during the study period. These two groups of patients have a mean age of 64 years, exhibit more comorbidities as evidenced by their higher diagnosis count, and generate greater reimbursement for hospitals. The in-hospital mortality rate for these patients is approximately 10%.

Table 3.2: Patient Statistics by Surgery Type (Hospitals near MVR threshold)

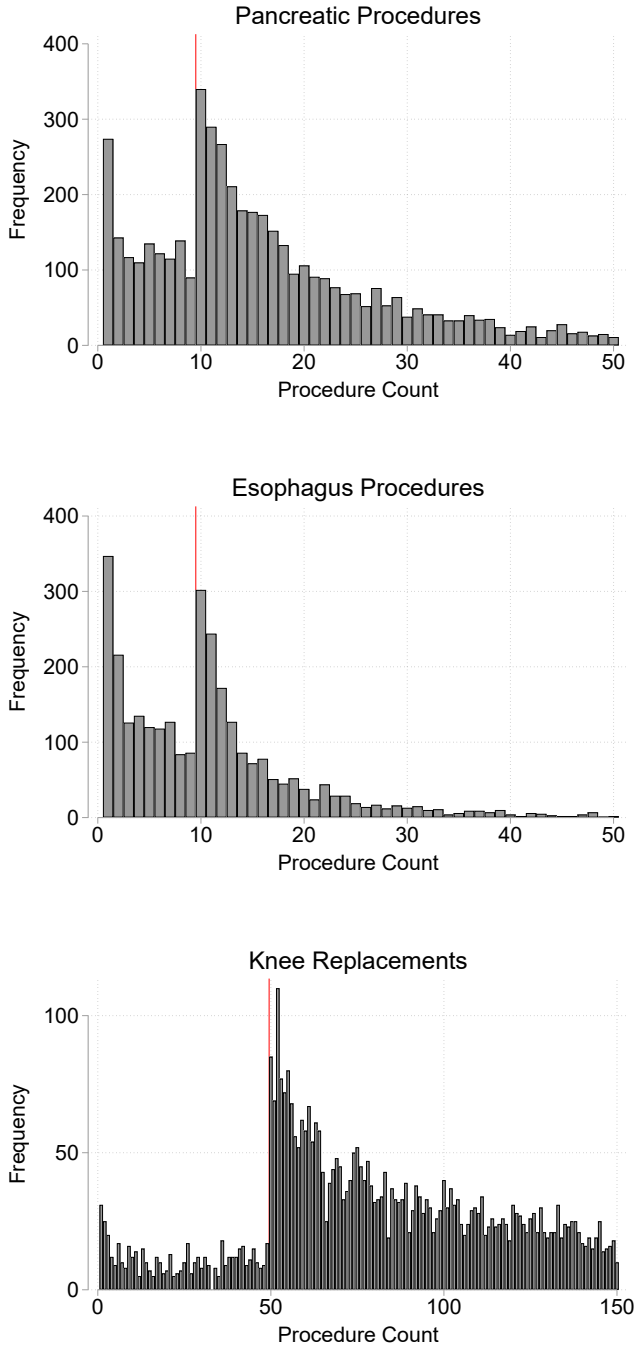
	Esophageal surgery		Pancreatic surgery		Knee replacement	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (years)	64.80	11.54	65.71	12.13	69.60	9.72
Length of stay (days)	30.91	26.37	30.23	21.71	13.35	7.26
Reimbursement (k EUR)	29.95	34.60	22.40	26.03	7.94	3.29
No. of diagnoses	14.17	9.51	14.10	8.85	6.00	4.10
No. of procedures	19.72	15.22	17.37	12.75	3.43	3.14
No. of cancer diagnoses	1.50	0.95	1.17	1.06	0.01	0.12
Shares:						
Female	0.25		0.44		0.65	
Lives close to hospital	0.54		0.71		0.66	
Surgery Complications	–		–		0.02	
Death	0.11		0.13		0.00	
N	13,999		21,650		147,110	

Notes: Data source: German hospital billing data, 2008–2017. This table includes only hospital-years where the total billed procedures are within 50% of the minimum volume requirement (MVR). For example, knee replacements with MVR = 50 include hospitals performing 26 to 75 procedures annually. "Lives close to hospital" indicates patient resides in the same district as the hospital. "Surgery Complications" are specific to knee replacement procedures (ICD code T.85). Reimbursement is in thousands of euros.

Table 3.2 replicates the statistics from Table 3.1, but only includes patients treated at hospitals potentially affected by Minimum Volume Requirements (MVR). We define these as hospital-years in which the respective annual procedure count falls within 50% of the MVR threshold. This subset represents hospitals for which the minimum thresholds are most relevant, typically smaller- and medium-sized institutions. Our analysis in subsequent Sections 3.5, 3.6, and 3.7 concentrates on these hospitals and the impact of MVR on their operations and patient outcomes.

The patient demographics at these MVR-sensitive hospitals broadly mirror the overall patient population, with some notable distinctions. Patients are marginally older across all procedures and experience slightly longer hospital stays. Notably, a substantially larger proportion of patients reside in the same

Figure 3.1: Distribution of Reported Procedure Totals



Notes: These histograms depict the frequency distribution of reported procedures in quality reports across all available data years for pancreas, esophagus, and knee replacement surgeries. Red vertical lines indicate the minimum volume thresholds (10 for pancreas and esophagus, 50 for knee replacements). The upper tails of the distributions are truncated for visual clarity. Distributions of other procedures covered by `MVR` can be found in the Appendix.

district as the hospital. This pattern aligns with the tendency for patients to seek care locally or, if traveling for treatment, to choose larger urban hospitals rather than similarly-sized rural institutions farther from home. For pancreatic and esophageal surgeries, we observe marginally higher mortality rates in this subset of hospitals.

Next, we examine data from German hospitals' quality reports. Figure 3.1 illustrates the distribution of reported annual hospital volumes for MVR-related procedures, encompassing all years in our study period when quality reporting was mandatory (excluding 2011-14 for knee replacements). The histograms, which display a subset of the full range, reveal distinct bunching just above the minimum volume threshold for each procedure. Notably, hospitals report volumes just meeting the threshold more than three times as frequently as volumes just below it. We investigate the extent of potential misreporting at these thresholds in Section 3.6.

3.5 Effect on Number of Procedures

3.5.1 Identification Strategy

To assess the impact of Minimum Volume Requirements (MVR) on hospital procedure volumes, we employ a quasi-experimental approach. Our identification strategy leverages two key aspects of the reform. First, we exploit the temporal structure of MVR. The minimum thresholds apply to calendar years, with failure to meet them theoretically resulting in no reimbursement (though this is rare, as noted in Section 3.2.3). This may incentivize hospitals to increase procedure volumes towards year-end to meet the required minimum. Second, we take advantage of the mid-sample four-year suspension of MVR for knee replacements. We compare within-year trends in knee replacement procedures during periods when MVR were in effect versus suspended. This

approach allows us to assess hospitals' responses to `MVR` incentives while controlling for other factors that might influence procedure volumes.

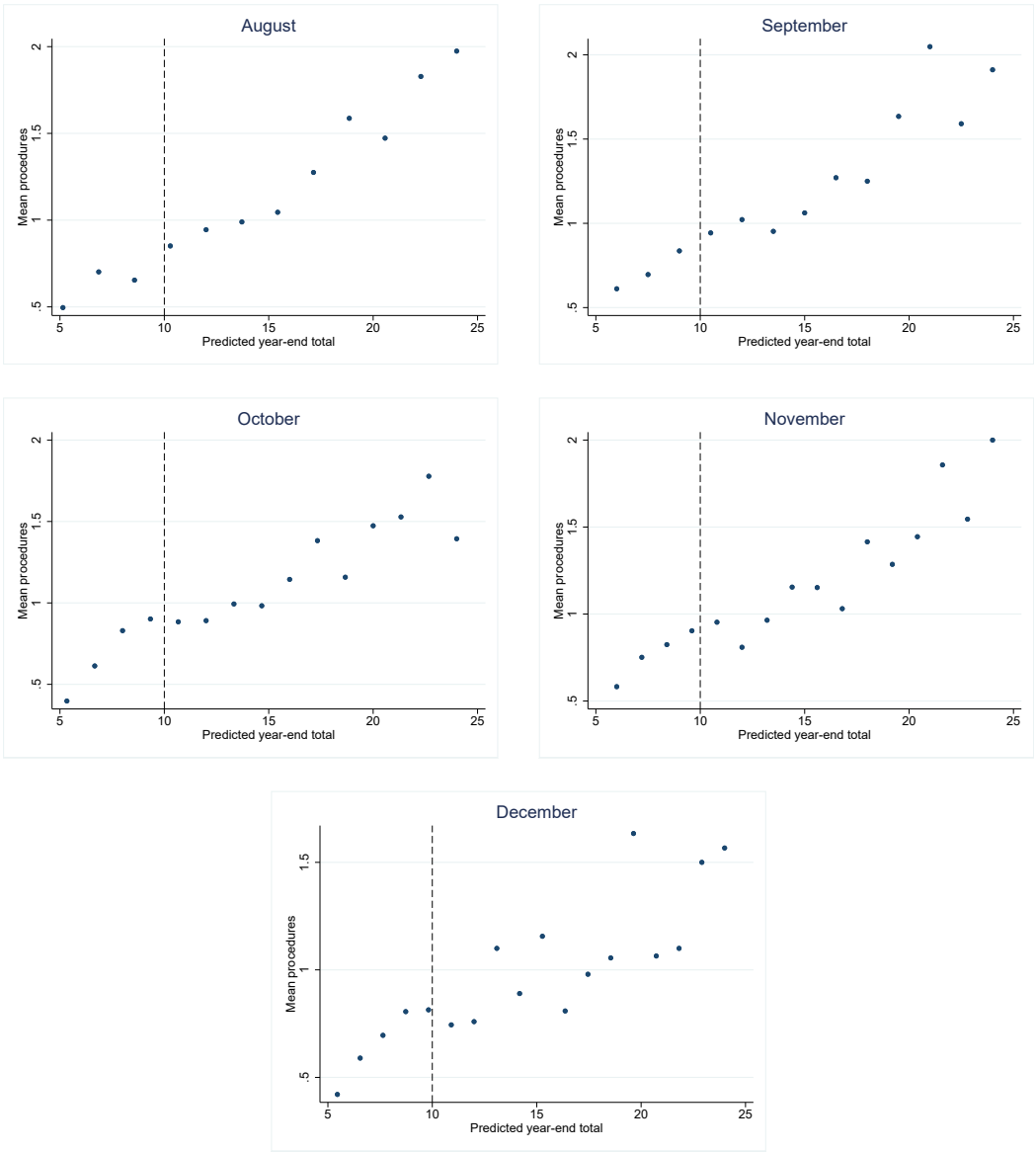
To visualize trends in our data, we extensively employ binned scatter plots (also known as binscatters), a graphical tool that combines data binning with scatter plots (Cattaneo et al. 2024). These plots divide the x-axis into intervals, or bins, and display the average outcome for all observations within each bin, effectively summarizing complex relationships in large datasets. In our analysis, the x-axis of these binscatters represents the cumulative number of procedures performed by a hospital within a calendar year, with each integer value forming a discrete bin. For example, all hospitals that have conducted five pancreatic surgeries up to a given point in the year are grouped in the same bin. The y-axis displays the average number of surgeries performed in the current month by all hospitals within that bin. Notably, these binscatters aggregate data across multiple years, allowing a single hospital to appear in different bins across years based on its procedure volume at any given point. This approach enables us to visualize trends in surgical volumes as hospitals approach the minimum volume threshold.

3.5.2 Within-Year perspective: Do hospitals react and when?

Figures 3.2 through 3.4 display binned scatterplots of monthly procedure volumes performed by hospitals, categorized by their previous cumulative procedure count. To enhance comparability across time, we transform the cumulative count z on the x-axis into a projected year-end total. This transformation assumes hospitals will maintain their current surgery rate throughout the year. We apply a linear transformation by multiplying z by $\frac{12}{m-1}$, where m represents the current month. For instance, in August ($m = 8$), a hospital with 7 procedures after 7 months would have a predicted year-end total of $7 \times \frac{12}{8-1} = 12$.

The y-axis depicts the average number of surgeries performed in the current

Figure 3.2: Monthly Esophagus Procedures in Relation to Cumulative Surgeries Performed



Notes: These graphs show monthly binned scatterplots. Hospitals are grouped along the x-axis based on their cumulative procedure count of compleax esophageal surgeries for the year up to the previous month (e.g., for August, procedures from January to July). This count is linearly transformed into a projected year-end total to make graphs comparable. The y-axis displays the mean number of procedures performed per bin for the given month. A dashed line indicates the minimum volume threshold. The data spans from 2008 to 2017.

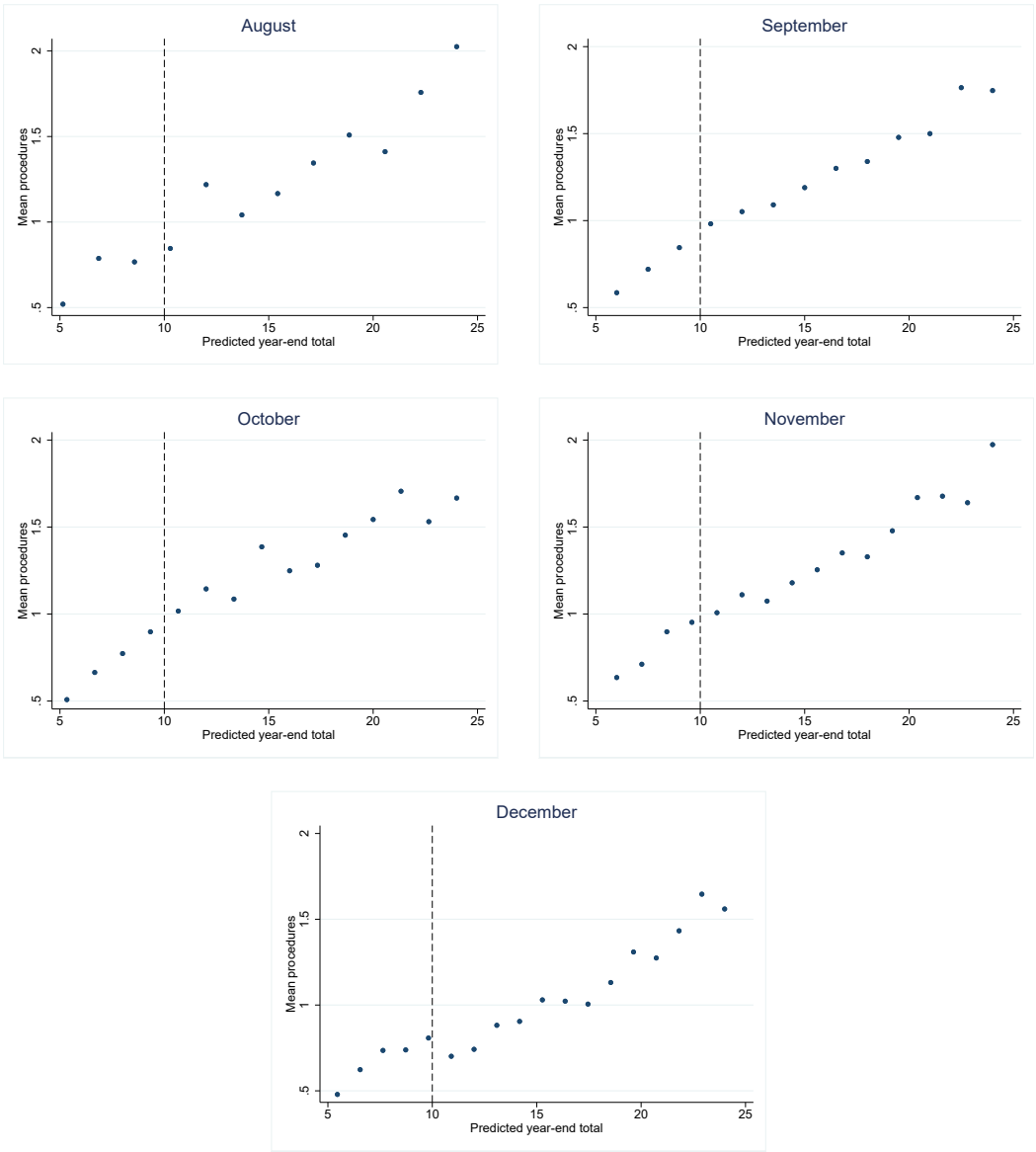
month by all hospitals within a bin. As an example, in the top-left panel of Figure 3.2, the point at x = 5.14 indicates that hospitals performing 3 surgeries by the end of July (projected to 5.14 annually) average 0.52 surgeries in August. A vertical dashed line marks the relevant minimum threshold. We adopt a monthly perspective to balance trend visibility and statistical power.

In the absence of an effect of MVR on the number of procedures, we would anticipate a positive and continuous relationship. Hospitals with higher previous procedure counts would, on average, perform more surgeries in subsequent months, even when accounting for the stochastic nature of surgical timing. However, if MVR alter treatment decisions, this relationship may change. By disaggregating newly performed surgeries by month, we can pinpoint when such changes occur. This approach is robust to seasonal trends (e.g., fewer December surgeries) and any shocks that affect all hospitals similarly. Note that individual hospitals may appear in different bins across years, further validating our comparative approach.

Figures 3.2 and 3.3, depicting esophagus and pancreas procedures, exhibit a positive and continuous relationship that does not noticeably change around the MVR threshold. A slight uptick is observable in the December graph for both procedures, but its magnitude is small.

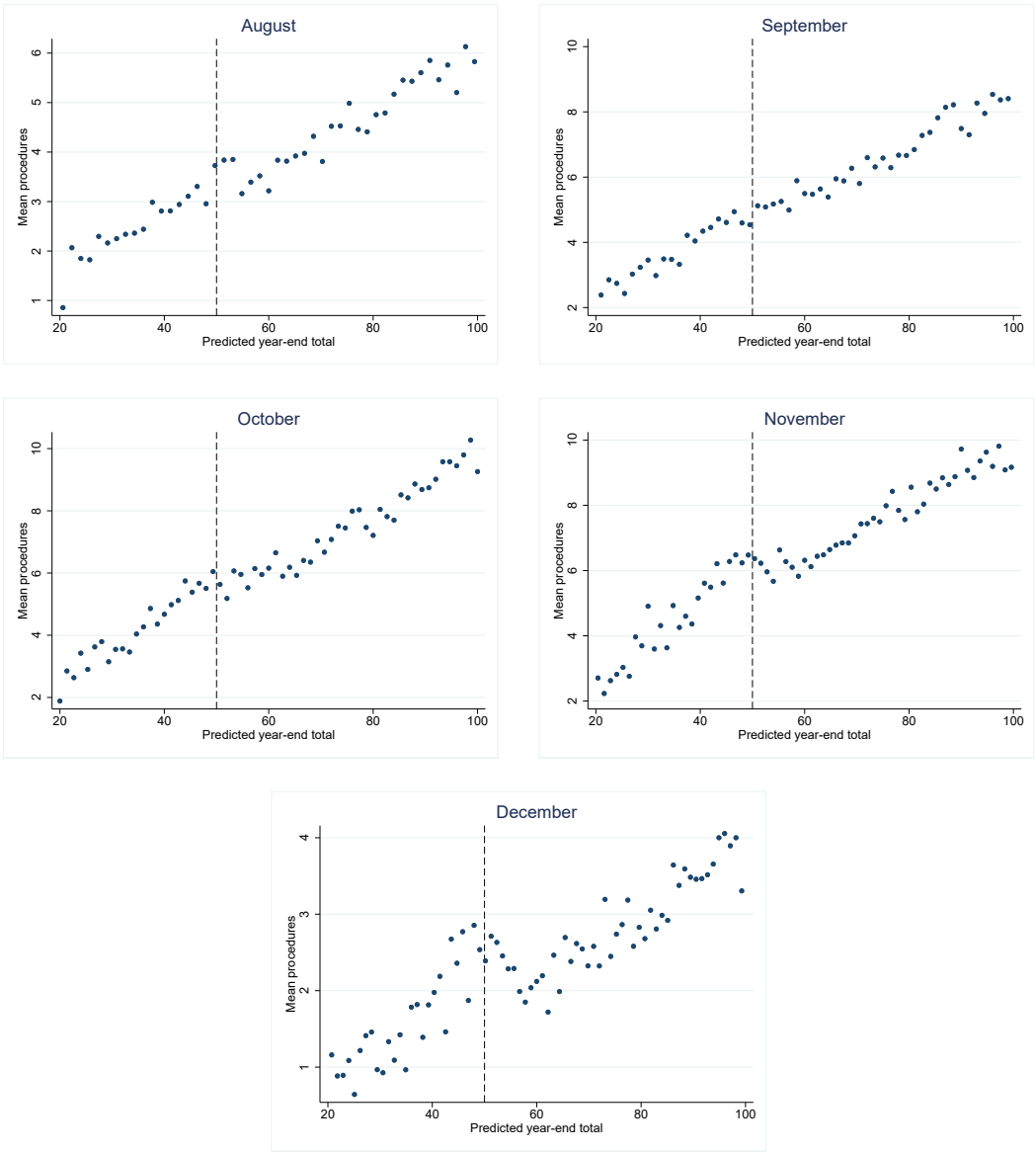
Figure 3.4, depicting knee replacement procedures, reveals a notably different pattern. Unlike the positive and continuous relationships observed for pancreas and esophagus procedures, distinct deviations emerge around the 50-procedure threshold for knee replacements in November and December. These bumps in the distribution suggest a potential response to the MVR policy. Figure 3.5 compares years when MVR were in effect to those when they were suspended. This comparison demonstrates a positive and continuous relationship during years of MVR suspension. In contrast, years with active MVR show a clear deviation from this pattern. During these periods, hospitals with a predicted year-end total just below 50 procedures perform approximately two

Figure 3.3: Monthly Pancreatic Procedures in Relation to Cumulative Surgeries Performed



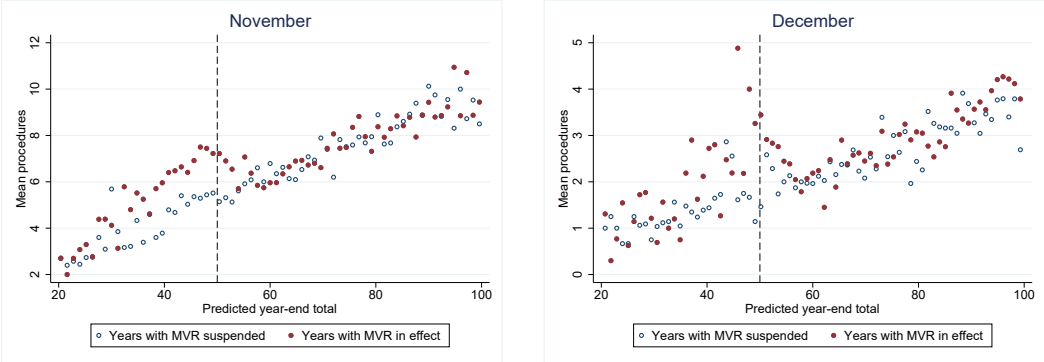
Notes: These graphs show monthly binned scatterplots. Hospitals are grouped along the x-axis based on their cumulative procedure count of complex pancreatic surgeries for the year up to the previous month (e.g., for August, procedures from January to July). This count is linearly transformed into a projected year-end total to make graphs comparable. The y-axis displays the mean number of procedures performed per bin for the given month. A dashed line indicates the minimum volume threshold. The data spans from 2008 to 2017.

Figure 3.4: Monthly Knee Replacement Procedures in Relation to Cumulative Surgeries Performed



Notes: These graphs show monthly binned scatterplots. Hospitals are grouped along the x-axis based on their cumulative procedure count of knee replacement surgeries for the year up to the previous month (e.g., for August, procedures from January to July). This count is linearly transformed into a projected year-end total to make graphs comparable. The y-axis displays the mean number of procedures performed per bin for the given month. A dashed line indicates the minimum volume threshold. The data spans from 2008 to 2017.

Figure 3.5: Late-Year Knee Replacement Procedures: Comparison of `MVR` and Non-`MVR` Periods



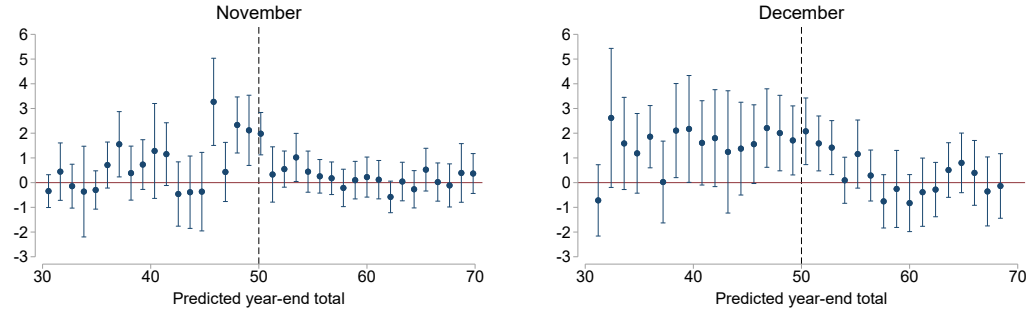
Notes: These graphs present monthly binned scatterplots for November and December. Hospitals are categorized into two groups: those operating during years when minimum volume requirements (`MVR`) were in effect (2008-2010, 2015-2017) and those operating during years when `MVR` were suspended (2011-2014). The x-axis represents the cumulative procedure count up to the previous month, linearly transformed into a projected year-end total. The y-axis shows the mean number of knee replacement surgeries performed per bin for the given month.

additional knee replacements per month in November and December. Figure 3.6 illustrates these differences across the number of previous procedures, showing that they are statistically significant at multiple points near the threshold[7].

One potential mechanism for the increase in knee replacement surgeries could be strategic scheduling. Hospitals might shift surgeries to earlier dates to meet the required minimum, thus performing fewer procedures in January of the following year. We would expect to see a corresponding dip in a *January(t+1)* graph based on procedure counts from *November(t)* that would mirror the bumps in Figure 3.5. Figure 3.7 presents these graphs for knee replacements, separated by years when `MVR` were in effect versus suspended. The absence of such opposite dips in both graphs suggests that year-to-year rescheduling can be ruled out as a main explanation for the previously detected increase in procedures.

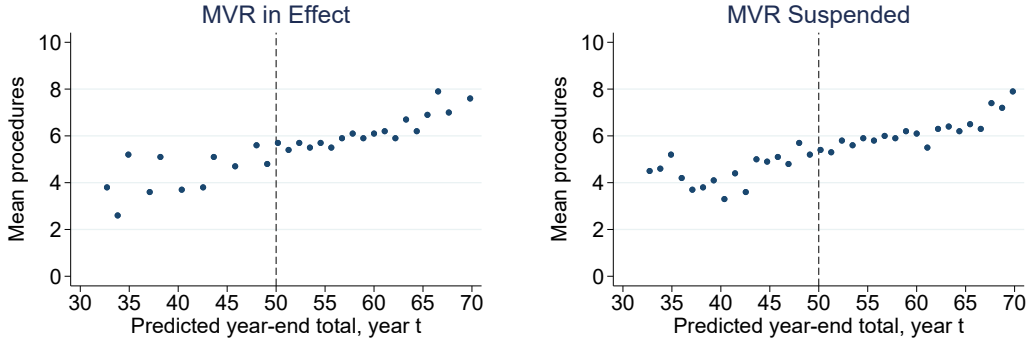
[7]In the December panel, the increase even above the minimum volume threshold is understandable when considering that hospitals typically perform fewer surgeries during the holiday season.

Figure 3.6: Difference in Late-Year Knee Replacement Procedures: November and December



Notes: This figure plots monthly knee replacement procedures differences, comparing periods when minimum volume requirements (`MVR`) were in effect versus suspended. Procedure totals are presented in Figure 3.5. The difference *procedures(MVR in effect) minus procedures(MVR suspended)* is displayed, along with 95% confidence intervals. Hospitals are binned along the x-axis based on their cumulative procedure count for the year up to the previous month.

Figure 3.7: Knee replacements in January of the subsequent year t+1



Notes: These graphs display binned scatterplots for January of year t+1. The x-axis bins hospitals by their procedure count through November of year t, mirroring the 'December' panels in Figure 3.4. The y-axis represents the mean number of knee replacement procedures performed per bin in January of year t+1. A dashed line indicates the minimum volume threshold. Data from 2008 to January 2018 are included.

Our analysis likely understates the true effect of MVR. Some hospitals may plan their procedure volumes well in advance, adjusting their entire yearly schedule to meet minimum thresholds. Such proactive planning would result in a more uniform distribution of procedures throughout the year, making hospitals' responses to MVR less detectable in our month-to-month analysis.

3.5.3 Magnitude of effect

We have demonstrated a response to Minimum Volume Requirements (MVR) for knee surgeries, but the magnitude remains unclear. To quantify the magnitude of hospitals' responses, we adapt methods from the literature on behavioral responses to policy thresholds (Chetty et al. 2011; Kleven 2016)[8]. Specifically, we estimate hypothetical counterfactual procedure distributions in the absence of MVR. This method allows us to compare the estimated counterfactual distribution to the observed data, allowing us to evaluate the extent of hospitals' reactions to the policy.

Our approach is formalized by estimating the following equation for each calendar month m with OLS:

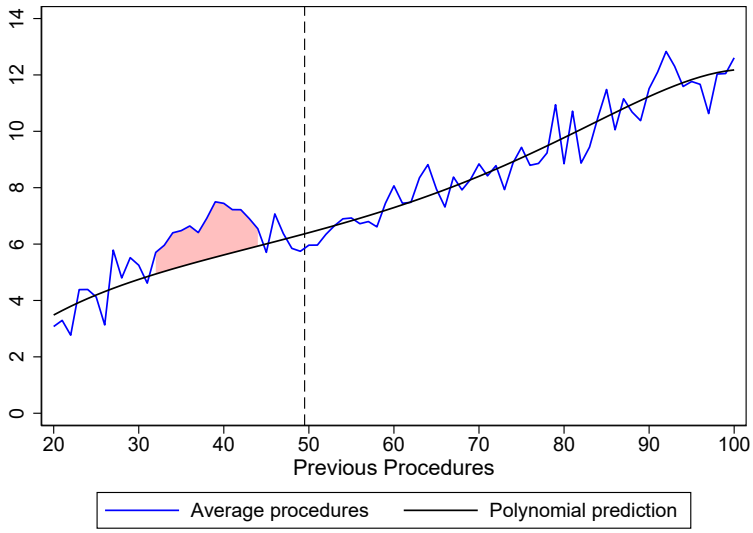
$$c_j = \sum_{i=0}^{p} \beta_i \cdot (z_j)^i + \sum_{i=z_-}^{z_+} \gamma_i \cdot \mathbf{1}\left[z_j = i\right] + v_j \tag{3.1}$$

In this equation, c_j represents the number of procedures in bin j, p is the order of the polynomial, z_j is the number of performed procedures until $m-1$ (where m is the month of a year), (z_-, z_+) is the excluded relevant range, and v_j is the error term.

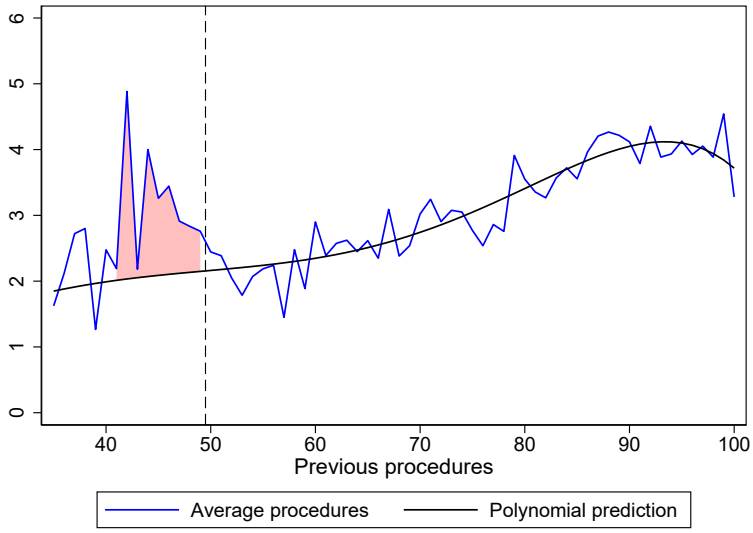
This equation aligns with our binscatter approach from the previous section: the dependent variable c represents the average number of procedures performed in month m for each bin, while z (the cumulative procedure count

[8] In its classical setting, one estimates counterfactual reactions to a tax schedule in the absence of kinks and notches in tax rates. True responses are then compared to estimate the responses to a kink or notch. In our setting, we estimate the counterfactual distribution, if MVR did not exist.

Figure 3.8: Observed Procedure Counts Vs. Estimated Counterfactual Distribution (Knee)



(a) November



(b) December

Notes: This figure compares the counterfactual distribution (black line) with observed data points (blue line) for knee surgeries in November and December. The black line is a counterfactual distribution derived from the methodology described in Section 3.5.3. The blue line represents binscatter results from Figure 3.5, corresponding to the red data points in that figure. The shaded red area indicates the increase in surgeries attributable to Minimum Volume Requirements (`MVR`). Data used in this analysis are from all years when `MVR` were in effect (2008-2010, 2015-2017).

up to month $m-1$) determines the bin assignment for each hospital-year.

To predict the counterfactual distribution, we use the left part of the equation:

$$\hat{c}_j = \sum_{i=0}^{p} \hat{\beta}_i \cdot (z_j)^i \tag{3.2}$$

Our data include a wide range of previous procedure counts $z \in [5, 300]$. Based on earlier results, we set the excluded ranges to $(32, 44)$ for November and $(40, 49)$ for December, which encompass the area around the threshold likely affected by the regulation. By fitting a polynomial to the observed distribution outside this excluded range, we can extrapolate what the distribution would look like in the absence of the `MVR`.

Figure 3.8 illustrates the results of our prediction exercise, comparing the estimated counterfactual distribution to the observed distribution of procedure counts. The shaded red area represents the increase in performed surgeries that can be reasonably attributed to `MVR` regulations.

To estimate the total impact of `MVR`, we calculate the difference between the counterfactual and observed procedures within the excluded range for each bin and then multiply it with the number of hospitals in that bin. Summing these results across all bins yields an estimate of additional procedures attributable to `MVR`.

This calculation reveals an average of approximately 140 additional knee replacement surgeries per year due to `MVR` during the seven years they were in effect. For context, around 150,000 knee replacement surgeries were performed annually during this time. While our analysis detects a behavioral response to `MVR`, the magnitude of this effect at the societal level is relatively small (+0.1%).

3.6 Effect on Misreporting in Quality Reports

The analysis based on billing data likely estimates a lower bound of the actual effect of Minimum Volume Requirements (MVR) on hospital behavior. This is primarily due to the fact that MVR is not strictly enforced in practice. However, there may be indirect effects as well. A hospital's income may be affected by the number of patients seeking treatment there, which in turn could be influenced by whether the hospital reaches the MVR threshold. Even if a hospital is only slightly above the threshold, this may positively affect physicians' and patients' perceptions of the hospital's performance.

Quality reports play a crucial role in this context. Since 2005, German law has required hospitals to publish these reports biennially (Geraedts et al. 2007). Designed to promote transparency and quality assurance, these reports are an integral component of the German healthcare system. Procedure volumes for surgeries subject to MVR are an important element of these reports.

Importantly, the information published and accessible to patients is based on these quality reports rather than billing data. Various patient information platforms, such as the *Weisse Liste* project, utilized these quality reports to help patients make informed decisions about their healthcare providers (Fischer 2015). These platforms have become increasingly important in patients' decision-making processes regarding hospital choice (Fischer et al. 2015). Non-governmental organizations and health insurance companies actively disseminate information about hospitals' MVR compliance through analytical reports (Hemschemeier et al. 2019) and interactive maps (AOK 2024). Consequently, reaching the MVR threshold can have significant reputational benefits for hospitals, potentially resulting in increased patient numbers.

Given these incentives, hospitals may strategically report meeting MVR thresholds in their quality reports, even if their actual procedure volumes fall short. This behavior could maintain their reputation and attract patients,

despite not meeting technical reimbursement requirements. To investigate such strategic reporting, particularly around `MVR` thresholds, we compare procedure counts from quality reports with those from billing data.

We merge quality report and billing data using institutional hospital identifiers. To ensure data quality, we compare annual procedure counts from both sources, applying criteria to retain reasonable matches while excluding likely mismatches[9]. A detailed description of the matching process can be found in the appendix.

In our matched sample, procedure counts from billing and quality reports generally align well, though small discrepancies are common. This is not unexpected, given the complexity of hospital management structures and the fact that billing and quality reporting are distinct processes, typically performed by different personnel at separate times. However, systematic discrepancies, particularly around the `MVR` thresholds, could indicate strategic reporting behavior.

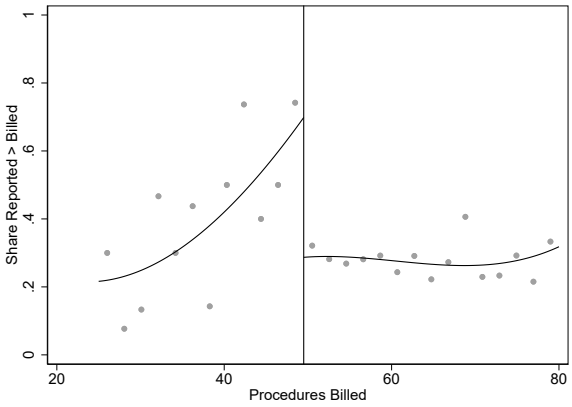
To investigate potential strategic reporting behavior, we present graphical evidence in the spirit of a regression discontinuity design. This approach allows us to examine whether hospitals are more likely to overreport procedure counts in their quality reports when their actual (billed) procedure counts are just below the `MVR` threshold.

Our analysis uses the count of billed procedures as the running variable and the share of hospitals reporting more procedures in their quality reports than they billed as the outcome. Specifically, we bin hospitals by their count of billed procedures and calculate, for each bin, the proportion of hospitals that report a larger procedure count in their quality report compared to their billing data.

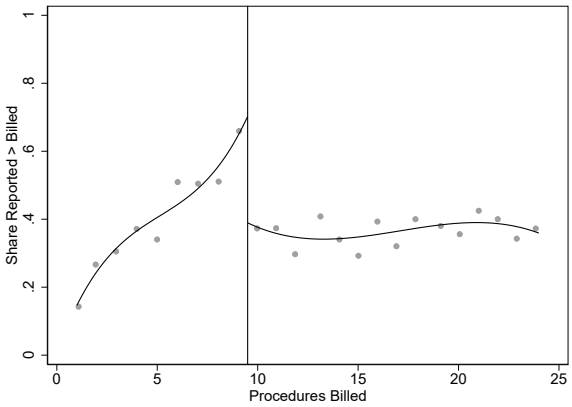
Figure 3.9 illustrates this analysis for knee replacements, pancreatic

[9]Nimptsch et al. (2016) document that hospital identifiers may vary over time and between data sources. De Cruppé (2024) describes challenges in defining and identifying hospital locations that have affected data accuracy over the years.

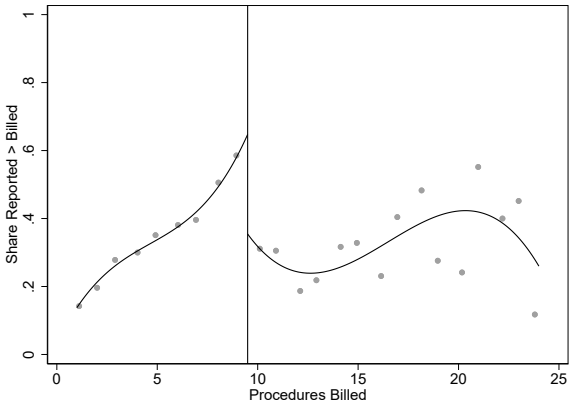
Figure 3.9: Share of Hospitals Overreporting Procedure Counts



(a) Knee



(b) Pancreas



(c) Esophagus

Notes: This graph shows the share of hospitals overreporting procedure totals. The x-axis categorizes hospitals by total procedures from billing data, with discrete bins for (b) and (c), and aggregated bins for (a) for data confidentiality. The y-axis displays the fraction of hospitals in each bin reporting higher counts in quality reports compared to their billing totals. The vertical line indicates the minimum threshold. A third-degree polynomial is fitted on both sides of the threshold.

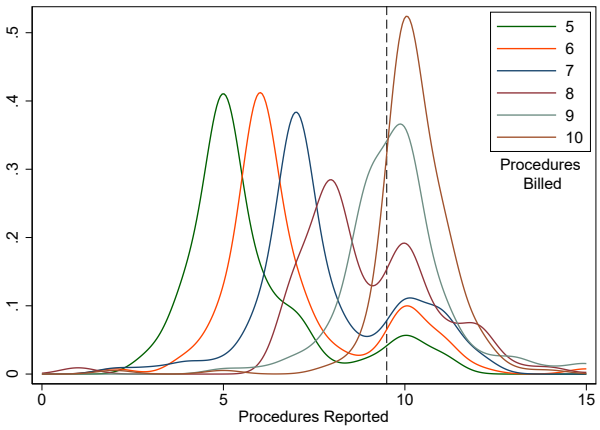
surgeries, and esophageal surgeries. While 20-40% of hospitals consistently report more procedures in quality reports than they bill, this proportion increases markedly for hospitals with billed procedure counts just below the MVR threshold. This discontinuity at the threshold suggests that hospitals billing slightly under the minimum requirement are significantly more likely to overreport their procedure counts in quality reports.

This pattern indicates a systematic response to MVR policies. Hospitals just below the MVR threshold appear more likely to inflate their reported procedure counts, aligning with our hypothesis that they have strong incentives to report meeting MVR thresholds.

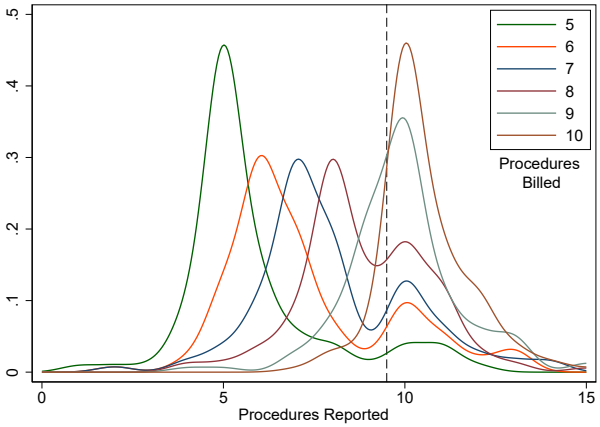
Figure 3.10 presents kernel density estimates of reported procedure totals for each surgery type. Each distribution represents hospitals that billed exactly n procedures. For most cases, the estimated distribution approximates a normal distribution centered around the billed total, indicating general consistency between billed and reported figures with some noise.

However, we consistently observe a bi-modal distribution with a secondary peak just above the MVR threshold. This pattern indicates that a subset of hospitals overstates their procedure volumes to appear compliant with the minimum threshold requirements. Notably, the proportion of such discrepancies increases as the billed total approaches the threshold. This clear relationship between threshold proximity and misreporting likelihood provides further evidence of strategic reporting behavior. Such behavior aligns with our earlier discussion of hospitals' incentives to maintain reputation and attract patients through favorable quality reports, even when actual procedure volumes fall short of MVR thresholds.

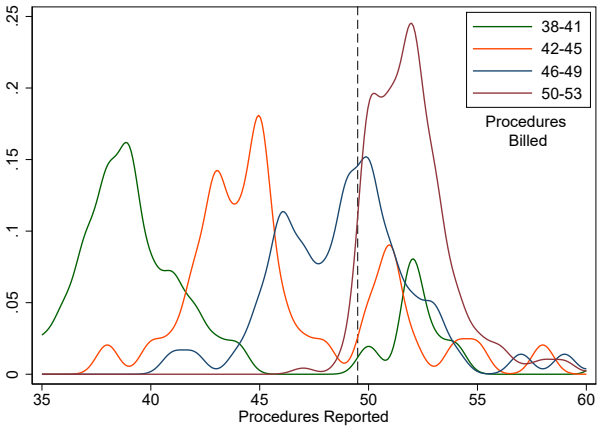
Figure 3.10: Distribution of Reported Procedures per Billed Procedure Count



(a) Esophagus



(b) Pancreas



(c) Knee

Notes: Each panel shows Gaussian kernel density estimates of reported procedure total distributions in quality reports, conditional on billed procedures (indicated by colors). Totals refer to calendar years. The dashed line represents the `MVR` threshold. Data include all years with mandatory quality reports, excluding knee replacements for the years when `MVR` were suspended.

3.7 Effects on Overtreatment and Health

We have shown that minimum volume requirements (MVR) affect hospital volume in the context of knee replacement surgeries, but the underlying mechanisms remain unclear. In Section 3.5.2, we only managed to rule out simple rescheduling explanations.

To explore the underlying mechanisms, we now examine whether the characteristics of the patients who receive surgery, particularly those at the margin of treatment, change in response to MVR. By marginal patients, we refer to individuals who might not have undergone surgery in the absence of hospitals' efforts to meet MVR thresholds. We focus on changes in the age and geographical distribution of patients. Hospitals might adapt to MVR by relaxing patient selection criteria or expanding their catchment area. Alternatively, they could intensify outreach efforts through existing channels, such as partnering orthopedists, potentially redirecting patients from other hospitals. Understanding these mechanisms is crucial: relaxed selection criteria could lead to overtreatment, where surgery is provided to patients who might not otherwise receive it, while patient redirection might only affect quality through changes in hospital volume.

Furthermore, we assess whether these potential changes lead to adverse health outcomes for these marginal patients. We do this by analyzing the effect on a proxy for post-surgery health (length of stay) and the incidence of complications related to the surgery.

To address these questions, we employ an instrumental variable approach to assess whether hospitals under pressure to meet MVR thresholds change their strategic behavior. We estimate the level of pressure hospitals face related to the count of previous procedures and the time of the year and use this as an instrument for the number of weekly knee replacement procedures.

3.7.1 Empirical Strategy

We first estimate a baseline of how hospital procedure counts vary as a function of the calendar week and the number of previously performed procedures in a scenario without minimum volume requirements (`MVR`). We then compare the resulting predictions to true procedure counts in the setting with `MVR` and define deviations from these predictions as *pressure* at the hospital level for any given week and count of previous procedures. This *pressure* measure captures both when and to what extent hospitals respond by increasing procedure volumes above predicted levels.

This *pressure* serves as an instrument for the number of weekly procedures. Subsequently, we utilize a two-stage least squares approach at the patient level to observe the effect of this pressure on individual characteristics and outcomes. 2SLS estimates can be interpreted as the effect of `MVR`-induced increases in procedure counts. Throughout our analysis, we will control for hospital- and year- fixed effects.

Our instrument, *pressure*, is relevant by design. A possible concern for the exclusion restriction is if *pressure* affects patient outcomes directly, not just through the number of procedures performed. However, this is unlikely because *pressure* is constructed solely in relation to procedure counts. Even if such a violation occurred — for example, if increased pressure led to lower quality of care independent of procedure count — a reduced-form version of our analysis would still be meaningful, as we are primarily interested in the overall effect of *pressure* on patient outcomes, capturing both direct and indirect pathways.

3.7.2 Implementation

We first create a hospital-level panel dataset with weekly procedure counts and a cumulative count of procedures performed at each hospital year-to-date.

This cumulative count is calculated by summing the procedure counts c of all previous weeks $1, w-1$:

$$z_{wj} = \sum_{t=1}^{w-1} c_{tj} \tag{3.3}$$

We then model the number of procedures in a given week as a function of the cumulative number of previous procedures and the time of year (week):

$$c_{wj} = f(w, z_{wj}) \tag{3.4}$$

c_{wj} represents the procedure count in week w for hospital j, w denotes the current calendar week, and z_{wj} is the stock of previously performed procedures at hospital j. To estimate this relationship, we employ a flexible specification that includes polynomials to account for non-linear trends:

$$c_{wj} = \beta_0 + \sum_{p \in 0.5,1,2,3,4} \beta_p z_{wj}^p + \alpha_w + \sum_{p \in 0.5,1,2,3,4} \delta_{pw} z_{wj}^p + \epsilon_{wj} \tag{3.5}$$

In this equation, β_0 represents the constant term, and β_p are coefficients for polynomials of z_{wj} (including a square root form $\sqrt{z_{wj}}$). Week fixed effects are captured by α_w, while δ_{pw} represents week fixed effects interacted with polynomials of z_{wj}. The error term is denoted by ϵ_{wj}.

To establish a baseline predictive pattern reflecting a hypothetical scenario without minimum volume requirements (`MVR`), we exclude hospital-week observations potentially influenced by `MVR`. We accomplish this by removing hospitals in years with `MVR` in effect that are on pace to narrowly reach or miss the minimum volume threshold. Using this adjusted sample, we estimate Equation 3.5.

We then use the full hospital-week-level sample and calculate the difference between observed values c_{wj} and predicted values from equation 3.5 $\hat{c}_{wj}$.

$$\Delta c_{wj} = c_{wj} - \hat{c}_{wj} \tag{3.6}$$

The underlying intuition is that hospitals reacting to minimum volume requirements (MVR) will systematically deviate from their predicted procedure counts. To estimate the timing and magnitude of this reaction, we use data from years when MVR were in effect and estimate the following equation:

$$\Delta c_{wj} = \beta_0 + (\sum_{p \in 0.5,1,2,3,4} \beta_p z_{wj}^p + \alpha_w + \sum_{p \in 0.5,1,2,3,4} \delta_{pw} z_{wj}^p) \times T_{wj} + \epsilon_{wj} \quad (3.7)$$

This equation is similar to Equation 3.5, but includes T_{wj}, a dummy variable that indicates whether hospital j has reached the minimum volume threshold in week w. We interact all variables with T_{wj} to estimate separate coefficients based on whether the MVR threshold has been met. This allows for a potential discontinuous change in pressure at the threshold. We continue to use a flexible polynomial form for z_{wj}.

The resulting predicted values, $\widehat{\Delta c_{wj}}$, provide an estimate of *pressure* for each combination of calendar week and previous procedure count. These values represent how much hospitals, on average, exceed their predicted surgery volumes based on their prior procedure count and the time of year. We interpret $\widehat{\Delta c_{w,j}}$ as the average hospitals' response in procedure counts to the minimum volume requirements.

3.7.3 Two-Stage Least Squares Estimation

We now use our measure of *pressure* as an instrument for weekly surgical volume to estimate the effect of MVR-influenced surgical volume on patient characteristics and patient health. For this analysis, we move to patient-level data, incorporating hospital and year fixed effects.

Our analysis focuses on two main sets of outcome variables. First, we examine patient characteristics to assess whether the marginal patient — the additional patient when a hospital is under pressure to meet the MVR — differs

from the average patient. We evaluate mean age and changes in the share of patients aged over 80 and under 60 to capture potential shifts in the age distribution. We also consider whether patients live close to the hospital (in the same district), as this can indicate whether hospitals are expanding their catchment areas.

By examining these variables, we aim to determine whether increased surgical volumes due to `MVR` pressure lead to overtreatment, potentially resulting in hospitals treating patients who are less suitable candidates for surgery or who might not have otherwise undergone the procedure.

Second, we assess health outcomes using the length of hospital stay and the incidence of surgical complications. While we cannot directly evaluate long-term outcomes or readmissions, the length of stay reflects recovery speed and patient mobility, while complications provide a direct measure of surgical quality. The within-hospital analysis helps to mitigate concerns that hospital-specific discharge policies could influence these results.

In this framework, *pressure* captures the extent to which hospitals increase procedure volumes in response to `MVR`, potentially leading to overtreatment. By instrumenting surgical volume with *pressure*, we can isolate the effect of `MVR`-induced volume increases on patient characteristics and outcomes.

The two-stage least squares estimation is structured as follows:

First stage:

$$c_{ywj} = \alpha_0 + \alpha_1 \widehat{\Delta c_{wj}} + \gamma_j + \delta_y + \nu_{ywj} \tag{3.8}$$

Second stage:

$$y_{iywj} = \beta_0 + \beta_1 \widehat{c_{ywj}} + \gamma_j + \delta_y + \epsilon_{iywj} \tag{3.9}$$

$c_{y,w,j}$ is the procedure count in hospital j in calendar week w in year y. $\widehat{\Delta c_{w,j}}$ is the previously constructed instrumental variable *pressure*. γ_j and δ_y are hospital and year fixed-effects. y_{iywj} is the outcome variable for patient i,

Table 3.3: Two-Stage Least Squares Estimates: Pressure and Outcomes

First Stage						
			Procedure Count			
Pressure			0.847***			
			(0.065)			
F-statistic			167.7			
Second Stage						
	(1) Lives Close to Hospital	(2) Age	(3) Age > 80	(4) Age < 60	(5) Complications	(6) Length of Stay
Procedure Count	-0.030*	0.160	0.009	0.001	0.002	0.121
	(0.012)	(0.249)	(0.009)	(0.010)	(0.003)	(0.166)
Observations	57,195	57,195	57,195	57,195	57,195	57,195

Notes: This table presents IV-FE estimates of the effect of weekly knee procedures attributable to minimum volume requirement pressure on various outcomes. The first stage shows the relationship between pressure and procedure count, with the F-statistic for instrument strength. The second stage presents IV estimates. Outcome variables: Lives Close to Hospital (dummy whether patient resides in same district), Age (in years and as dummies), Surgery complications (ICD Code T.85), and Length of Stay (in days). Hospital- and Year-Fixed Effects included. Standard errors clustered at hospital level. Data includes only years with `MVR` in effect, excluding hospitals with predicted year-end totals exceeding 100 procedures. Standard errors in parentheses. * p<0.05, ** p<0.01, *** p<0.001.

while ν_{ywj} and ϵ_{iywj} are the error terms.

Table 3.3 presents the main results of our instrumental variable analysis. The first-stage results demonstrate the relationship between our instrument and the number of surgeries performed. As expected, this relationship is strong by construction (F=167.7).

The coefficients of interest β_1 can be interpreted as the effect of one additional weekly patient attributable to minimum volume regulations (`MVR`) on the average patient characteristics at a hospital in a given week. It's crucial to note that the observed increases in patient numbers due to `MVR` were approximately 2 patients per month, or 0.5 patients per week. For context, the mean number of weekly procedures was around 1.5 in November and 0.6 in December (Figure 3.4). Consequently, these coefficients likely underestimate the effects on the marginal patient.

Column (1) reveals a statistically significant decrease of 3 percentage

points in the share of patients living close to the hospital. This indicates that hospitals might be expanding their catchment area in response to MVR pressure, potentially to meet volume requirements. Columns (2) through (4) report effects on age-related variables, all of which are small and statistically insignificant. Columns (5) and (6) show the effects on health outcomes. We find no statistically significant effect on complications, and the estimated increase in length of stay is statistically indistinguishable from zero.

In sum, we detect changes in patient composition due to MVR pressure, particularly in terms of geographical distribution, with hospitals drawing patients from a broader area as indicated by the decrease in the proportion of patients living close to the hospital. However, we find no significant effects on health outcomes, such as complications or length of stay, and no evidence that hospitals are lowering their selection criteria in terms of patient age.

These findings suggest that hospitals respond to MVR pressure by expanding their reach without compromising patient selection standards or health outcomes, providing little evidence of overtreatment.

We acknowledge limitations inherent to our approach. First, statistical power is constrained due to the limited response of hospitals to the MVR. Second, data limitations preclude us from making statements about pre-surgery dynamics or post-surgery long-term outcomes.

3.8 Discussion

This study reveals several important effects of Minimum Volume Requirements (MVR) on hospitals. We observe a substantial increase in knee replacement procedures performed by hospitals near the minimum threshold at year-end, suggesting strategic behavior to meet volume requirements. However, this effect is not present for complex pancreatic and esophageal surgeries.

These findings demonstrate that while MVR aim to improve patient outcomes

by centralizing care, hospitals may respond in ways that prioritize meeting volume thresholds over enhancing surgical quality. The differential response between knee replacements and complex pancreatic and esophageal surgeries highlights how the complexity and interdisciplinary nature of decision-making in the latter may limit hospitals' ability to adjust volumes in response to MVR.

Our discovery of misreporting in hospital quality reports near the minimum volume threshold raises concerns about their integrity, potentially misleading patients and compromising policymakers' ability to accurately assess healthcare quality.

The marginal additional patients due to MVR pressure are more likely to live farther from the hospital, indicating that hospitals are expanding their catchment areas to meet volume requirements without altering clinical eligibility criteria. The age profiles of patients treated under increased MVR pressure remain consistent with the overall patient population, indicating that hospitals maintain consistent clinical criteria for treatment eligibility. This suggests that hospitals are adapting their operational strategies rather than modifying clinical criteria to meet MVR thresholds. Such behavior indicates that MVR policies may be inducing changes in hospital management practices and strategies to attract patients.

Overall, hospitals employ two main strategies in response to MVR: increasing procedure volumes near year-end to meet thresholds, and misreporting data in quality reports to appear compliant. The former suggests a direct response to potential reimbursement incentives, despite their infrequent enforcement, while the latter likely stems from reputational considerations, as these reports serve as a primary source of information for public quality assessment platforms. The coexistence of these strategies highlights the multi-faceted nature of strategic hospital responses to MVR, encompassing both actual changes in medical practice and manipulation of reported data.

While our study reveals some unintended consequences of MVR, it is

important to consider these in the broader context of healthcare policy. Despite these challenges, the observed reduction in low-volume providers suggests that MVR effectively promote centralization, potentially improving patient outcomes as intended. Figure 3.10 shows that not all hospitals misreport, and Figure 3.1 can be interpreted as evidence of successful centralization via MVR.

Thus, MVR appear to be an effective centralization policy, as their benefits in concentrating procedures may outweigh the modest unintended consequences we've identified. MVR may avoid certain drawbacks associated with alternatives such as central planning or certification systems, including reduced patient choice and increased administrative burden on hospitals.

Policymakers should consider enhancing enforcement mechanisms, such as regular audits, and improving transparency in quality reporting to ensure compliance and maintain public trust. Weak enforcement and inadequate transparency can undermine the effectiveness of MVR policies, making it difficult for patients and providers to navigate the system. By strengthening enforcement and improving reporting accuracy, policymakers can maximize the benefits of centralization and reduce unintended negative consequences.

References

Alexander, D. (2020). "How Do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs". *Journal of Political Economy* 128.11: 4046–4096.

AOK (2024). "Mindestmengen-Transparenzliste". Technical Report.

Arias-de la Torre, J., Valderas, J. M., Evans, J. P., Martín, V., Molina, A. J., Muñoz, L., Pons-Cabrafiga, M., Espallargues, M., and Catalan Arthroplasty Register Steering Committee (RACat) (2019). "Differences in Risk of Revision and Mortality Between Total and Unicompartmental Knee Arthroplasty. The Influence of Hospital Volume". *The Journal of*

Arthroplasty 34.5: 865–871.

Auras, S., Cruppé, W. de, Blum, K., and Geraedts, M. (2012). "Mandatory quality reports in Germany from the hospitals' point of view: a cross-sectional observational study". *BMC Health Services Research* 12.1: 378.

Batty, M. and Ippolito, B. (2017). "Financial Incentives, Hospital Care, and Health Outcomes: Evidence from Fair Pricing Laws". *American Economic Journal: Economic Policy* 9.2: 28–56.

Birkmeyer, J. D. and Lucas, F. L. (2002). "Hospital Volume and Surgical Mortality in the United States". *The New England Journal of Medicine* 346.15: 1128–1137.

Cattaneo, M. D., Crump, R. K., Farrell, M. H., and Feng, Y. (2024). "On Binscatter". *American Economic Review* 114.5: 1488–1514.

Červený, J. (2023). "Selective referral or learning by doing? An analysis of hospital volume-outcome relationship of vascular procedures". *Health Economics* 32.6.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from Danish tax records". *The quarterly journal of economics* 126.2: 749–804.

Dafny, L. S. (2005). "How Do Hospitals Respond to Price Changes?" *American Economic Review* 95.5: 1525–1547.

De Cruppé, W. (2024). "Umsetzung der Mindestmengenregelung in deutschen Krankenhäusern". *Dissertation*.

De Cruppé, W. and Geraedts, M. (2018). "Mindestmengen unterschreiten, Ausnahmetatbestände und ihre Konsequenzen ab 2018. Komplexe Eingriffe am Ösophagus und Pankreas in deutschen Krankenhäusern im Zeitverlauf von 2006 bis 2014". *Zentralblatt für Chirurgie* 143.03: 250–258.

De Cruppé, W., Ortwein, A., Kraska, R. A., and Geraedts, M. (2020). "Impact

of suspending minimum volume requirements for knee arthroplasty on hospitals in Germany: an uncontrolled before–after study". *BMC Health Serv Res* 20.1109.

Dudley, R. A., Johansen, K. L., Brand, R., Rennie, D. J., and Milstein, A. (2000). "Selective Referral to High-Volume Hospitals: Estimating Potentially Avoidable Deaths". *JAMA* 283.9: 1159–1166.

El Saghir, N. S., Keating, N. L., Carlson, R. W., Khoury, K. E., and Fallowfield, L. (2014). "Tumor Boards: Optimizing the Structure and Improving Efficiency of Multidisciplinary Management of Patients with Cancer Worldwide". *American Society of Clinical Oncology Educational Book* 34: e461–e466.

Eliason, P. J., Grieco, P. L. E., McDevitt, R. C., and Roberts, J. W. (2018). "Strategic Patient Discharge: The Case of Long-Term Care Hospitals". *American Economic Review* 108.11: 3232–3265.

Ettelt, S. (2017). "The Politics of Evidence Use in Health Policy Making in Germany—the Case of Regulating Hospital Minimum Volumes". *J Health Polit Policy Law* 42.3: 513–538.

Fischer, S. (2015). "Project "Weisse Liste": A German Best Practice Example for Online Provider Ratings in Health Care". *Challenges and Opportunities in Health Care Management*. Ed. by S. Gurtner and K. Soyez. Cham: Springer International Publishing: 339–346.

Fischer, S., Pelka, S., and Riedl, R. (2015). "Understanding patients' decision-making strategies in hospital choice: Literature review and a call for experimental research". *Cogent Psychology* 2.1.

Geraedts, M., Schwartze, D., and Molzahn, T. (2007). "Hospital quality reports in Germany: patient and physician opinion of the reported quality indicators". *BMC Health Services Research* 7.1: 157.

Hamilton, B. H. and Ho, V. (1998). "Does Practice Make Perfect?: Examining the Relationship between Hospital Surgical Volume and Outcomes for Hip

Fracture Patients in Quebec". *Medical Care* 36.6: 892–903.

Hemschemeier, M., Bittkowsk, M., and Stollorz, V. (2019). "Mindestmengen im Krankenhaus-Bilanz und Neustart". Technical Report. Bertelsmann Stiftung.

Hentschker, C. and Mennicken, R. (2018). "The Volume-Outcome Relationship Revisited: Practice Indeed Makes Perfect". *Health Services Research* 53.1: 15–34.

Kleven, H. J. (2016). "Bunching". *Annual Review of Economics* 8: 435–464.

Krautz, C., Nimptsch, U., Weber, G. F., Mansky, T., and Grützmann, R. (2018). "Effect of Hospital Volume on In-hospital Morbidity and Mortality Following Pancreatic Surgery in Germany". *Annals of Surgery* 267.3: 411–417.

Kugler, C. M., Goossen, K., Rombey, T., De Santis, K. K., Mathes, T., Breuing, J., Hess, S., Burchard, R., and Pieper, D. (2022). "Hospital volume–outcome relationship in total knee arthroplasty: a systematic review and dose–response meta-analysis". *Knee Surgery, Sports Traumatology, Arthroscopy* 30.8: 1795.

Levaillant, M., Marcilly, R., Levaillant, L., Michel, P., Hamel-Broza, J.-F., Vallet, B., and Lamer, A. (2021). "Assessing the hospital volume-outcome relationship in surgery: a scoping review". *BMC Medical Research Methodology* 21.1: 204.

Lidsky, M. E., Sun, Z., Nussbaum, D. P., Adam, M. A., Speicher, P. J., and Blazer, D. G. (2017). "Going the Extra Mile: Improved Survival for Pancreatic Cancer Patients Traveling to High-volume Centers". *Annals of Surgery* 266.2: 333–338.

Luft, H. S., Bunker, J. P., and Enthoven, A. C. (1979). "Should Operations Be Regionalized? The Empirical Relation between Surgical Volume and Mortality". *New England Journal of Medicine* 301.25: 1364–1369.

Lützner, J., Lange, T., Schmitt, J., Kopkow, C., Aringer, M., Böhle, E.,

Bork, H., Dreinhöfer, K., Friederich, N., Gravius, S., Heller, K.-D., Hube, R., Gromnica-Ihle, E., Kirschner, S., Kladny, B., Kremer, M., Linke, M., Malzahn, J., Sabatowski, R., Scharf, H.-P., Stöve, J., Wagner, R., and Günther, K.-P. (2018). "The S2k guideline: Indications for knee endoprosthesis: Evidence and consent-based indications for total knee arthroplasty". *Der Orthopäde* 47.9: 777–781.

McGuigan, A., Kelly, P., Turkington, R. C., Jones, C., Coleman, H. G., and McCain, R. S. (2018). "Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes". *World Journal of Gastroenterology* 24.43: 4846–4861.

Morche, J., Renner, D., Pietsch, B., Kaiser, L., Brönneke, J., Gruber, S., and Matthias, K. (2018). "International comparison of minimum volume standards for hospitals". *Health Policy* 122.11: 1165–1176.

Nimptsch, U., Haist, T., Krautz, C., Grützmann, R., Mansky, T., and Lorenz, D. (2018). "Hospital Volume, In-Hospital Mortality, and Failure to Rescue in Esophageal Surgery". *Deutsches Ärzteblatt international* 7.e016184.

Nimptsch, U. and Mansky, T. (2017). "Hospital volume and mortality for 25 types of inpatient treatment in German hospitals: observational study using complete national data from 2009 to 2014". *BMJ Open* 7.9: e016184.

Nimptsch, U., Peschke, D., and Mansky, T. (2017). "Mindestmengen und Krankenhaussterblichkeit – Beobachtungsstudie mit deutschlandweiten Krankenhausabrechnungsdaten von 2006 bis 2013". *Das Gesundheitswesen* 79.10: 823–834.

Nimptsch, U., Wengler, A., and Mansky, T. (2016). "Kontinuität der Institutionskennzeichen in Krankenhausabrechnungsdaten – Analyse der bundesweiten DRG-Statistik von 2005 bis 2013". *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen* 117: 38–44.

Pieper, D., Mathes, T., Neugebauer, E., and Eikermann, M. (2013). "State of evidence on the relationship between high-volume hospitals and outcomes

in surgery: a systematic review of systematic reviews". *Journal of the American College of Surgeons* 216.5: 1015–1025e18.

Pillay, B., Wootten, A. C., Crowe, H., Corcoran, N., Tran, B., Bowden, P., Crowe, J., and Costello, A. J. (2016). "The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: A systematic review of the literature". *Cancer Treatment Reviews* 42: 56–72.

Ratnayake, B., Pendharkar, S. A., Connor, S., Koea, J., Sarfati, D., Dennett, E., Pandanaboyana, S., and Windsor, J. A. (2022). "Patient volume and clinical outcome after pancreatic cancer resection: A contemporary systematic review and meta-analysis". *Surgery* 172.1: 273–283.

RDC of the Federal Statistical Office and Statistical Offices of the Federal States (2017). "DRG-Statistic, 2008 to 2017".

Scharfe, J., Pfisterer-Heise, S., Kugler, C. M., Shehu, E., Wolf, T., Mathes, T., and Pieper, D. (2023). "The effect of minimum volume standards in hospitals (MIVOS) — protocol of a systematic review". *Systematic Reviews* 12.1: 11.

Schlottmann, F., Strassle, P. D., Charles, A. G., and Patti, M. G. (2018). "Esophageal Cancer Surgery: Spontaneous Centralization in the US Contributed to Reduce Mortality Without Causing Health Disparities". *Annals of Surgical Oncology* 25.6: 1580–1587.

Urbach, D. R. (2015). "Pledging to Eliminate Low-Volume Surgery". *New England Journal of Medicine* 373.15: 1388–1390.

Vonlanthen, R., Käser, S., and Clavien, P.-A. (2021). "Centralization in Surgery in European Countries". *Volume-Outcome Relationship in Oncological Surgery*. Springer International Publishing: 145–159.

Vonlanthen, R., Lodge, P., Barkun, J. S., Farges, O., Rogiers, X., Soreide, K., Kehlet, H., Reynolds, J. V., Käser, S. A., Naredi, P., Borel-Rinkes, I., Biondo, S., Pinto-Marques, H., Gnant, M., Nafteux, P., Ryska, M.,

Bechstein, W. O., Martel, G., Dimick, J. B., Krawczyk, M., Oláh, A., Pinna, A. D., Popescu, I., Puolakkainen, P. A., Sotiropoulos, G. C., Tukiainen, E. J., Petrowsky, H., and Clavien, P.-A. (2018). "Toward a Consensus on Centralization in Surgery". *Annals of Surgery* 268.5: 712–724.

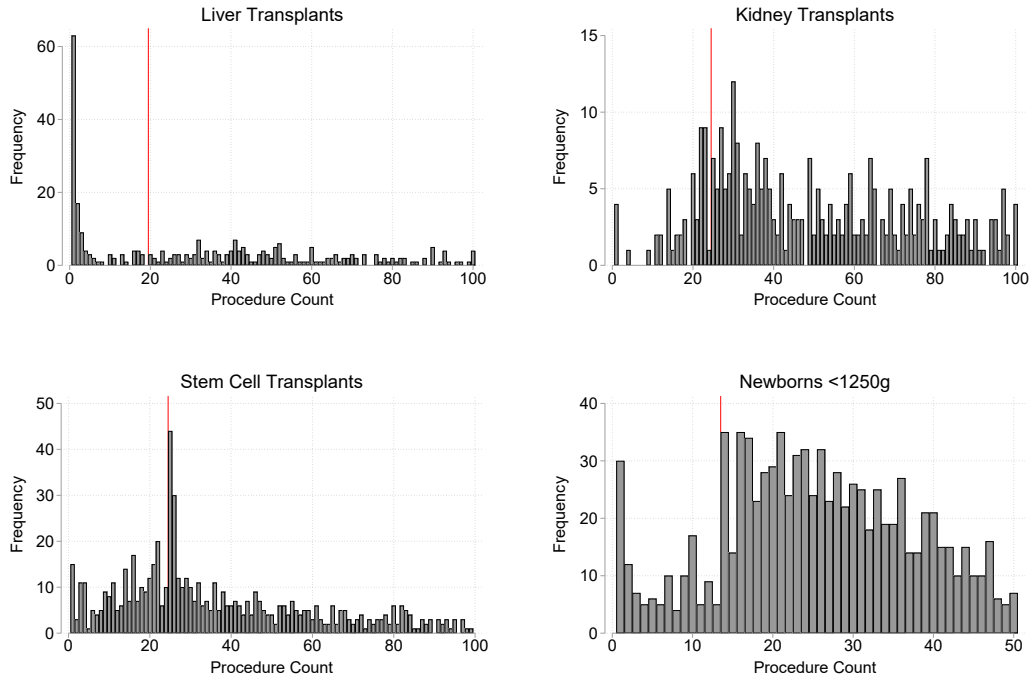
Wengler, A., Nimptsch, U., and Mansky, T. (2014). "Hip and Knee Replacement in Germany and the USA". *Dtsch Arztebl International* 111.23-24: 407–416.

Yang, J., Liu, X., Cao, S., Dong, X., Rao, S., and Cai, K. (2020). "Understanding Esophageal Cancer: The Challenges and Opportunities for the Next Decade". *Frontiers in Oncology* 10.

Appendix to Chapter 3

Figures

Figure 3.11: Distribution of Reported Procedure Totals



Notes: These histograms depict the frequency distribution of reported procedures in quality reports across all available data years for liver transplants, kidney transplants, stem cell transplants, and newborn care under 1250g. Red vertical lines indicate the minimum volume thresholds relevant for each procedure. The upper tails of the distributions are truncated for visual clarity.

Data Matching Process For Comparison of Procedure Totals

We matched two data sources: Quality Reports (Quality Report Data) and DRG (Diagnosis-Related Group) data (DRG Billing Data). The matching process aimed to link hospital records across these sources using hospital identifiers ("IK-Nummer") and year.

Data Overview

Quality Report Data:

- Esophageal Surgery (8 years): 2,246 observations (approximately 281 hospitals per year)
- Pancreatic Surgery (8 years): 3,685 observations (approximately 461 hospitals per year)
- Knee Replacement (5 years, only when MVR in effect): 4,313 observations (approximately 863 hospitals per year)

DRG Billing Data:

- Esophageal Surgery (10 years): 4,339 observations (approximately 434 hospitals per year)
- Pancreatic Surgery (10 years): 6,814 observations (approximately 681 hospitals per year)
- Knee Replacement (10 years): 10,814 observations (approximately 1,081 hospitals per year)

Matching Process and Algorithm

The matching process involved an initial merge based on hospital identifier and year, followed by the creation of deviation variables to assess match quality. We calculated absolute and percentage deviations for each procedure and computed a total deviation quotient across all procedures. When multiple potential matches existed, we retained the matches with lower deviation quotients. We applied stricter matching criteria for cases with higher deviations or missing data, including the retention of best matches and the application of threshold criteria to identify and handle discrepancies.

Chapter 4

Geographic Inequality in Income and Mortality in Germany

This chapter is co-authored with Amelie Wuppermann, Joachim Winter, Hannes Schwandt, and Janet Currie. It was published in Fiscal Studies, vol. 42, no. 1 (2021).

FISCAL STUDIES, vol. 42, no. 1, pp. 147–170 (2021) 0143-5671

Geographic Inequality in Income and Mortality in Germany*

Peter Redler,† Amelie Wuppermann,‡ Joachim Winter,† Hannes Schwandt§ and Janet Currie◊

†*University of Munich*
(peter.redler@econ.lmu.de, winter@lmu.de)
‡*University of Halle-Wittenberg*
(amelie.wuppermann@wiwi.uni-halle.de)
§*Northwestern University*
(schwandt@northwestern.edu)
◊*Princeton University*
(jcurrie@princeton.edu)

Abstract

We use data from the German Federal Statistical Office on population counts, births, deaths and income to study the development of socio-economic inequality in mortality rates from 1990 to 2015 for different age groups and both genders. Ranking the 401 German districts by average disposable income per capita, we observe large inequalities in district-level mortality rates in 1990, which had almost disappeared, or at least been flattened considerably, by 2015 particularly for infants, children and the very old. The most important driver of this reduction in inequality is German reunification in 1990. As indicated by more detailed analyses comparing districts in the former East and the former West, even five years after reunification there was a large gap in disposable income, with all Eastern districts considerably poorer than the

*Submitted November 2020.
P. Redler gratefully acknowledges support from the Elite Network of Bavaria within the Evidence-Based Economics programme. Daniel Saggau provided excellent research assistance.
Keywords: Germany, health inequality, income, inequality, mortality, SES.
JEL classification numbers: I12, I14, I18, J10.

poorest district in the West. At the same time, mortality rates were higher for all age groups and both genders in the East. Income has caught up, to the extent that there are equally poor districts in the East and West in most recent years (although the West is still much richer on average). Mortality rates in the East have improved considerably and are even below mortality rates for similarly poor districts in the West in the most recent data.

I. Introduction

Inequalities in mortality provide a stark measure of unequal life chances and hence are of fundamental interest and importance. Yet, they can be difficult to measure. Currie and Schwandt (2016a, 2016b) proposed an alternative method of examining inequalities in mortality, which allowed all of the deaths for the entire population to be included. As described in the introduction to this special issue, they first ranked areas from richest to poorest and then grouped them into 20 'bins' of approximately equal size. Age-specific mortality rates were then examined by age and gender groups for each bin in census years. This way of examining inequality in mortality has the advantage that it is able to include all deaths in a consistent way over time.

In this paper, we apply these methods to data from Germany for 1990, 2003 and 2015. This time period is of special interest in German history because it follows reunification. As of 3 October 1990, the German Democratic Republic (GDR, East Germany) no longer existed and the Federal Republic of Germany absorbed five new states from the East. East and West Berlin were also reunited and the unified city became the capital of the newly united country. However, as we show below, stark differences in income and mortality still divided the East and the West, though the East began to catch up. Hence, Germany over these 25 years offers a unique setting for examining the relationship between income and mortality.

We find that inequality in mortality rates has decreased considerably since 1990, driven mainly by rapid reductions in mortality in the East. Our most striking finding is that although the East remains poorer, for many age and gender groups, the East has overtaken the West in terms of lowering mortality. Conditional on income, the East has lower infant mortality, and lower mortality for adult women aged 20–79. Men in the East now largely match men in the West in terms of mortality, conditional on income.

These results suggest once again that although higher income tends to be linked to better health, the two do not move in lock step. It was possible for the East to close large gaps in mortality rates even though gaps in income remained. A better understanding of how the East has been able to lower mortality, especially for adult women and infants, could shed light on how to eliminate the link between socio-economic status and health more generally.

The rest of the paper proceeds as follows. Section II offers a discussion of the evolution of inequality in mortality for men and women across six broad age groups. Section III probes these results in greater detail. In Section IV, we offer an analysis of the evolution of mortality between the German East and West. We end with a summary and conclusions.

II. Evolution of inequality in mortality in six broad age groups

Part A

1. Data

We use data on population counts, deaths, births and income by the German Federal Statistical Office on a district (*Kreise und kreisfreie Städte*) level from 1990 to 2016. There were 401 districts in Germany in 2016 with an average of around 200,000 inhabitants each. The district level is suitable for our analysis as this is the lowest level at which data on mortality and income are easily available for the full population.

The official population counts are based on national census data and yearly large-scale survey data (*Mikrozensus*). The population counts represent the population at the beginning of the respective year.[1] These data contain population totals by five-year age groups with the exception of the youngest age groups (0–2, 3–5 and 6–9). Before 2012, these categories are broader for older ages (65–74 and 75 and older). Because the number of deaths is published for infants below one year of age and mortality in young children is highest in the first year, we supplement these population data with the number of births in a district in a given year to estimate the number of infants.

Mortality data include the number of deaths by age group at the time of death in a given year. Age groups span five years, with the exception of the 0–4 age group, which is split into infant mortality (age 0) and mortality from age 1 to 4.[2]

To construct mortality rates within age groups, we divide the number of deaths in a given year by the total population at the beginning of the year.

[1]The census in 2011 revealed that population estimates for most areas were too high (Scholz and Kreyenfeld, 2016). The downward correction amounted to a drop of around 1.8 per cent of the national population from 2010 to 2011. Crucially, the Federal Statistical Office did not retrospectively adjust population estimates. Klüsener et al. (2018) estimate population counts by state and age for the entire intercensal period from 1987 to 2010. We use their data to adjust population estimates on a district/year/age group level.

[2]For some years and age groups, the age groups for population counts and for the number of deaths do not match perfectly. In this case, we estimate the population of finer age groups using the state-level estimates of Klüsener et al. (2018). We conduct the same procedure to create larger age groups where necessary. The resulting measurement error at the district level is mitigated by binning districts in the subsequent analysis. We proceed in a similar way for the number of deaths for a few states in which they are aggregated at 75 years and older.

When we aggregate five-year age groups, we ensure comparability over time by age-adjusting mortality rates to the German population in 2015. The reason for age adjustment is that in a relatively wide age category such as individuals aged 20–49, a sample that had more individuals aged 49 would be expected to have higher mortality rates only because they were older. Therefore, changes in the age distribution within age categories could complicate the interpretation of changes over time. Note that, due to data limitations, no age adjustment is available beyond age 85.

As a measure of socio-economic status (SES), we use the per capita disposable income in a district. This measure captures post-transfer income (i.e. after government benefits and income taxes). This measure is available from 1995 to 2017. For the three states of Niedersachsen, Mecklenburg-Vorpommern and Schleswig-Holstein, the measure is only available at the state level before 2000. For these states, we extrapolate the pre-2000 state trend to the district level and impute the missing values. We also impute recent district data (2014–17) for one state (Sachsen) in an identical way.

The SES data are available for districts in their current geographical configuration. Over the last 30 years, five states in the East that were part of the former GDR performed major district reforms, mostly consolidating smaller districts into larger ones. We map the districts that no longer exist to their current districts in order to link the SES measures.

2. Income inequality in Germany

Individual income inequality in Germany rose to some extent in the 1990s and more strongly in the early 2000s.[3] This increase was stronger for the East, which was less unequal after the reunification in 1990. The increase for disposable income per capita (our SES variable) was lower than the increase of market income inequality due to the tax and transfer system.[4] The increase in inequality was mainly driven by an increase in wage inequality conditional on employment. This trend stopped around 2005.[5]

Geographic inequality in our SES variable, measured by a standard Gini coefficient, declined over time in our analysis period. This decline is driven by a catch-up of Eastern districts while inequality across districts in the West was stable over time. Inequality across districts in the East also declined (detailed results available upon request).

[3] Fuchs-Schündeln, Krueger and Sommer, 2010; Dustmann, Ludsteck and Schönberg, 2009.
[4] Fuchs-Schündeln , Krueger and Sommer, 2010.
[5] Biewen, Ungerer and Löffler, 2017.

3. Health care provision in Germany

Health care in Germany is provided in a multi-payer system with universal mandatory health insurance. There are two health insurance systems that exist side by side. The public health insurance system (statutory health insurance) covered around 85 per cent of individuals during our study period.[6] In this system, insurance premiums are based on individuals' income and are not risk-rated. Dependants are covered at no charge. Low co-payments for drugs and some treatments were introduced in the mid-1990s. High earners, civil servants and the self-employed can opt out of the public system and buy private health insurance with risk-rated insurance premiums. Health care provision is mostly private, but prices are heavily regulated. Choice of health care provider is not restricted. Health care spending per capita increased from 8 per cent of GDP in 1990 to around 11 per cent of GDP in 2015.[7]

4. Results

Figure 1 displays the distribution of poverty across Germany in the years 1995, 2003 and 2015. As described above, we first rank all German districts by average disposable income per capita in each year and then group districts into 'bins' that each contain roughly 5 per cent of the population. Due to large districts (e.g. the big cities of Hamburg, Berlin and Munich), the exact population represented by a single bin varies between 2.93 and 4.96 million in 1990, 3.67 and 4.38 million in 2003, and 2.56 and 5.23 million in 2015.

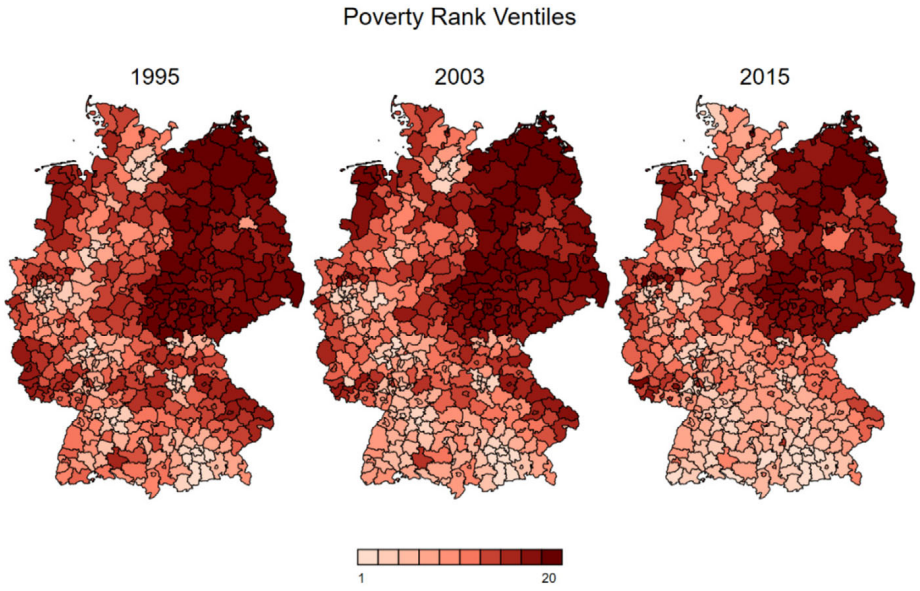
As Figure 1 indicates, in 1995 (i.e. five years after German reunification), the poorest districts in Germany were all exclusively in the East of Germany (the former GDR). Berlin, which belonged in part to the former GDR and in part to the Federal Republic of Germany (BRD) before reunification, is an exception, with much higher income on average than the rest of the former East in 1995. The ranking of districts changed considerably over time, as shown in Figure 1, though even by 2015 the East is still relatively poor compared with the West. However, in 2015, there are equally poor districts in the West, such as in the Ruhr valley where deindustrialisation has had a large impact.[8]

Figure 2 displays the association between a district's poverty rank and mortality rates (per 1,000 inhabitants) for men (top panel) and women (bottom panel). There is a separate figure for each age group (0–4, 5–19, 20–49, 50–64, 65–79 and 80+). Each figure has a line for each of the three years: 1990, 2003 and 2015. Because of the lack of other data, we use disposable income in 1995 to rank districts in 1990. The rankings in 2003 and 2015 are

[6]Busse and Blümel, 2014.
[7]Papanicolas, Woskie and Jha, 2018.
[8]Oei, Brauers and Herpich, 2020.

FIGURE 1

Regional distribution of poverty ranks by year, where darker colours represent lower socio-economic status



based on income data for those years. Mortality rates in 2003 and 2015 are based on three-year averages (2002–04 and 2014–16, respectively), while data restrictions force us to use data from 1990 only to construct the 1990 mortality rate, possibly leading to higher variability in the 1990 rates. Table 1 displays average mortality rates (per 1,000 inhabitants) for the most extreme poverty bins for 1990, 2003 and 2015, separately by age and gender, as well as the slopes of the regression lines depicted in Figure 2. Furthermore, it reports *p*-values for tests of equality of the slopes across years.

The graphs in Figure 2 and the results in Table 1 show that mortality rates dropped considerably between 1990 and 2003, for both genders and across all age groups and all districts. For most age groups and districts, there are further but smaller reductions in mortality between 2003 and 2015. As the overall decline in mortality rates between 1990 and 2015 is particularly large for the poorer districts, inequality in mortality has dropped considerably in Germany, with no inequality remaining for children (5–19) and girls (0–4), for whom the gradient in mortality across districts is not significant in 2015 (Table 1), and older women (65+), for whom the gradient is still significantly different from zero in 2015 but comparably small.

It is striking that in many of the graphs the three points representing regions at the 85th percentile of poverty or higher showed much higher mortality

FIGURE 2

Yearly mortality per 1,000 by age group, gender and poverty rank

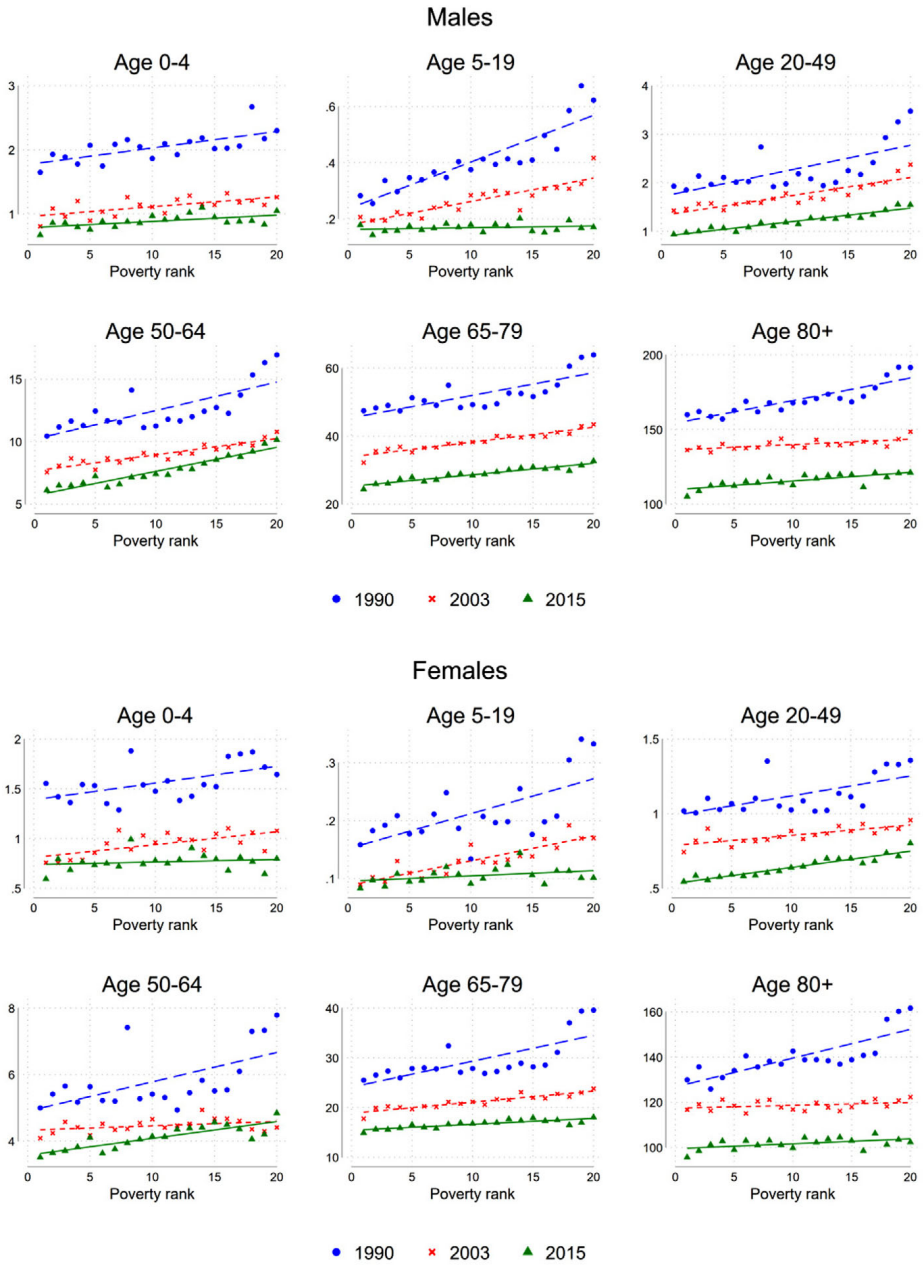


TABLE 1

Age-specific one-year mortality in lowest and highest poverty bins and change in inequality

	Lowest poverty bin			Highest poverty bin			Slope of regression line			p-value	
	1990	*2003*	*2015*	*1990*	*2003*	*2015*	*1990*	*2003*	*2015*	Δ_{1990}^{2003}	Δ_{2003}^{2015}
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*
Men											
0–4	1.651	0.809	0.672	2.301	1.263	1.048	0.026***	0.015***	0.010***	0.176	0.309
							(0.006)	(0.004)	(0.003)		
5–19	0.283	0.206	0.178	0.623	0.417	0.170	0.017***	0.008***	0.001	0.001	0.000
							(0.002)	(0.001)	(0.001)		
20–49	1.931	1.429	0.936	3.478	2.379	1.546	0.053***	0.039***	0.029**	0.374	0.077
							(0.014)	(0.005)	(0.002)		
50–64	10.452	7.556	6.112	16.951	10.795	10.140	0.228***	0.129***	0.193***	0.045	0.001
							(0.047)	(0.013)	(0.014)		
65–79	47.495	32.243	24.359	63.940	43.453	32.679	0.664***	0.435***	0.338***	0.084	0.068
							(0.122)	(0.034)	(0.027)		
80+	159.938	136.255	105.117	191.549	148.533	121.094	1.513***	0.359***	0.583***	0.000	0.130
							(0.187)	(0.093)	(0.110)		

(*Continued*)

TABLE 1
(Continued)

	Lowest poverty bin			Highest poverty bin			Slope of regression line			p-value	
	1990	*2003*	*2015*	*1990*	*2003*	*2015*	*1990*	*2003*	*2015*	Δ_{1990}^{2003}	Δ_{2003}^{2015}
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*
Women											
0–4	1.555	0.759	0.593	1.645	1.079	0.798	0.017**	0.013***	0.003	0.511	0.028
							(0.006)	(0.003)	(0.003)		
5–19	0.159	0.090	0.083	0.333	0.170	0.102	0.006***	0.004***	0.001	0.320	0.000
							(0.002)	(0.001)	(0.001)		
20–49	1.019	0.744	0.545	1.357	0.957	0.802	0.013***	0.007***	0.011***	0.065	0.033
							(0.004)	(0.001)	(0.001)		
50–64	4.998	4.088	3.510	7.791	4.412	4.842	0.088***	0.013*	0.051***	0.005	0.001
							(0.028)	(0.007)	(0.008)		
65–79	25.522	17.783	14.926	39.598	23.822	18.052	0.521***	0.222***	0.120***	0.019	0.002
							(0.116)	(0.022)	(0.018)		
80+	129.974	116.712	95.624	161.674	122.341	102.352	1.272***	0.127	0.217**	0.000	0.440
							(0.220)	(0.080)	(0.086)		

Note: Columns 1–6 report one-year mortality rates for each gender and age group in 1990, 2003 and 2015, in the bin of districts with the lowest and highest poverty rank. Rates for 2003 and 2015 are smoothed across three years. The rates are age-adjusted to the 2015 population. Columns 7–9 report the coefficient of the fitted regression line in each year. Columns 10 and 11 report the *p*-value for the null hypothesis that the slopes are equal in the given years. ***, ** and * denote significance at the 1, 5 and 10 per cent levels, respectively.

rates than other districts in 1990. From Figure 1, we know that these districts were all in the East. Hence, higher mortality rates in the poorest parts of the East were a significant driver of overall inequality in mortality in Germany. By 2003, these points are no longer such extreme outliers, suggesting that reductions in mortality in these particular districts helped to equalise mortality rates.

While inequality in mortality rates decreased across the board between 1990 and 2015, the picture is more nuanced for the subperiod 2003–15. During this subperiod, inequality decreased for children (girls 0–4, and both genders 5–19) but increased for the 50–64 age group. The results suggest that, for children, mortality rates declined more for poorer districts than for richer districts between 2003 and 2015, leading to a reduction in inequality. At the same time, there were either no improvements (for women aged 50–64) or smaller improvements (for women aged 20–49 and men aged 50–64) in mortality rates for poorer districts compared with richer districts for other age groups, leading to increases in inequality. For some age groups, mortality gradients remained constant (men aged 0–4 and 20–49, and men and women aged 80+; see Table 1).

III. Robustness and additional results

Part B

Figure 2 shows inequality in mortality rates obtained by re-ranking districts in each year using the average disposable income per capita for that year. Hence, the poverty rank is for a specific year. This re-ranking means that changes in mortality gradients over time could result from districts changing ranks, or from changes in mortality within districts with a fixed rank. In order to investigate what is behind these changes, we re-estimate our gradients using a fixed ranking based on the 2015 income data for all three years. The results are shown in Figure 3. Overall, the relationships look very similar whether based on re-ranking or using a fixed rank.

An interesting observation is that mortality rates for 1990 are more linear based on the 2015 ranking (Figure 3) compared with the re-ranked results in Figure 2. One reason is that Berlin, which is an outlier in terms of income in 1990 (high income compared with the East), had relatively high mortality rates that were similar to those of other Eastern districts. Based on the 2015 ranking, Berlin sticks out less. Another reason is that, as Figure 1 indicated, many poor regions in the East moved to a higher SES in relative terms over time. Thus, the poorest bins in 2015 consist of both Western and Eastern regions. As results in Section IV underline further, mortality was much higher in the East in 1990 across all age groups, so that combining mortality rates for Eastern and Western districts for 1990 (by using the fixed 2015 rankings) reduces extreme mortality rate outliers.

FIGURE 3

Yearly mortality per 1,000 (fixed 2015 poverty rank)

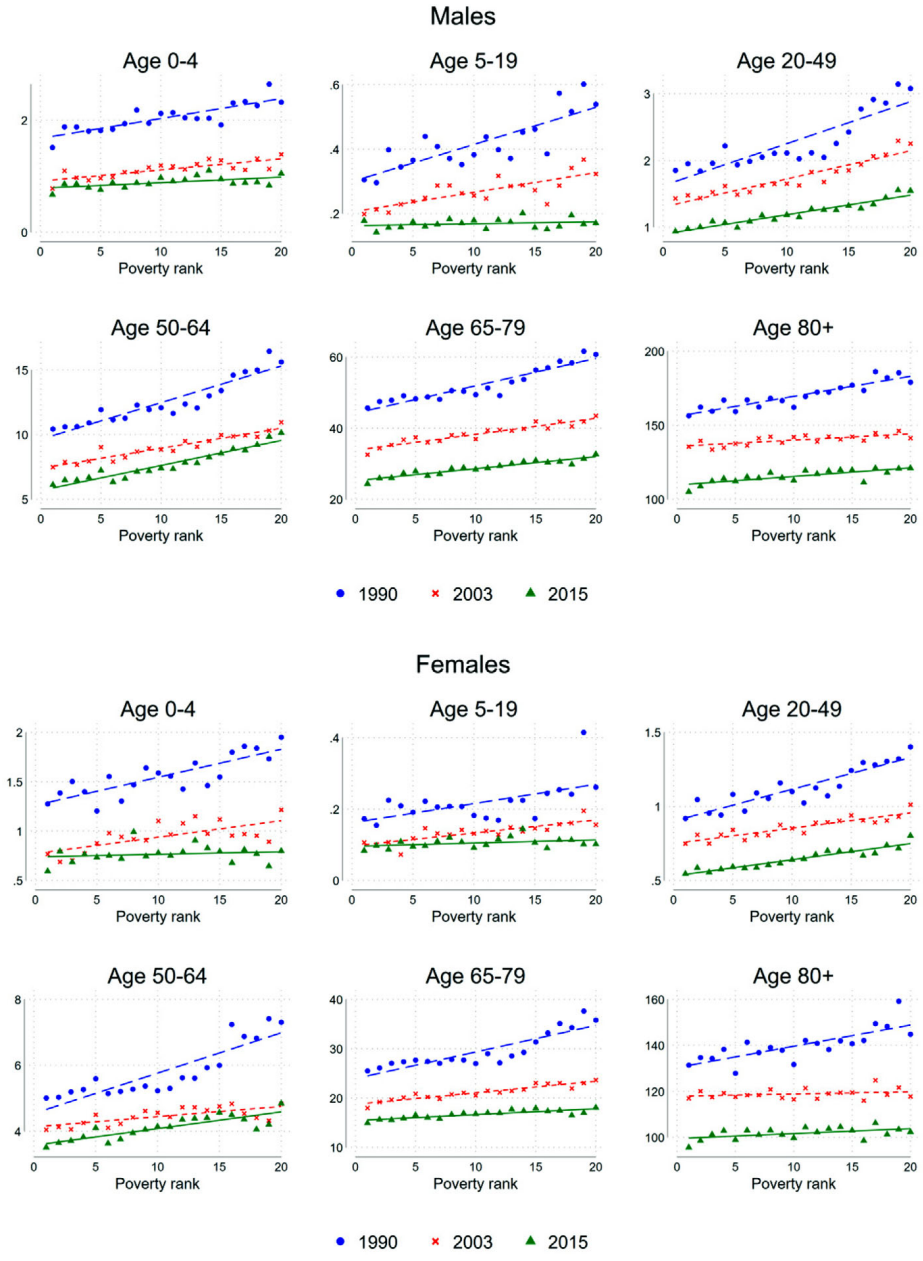


FIGURE 4

Mid-age yearly mortality per 1,000

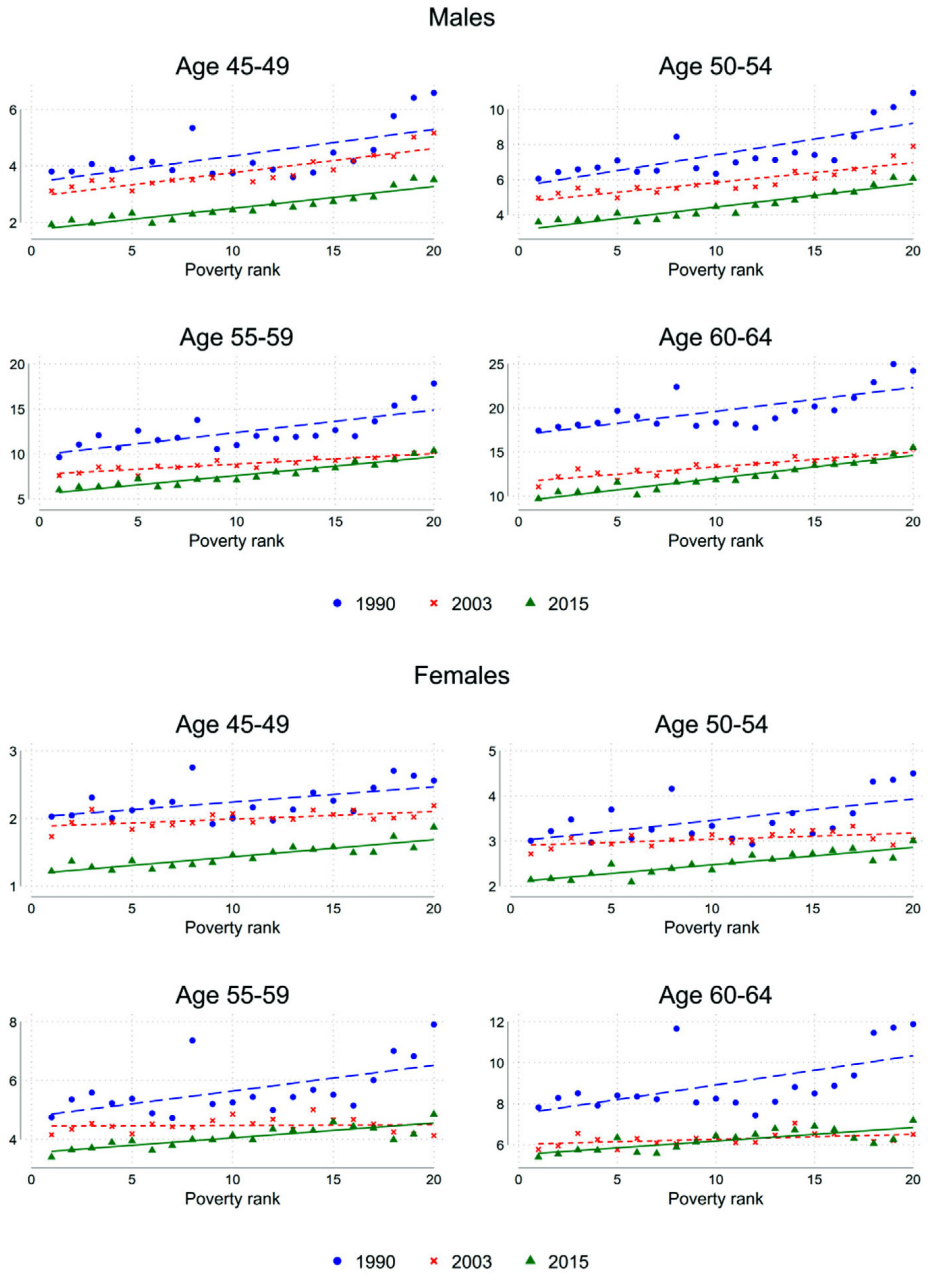


Figure A.1 displays the same results as Figure 2 without adjusting for age. The results in Figure 2 were age-adjusted to the 2015 population using the population counts in the finer five-year age groups discussed above. However, Figure A.1 suggests that age adjustment is not a major concern here – the general trends are similar in Figures 2 and A.1. The one difference is that in Figure A.1 the mortality gradient for young children in 1990 is almost flat. The underlying reason for the discrepancy in this case is the low number of births in the East in 1990 (because, within the 0–4-year-old group, infants have much higher mortality). This example shows that it is necessary to adjust for lower fertility in the low-SES areas (i.e. mostly East German districts) in 1990 in order to draw valid conclusions about the youngest age group.

Figure 4 explores mortality rates in middle age in more detail. It displays mortality rates for the smallest available age groups (45–49, 50–54, 55–59 and 60–64) and re-ranks districts over time, as in Figure 2. The trends by age group differ over time. Mortality rates fell for both genders and all age groups from 1990 to 2015. For the age groups between 55 and 64, this decrease was largest in the period 1990–2003, while for the age groups between 45 and 54, the decrease was largest from 2003 to 2015. Here, underlying cohort effects might be playing a role as the cohorts born during and shortly after World War II (i.e. ages 40–49 in 1990 or 55–64 in 2003) faced difficult conditions in early childhood.[9] However, this should not influence the SES gradient in mortality rates, which is trending similarly for men and women. There was no improvement in mortality rates from 2003 to 2015 for lower-SES districts in the age groups between 55 and 64. For women aged 60–64, we even see the linear fit lines for 2003 and 2015 crossing (albeit narrowly): lower SES groups are worse off while higher SES groups are improving. In contrast, mortality rates have been decreasing for age groups between 45 and 54 across the board.

IV. Discussion: East–West differences and trends

1. Background and previous literature

Differences between the former Communist GDR and the West are widely studied in economics and other fields. While GDP per capita had been consistently higher in the West since the 1950s, life expectancy only started diverging in the 1970s.[10] The East caught up in life expectancy rather quickly after reunification. In this section, we explore these differences further by examining age-specific mortality rates over time.

Various studies in the fields of epidemiology, sociology and demography have analysed mortality differences in Germany. Van Raalte et al. (2020) find

[9]Kesternich et al., 2014.
[10]Becker, Mergele and Woessmann, 2020.

that regional inequality in mortality, particularly among women, is low for Germany in an international comparison in absolute terms. Von Gaudecker and Scholz (2007) compare male pensioners' life expectancy at age 65 by lifetime earnings and find a positive association between income and longevity for both East and West German men. Kibele, Jasilionis and Shkolnikov (2013) also analyse male pensioners and find an increase in the income–mortality gradient from the 1990s to the 2000s that was more pronounced in East Germany. Rau and Schmertmann (2020) estimate life expectancy in 2015–17 at the district level. They find that life expectancy correlates with various local economic indicators while it does not correlate strongly with physician density and population density.

2. Health care in the East

The health care system in the former GDR was centralised and mostly state-operated. It was split up into ambulatory and hospital services. Up until the 1960s, it was considered a model health care system within the socialist countries. Since the 1970s, shortages of equipment and medical personnel had played a larger role.[11] While this was not disclosed to the public, internal documents later showed the lack of equipment, hospital beds and personnel, as well as the bad shape of the hospitals in the 1970s and 1980s. Medical technology in the 1980s was described as '10 years behind Western technology' in internal documents in the 1980s.[12]

As summarised by Busse and Blümel (2014), after the reunification in 1990, the East quickly and fully adopted the health care system of the West. An aid package of several billion euros directed at Eastern hospitals and nursing homes helped to modernise health care facilities in the East. Over time, the densities of hospital beds and medical personnel in the East have risen to levels that are comparable with the West. The net migration of 1.45 million people from the East to the West also played a role in this adaptation process. Today, there are marginal differences in health care provision between the East and the West in general. Differences are much more pronounced between rural and urban areas.

3. Trends within the East and the West by age group

Figure 5 explores differences in the evolution of inequality in mortality between the East and the West. Districts are binned according to their disposable income per capita (in prices of 2015). West German districts are grouped into 20 bins while former GDR districts (excluding Berlin) make up

[11]Busse and Blümel, 2014.
[12]Erices and Gumz, 2014.

FIGURE 5A

East versus West: trends in income and mortality rates, age 0–4

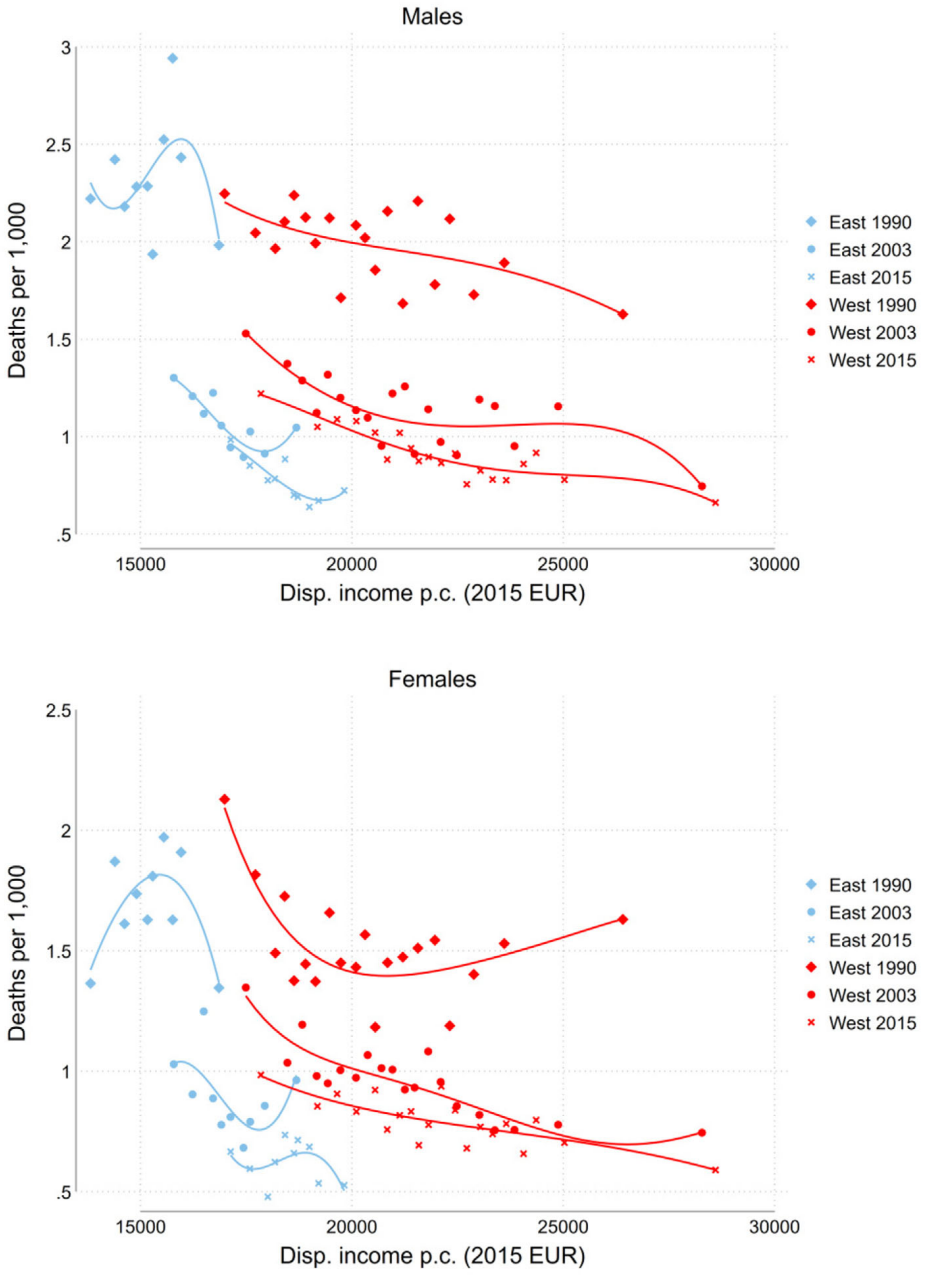


FIGURE 5B

East versus West: trends in income and mortality rates, age 5–19

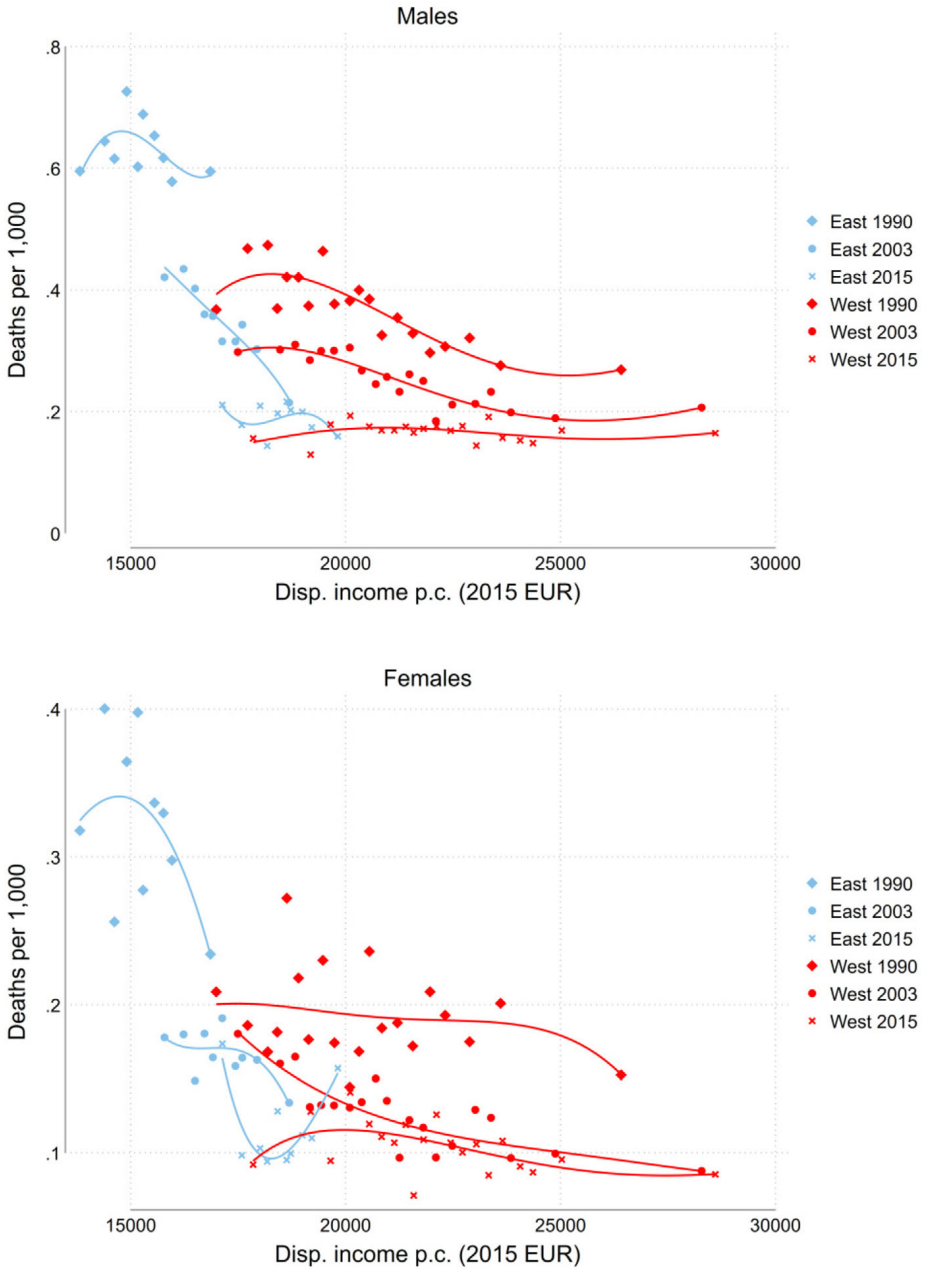


FIGURE 5C

East versus West: trends in income and mortality rates, age 20–49

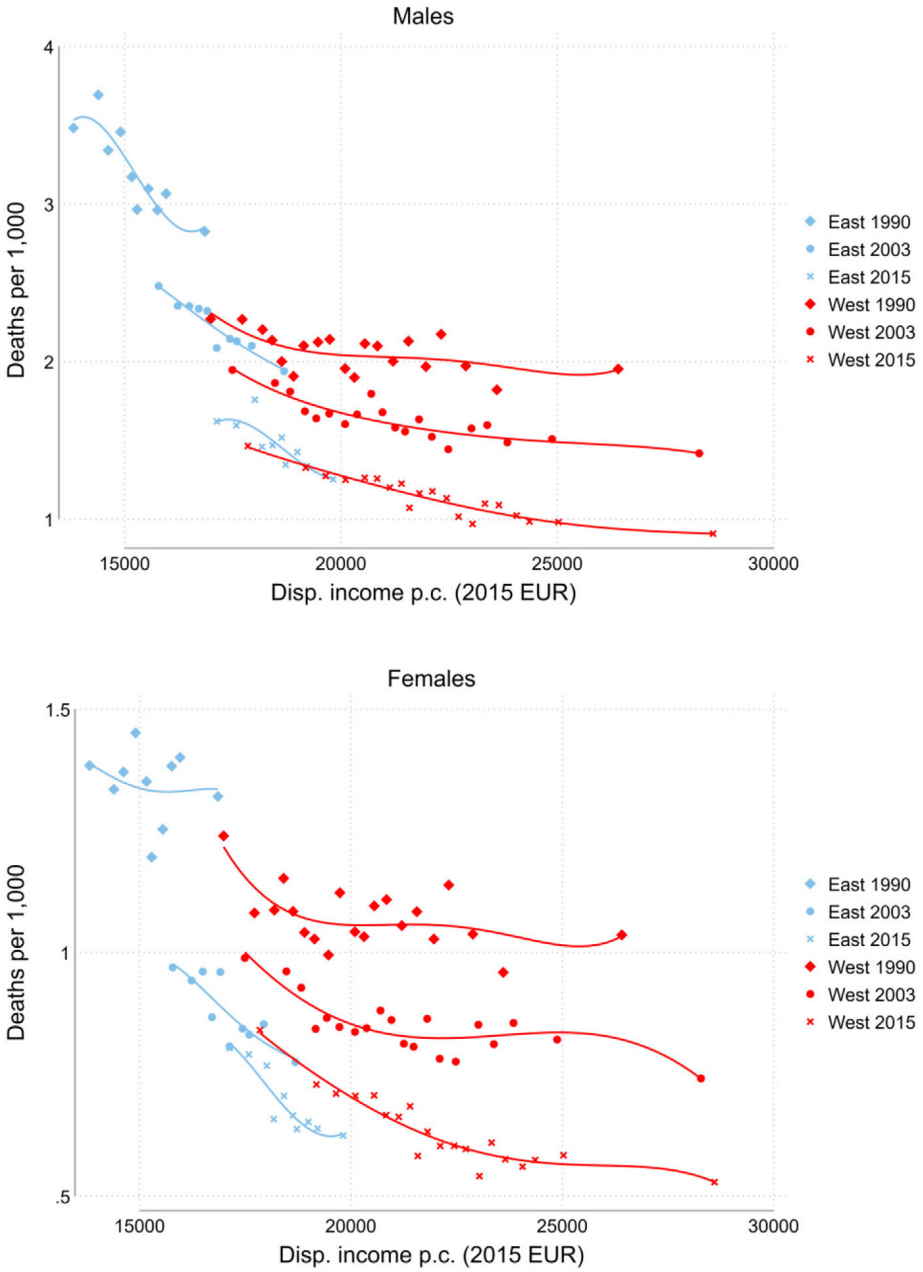


FIGURE 5D

East versus West: trends in income and mortality rates, age 50–64

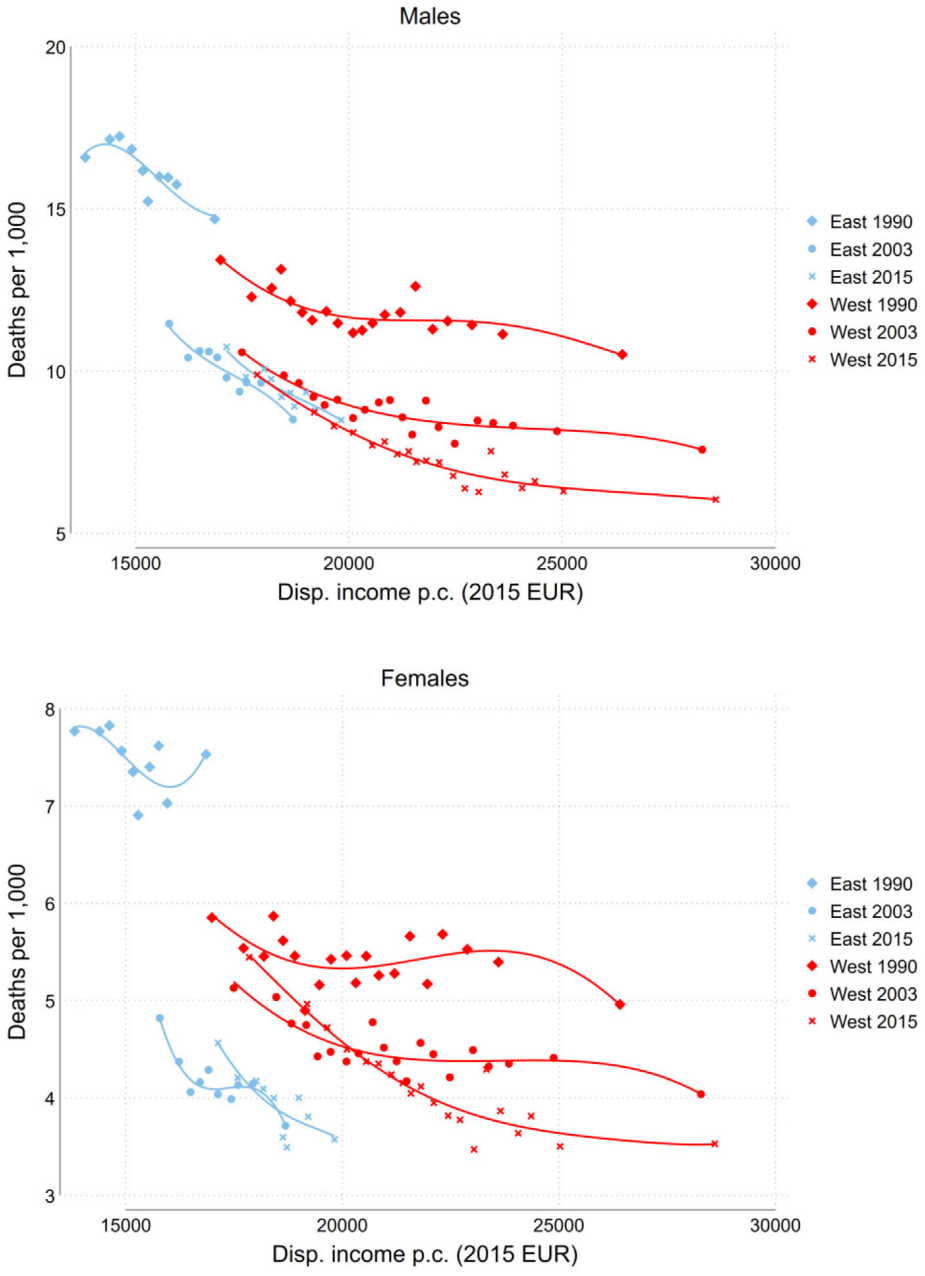


FIGURE 5E

East versus West: trends in income and mortality rates, age 65–79

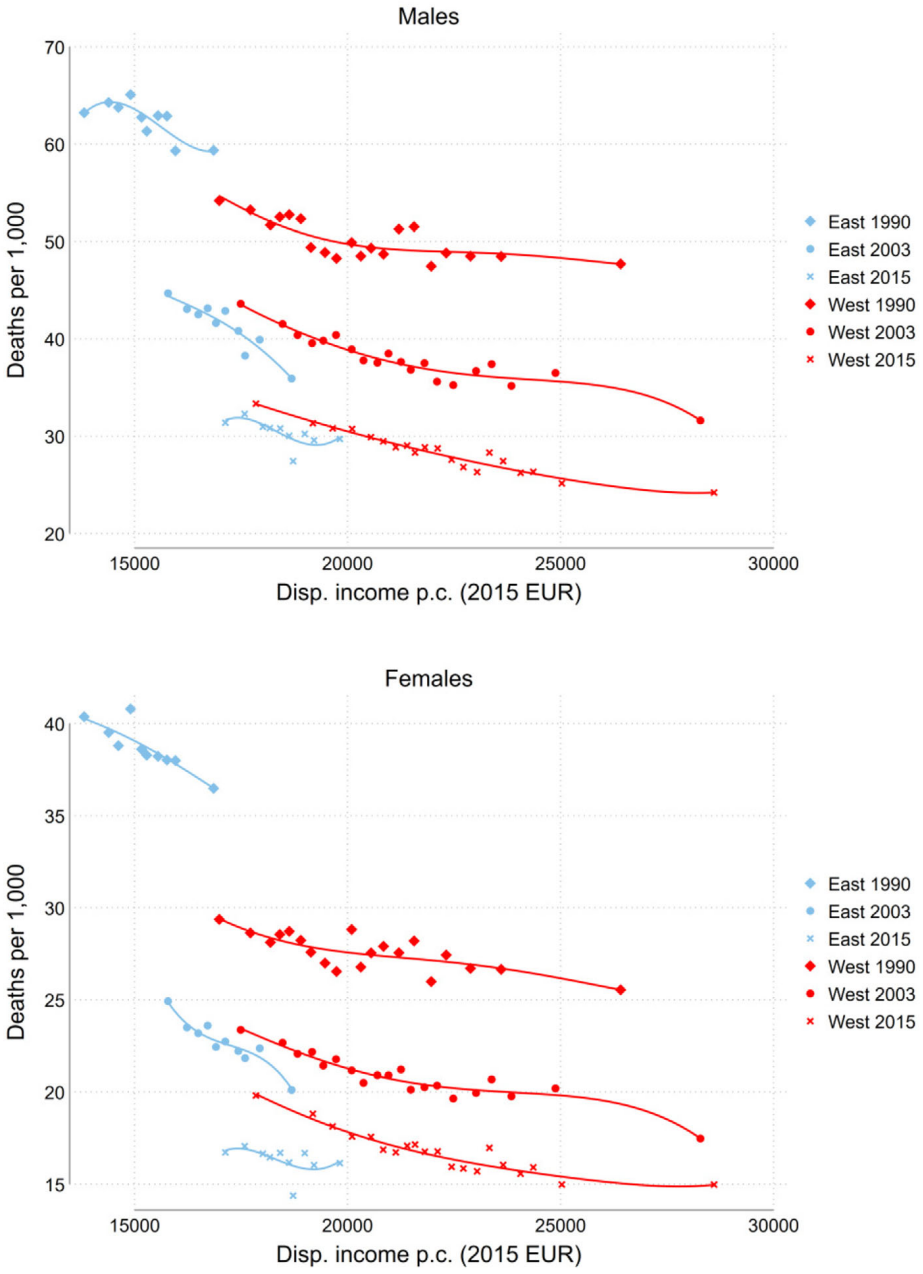
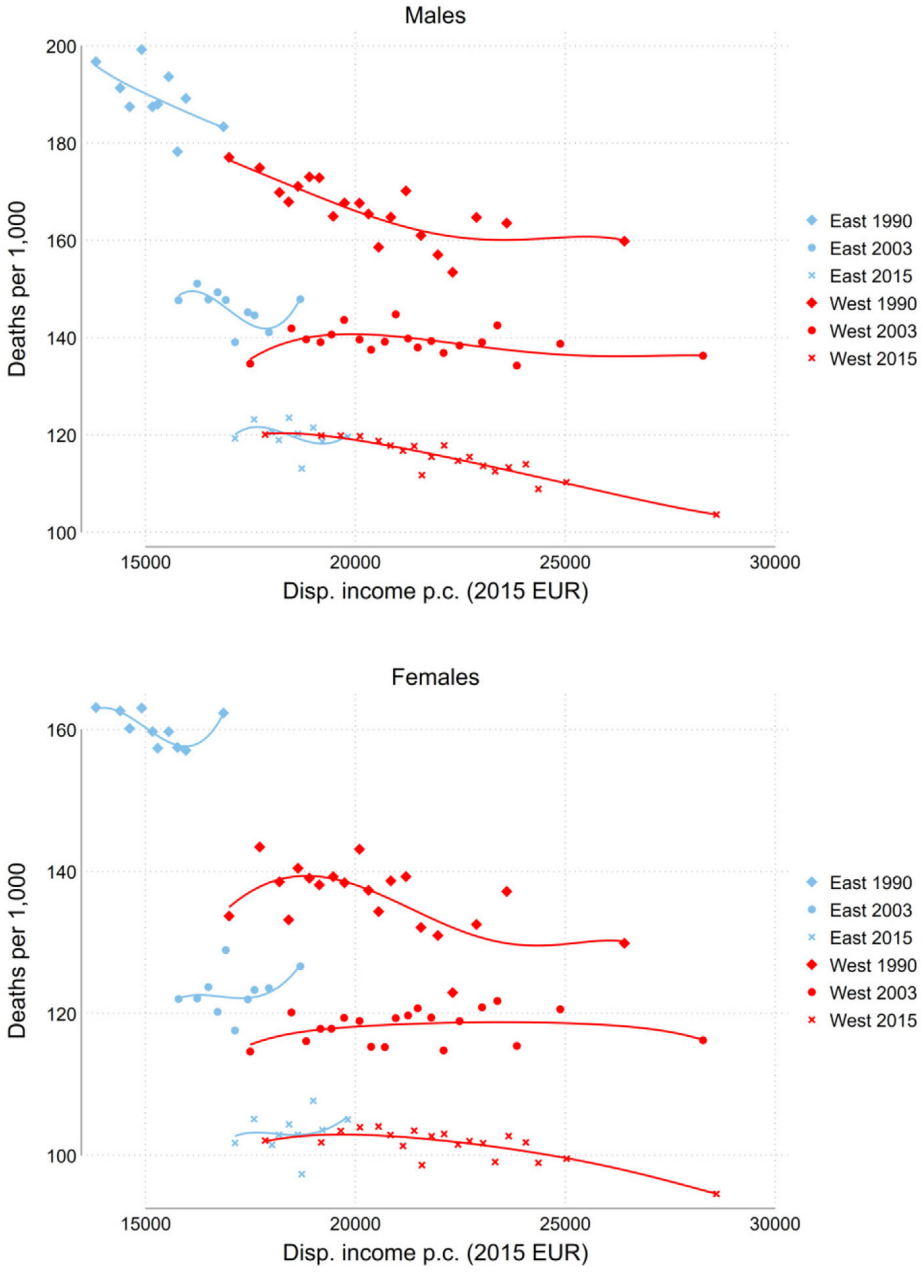


FIGURE 5F

East versus West: trends in income and mortality rates, age 80+



10 bins. Over time, movement to the right indicates improvements in income while a downward shift represents a lower mortality rate. Trend fit lines are polynomic to account for possibly non-linear trends.

Figure 5 indicates that there was no overlap in 1995 income between the East and the West (our SES measure for 1990). The richest decile in the East had lower disposable income on average than the poorest ventile in the West. This is not the case in 2003 and 2015 where we see some convergence in income. However, strong inequality persists as the highest decile in the East is still only at the level of the 16th ventile of the West in 2015. Income inequality within the West and the East also shows differing trends. While the range of incomes in the East is decreasing over time, the top and bottom income bins in the West diverge as a sign of widening regional economic inequality.

The most striking thing about Figure 5 may be that in many age and gender groups, mortality rates in the East and the West do not appear to be on the same line in 1990. That is, in 1990 the higher mortality in the East could not be explained simply by the fact that these districts had lower income. The discontinuity in the relationship between income and mortality is seen most strongly for men aged 5–49, and for women aged over 50. Looking over time, across all age groups, the reduction of inequality in mortality in Germany since 1990 is most strongly driven by improvements in mortality rates in Eastern districts. By 2015, mortality rates in the East are largely what one would expect given their lower income levels.

Kibele, Klüsener and Scholz (2015) argue that excess mortality in the East after reunification was attributable to traffic-related mortality among younger men and to excess cardiovascular mortality among people aged 60 and older; deaths from both of these causes were greatly reduced during the 1990s.

For young children (Figure 5A), the East–West gap was low in 1990, with slightly lower mortality rates in the West. As discussed above, mortality rates in this age group are mainly driven by infant mortality, which was 7.1 deaths per 1,000 live births in the West and 7.3 deaths per 1,000 live births in the East in 1990, according to the German Federal Institute for Population Research. In the late 1990s, the infant mortality rate in the East dropped below the rate in the West and remains lower in the East today (2.5 in the East compared to 3.3 in the West in 2019). This is especially remarkable as the SES disadvantage persists. In this age group, mortality has also dropped most in relative terms. It is possible that the difference in infant mortality rates may reflect a difference in which women select into child birth.

For those aged 5–19, Figure 5B shows the flattening of the SES mortality gradient in more detail. Because of the low number of deaths, the graphs are relatively noisy. By 2015, the West and the East had similar mortality rates and there was almost no gradient with respect to disposable income, which represented a remarkable improvement since 1990, especially in the East.

For adults aged 20–49, the mortality disadvantage in the East was particularly pronounced for men in 1990, while for women, the biggest mortality disadvantage in the East opened up after age 50. These large gaps had largely closed by 2003 for men in both age groups whereas by 2003 women aged 50–79 in the East have lower mortality rates than their Western counterparts with similar income. In 2015, the picture is similar: men in the East die at comparable rates to similar income groups in the West while women in the East have a lower mortality rate conditional on income category. Among men aged 65–79, the clear 1990 East–West mortality rate gap had mostly closed by 2003 and fully closed by 2015.

Figure A.2 compares the development of SES gradients in the West and the East from 1990 to 2015 for more years (in five-year intervals). Here, districts in the West and in the East are grouped into five bins each. One-year mortality rates are shown. As districts are ranked within East and West, we can compare, for example, mortality in the poorest places in the West with mortality in the poorest places in the East. Conducting this type of comparison reveals a puzzle – although mortality rates for men are largely similar conditional on poverty rank, women aged 5–19 (Panel B2) and 50–64 (Panel D2) in the poorest places in the West have higher mortality than those in the poorest places in the East. Hence, pockets of high poverty and high mortality appear to be emerging in the West. Kibele et al. (2015) also observe that many economically disadvantaged regions in the West have become high-mortality hot spots.

A potential limitation of our analysis is that there is undoubtedly a good deal of inequality in income and mortality rates within districts as well as across districts. Rau and Schmertmann (2020) find that deprivation indicators correlate most strongly with life expectancy at the district level and more so than GDP per capita. This suggests that individual-level associations between income and mortality rates, which have been demonstrated for subgroups,[13] are likely to be larger in magnitude than our district-level gradients.

V. Summary and conclusions

Income-related inequality in mortality rates across German districts has decreased considerably since 1990, for all age groups and both genders. This decrease is mainly driven by the fact that the East has caught up. The East remains considerably poorer, on average, than the West 25 years after reunification, but the large gap in disposable income that existed five years after reunification has closed and there are now equally poor districts in the East and West.

[13] See, for example, Von Gaudecker and Scholz (2007).

In contrast to the slow progress in terms of closing the income gap, the East has largely caught up and even overtaken the West in terms of reducing mortality. In 2015, mortality rates for many demographic groups (particularly for infants, but also for women aged 20–79) are below those of comparably poor districts in the West. For men, mortality rates conditional on district income are now quite similar in East and West.

Our results align with findings in the literature[14] that life expectancy does not differ significantly between the East and West, while GDP in the East is still significantly lower than in the West. Our results provide further detail about where the catch-up in life expectancy is coming from: the main driver for males is lower mortality rates among children aged 0–4 in the East compared with similarly poor districts in the West. For females, mortality in the East is also lower for age groups 20–49, 50–64 and 65–79.

These results speak to the larger literature about the relationship between inequality in income and inequalities in health. Germany offers a case study in which gaps in mortality were closed and even reversed, even while gaps in income proved stubbornly resistant to change. Understanding the drivers of this tremendous progress in the East, coupled with the reasons for the emergence of mortality hot spots in the West, is an important topic for future research.

Article Funding

Open access funding enabled and organized by Projekt DEAL.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

• Appendix

References

Becker, S. O., Mergele, L. and Woessmann, L. (2020), 'The separation and reunification of Germany: rethinking a natural experiment interpretation of the enduring effects of communism', *Journal of Economic Perspectives*, 34(2), 143–71.

Biewen, M., Ungerer, M. and Löffler, M. (2017), 'Why did income inequality in Germany not increase further after 2005?', *German Economic Review*, 20, 471–504.

Busse, R. and Blümel, M. (2014), 'Germany: health system review', *Health Systems in Transition*, 16(2), 1–296.

[14]Becker, Mergele and Woessmann, 2020.

Currie, J. and Schwandt, H. (2016a), 'Inequality in mortality decreased among the young while increasing for older adults, 1990–2010', *Science*, 352, 708–12.
— and — (2016b), 'Mortality inequality: the good news from a county-level approach', *Journal of Economic Perspectives*, 30(2), 29–52.
Dustmann, C., Ludsteck, J. and Schönberg, U. (2009), 'Revisiting the German wage structure', *Quarterly Journal of Economics*, 124, 843–81.
Erices, R. and Gumz, A. (2014), 'Health-care services in the GDR during the 1980s: a status report based on the files of the state security agency', *Gesundheitswesen*, 76, 73–8.
Fuchs-Schündeln, N., Krueger, D. and Sommer, M. (2010), 'Inequality trends for Germany in the last two decades: a tale of two countries', *Review of Economic Dynamics*, 13, 103–32.
Kesternich, I., Siflinger, B., Smith, J. P. and Winter, J. (2014), 'The effects of World War II on economic and health outcomes across Europe', *Review of Economics and Statistics*, 96, 103–18.
Kibele, E., Jasilionis, D. and Shkolnikov, V. M. (2013), 'Widening socioeconomic differences in mortality among men aged 65 years and older in Germany', *Journal of Epidemiology and Community Health*, 67, 453–7.
—, Klüsener, S. and Scholz, R. (2015), 'Regional mortality disparities in Germany: long-term dynamics and possible determinants', *Kölner Zeitschrift für Soziologie*, 67, 241–70.
Klüsener, S., Grigoriev, P., Scholz, R. D. and Jdanov, D. A. (2018), 'Adjusting intercensal population estimates for Germany 1987–2011: approaches and impact on demographic indicators', *Comparative Population Studies*, 43, 31–64.
Oei, P., Brauers, H. and Herpich, P. (2020), 'Lessons from Germany's hard coal mining phaseout: policies and transition from 1950 to 2018', *Climate Policy*, 20, 963–79.
Papanicolas, I., Woskie, L. R. and Jha, A. K. (2018), 'Health care spending in the United States and other high-income countries', *JAMA*, 319, 1024–39.
Rau, R. and Schmertmann, C. P. (2020), 'District-level life expectancy in Germany', *Deutsches Ärzteblatt International*, 117, 493–9.
Scholz, R. D. and Kreyenfeld, M. (2016), 'The register-based census in Germany: historical context and relevance for population research', *Comparative Population Studies*, 41, 175–204.
Van Raalte, A., Klüsener, S., Oksuzyan, A. and Grigoriev, P. (2020), 'Declining regional disparities in mortality in the context of persisting large inequalities in economic conditions: the case of Germany', *International Journal of Epidemiology*, 49, 486–96.
Von Gaudecker, H. M. and Scholz, R. D. (2007), 'Differential mortality by lifetime earnings in Germany', *Demographic Research*, 17, 83–108.

Appendix to Chapter 4

Geographic inequality in income and mortality in Germany

PETER REDLER, AMELIE WUPPERMANN, JOACHIM WINTER, HANNES SCHWANDT AND JANET CURRIE

Appendix

FIGURE A.1

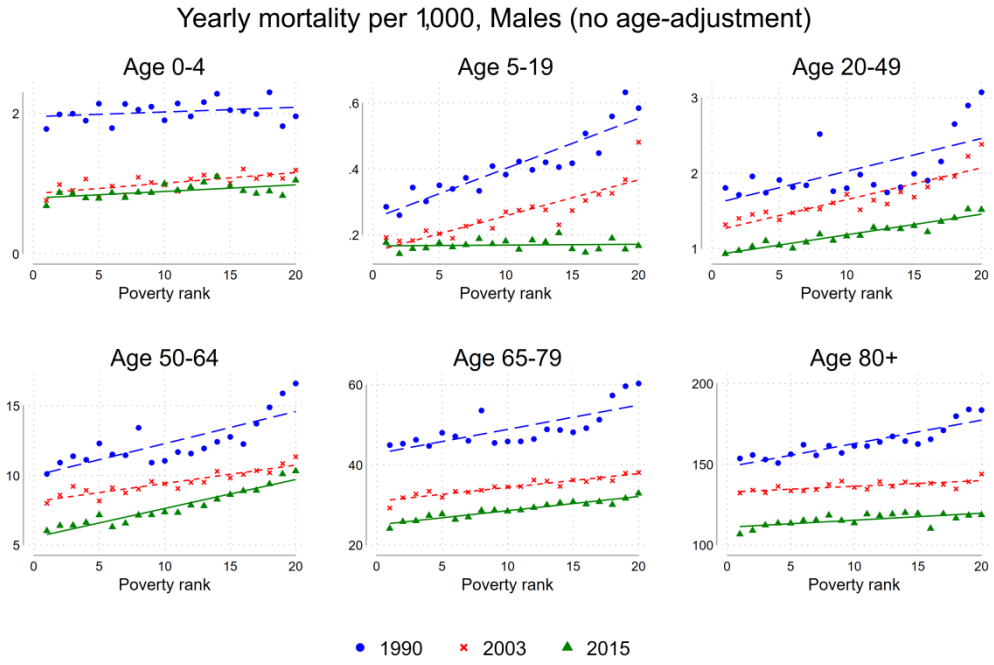


FIGURE A.1
(continued)

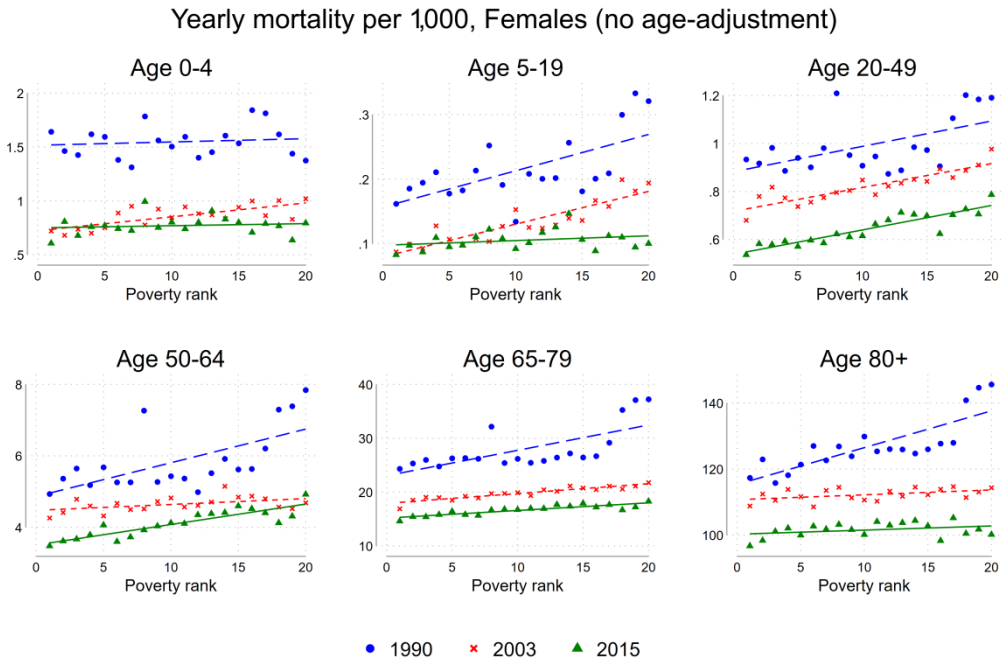


FIGURE A.2 (PANELS A)

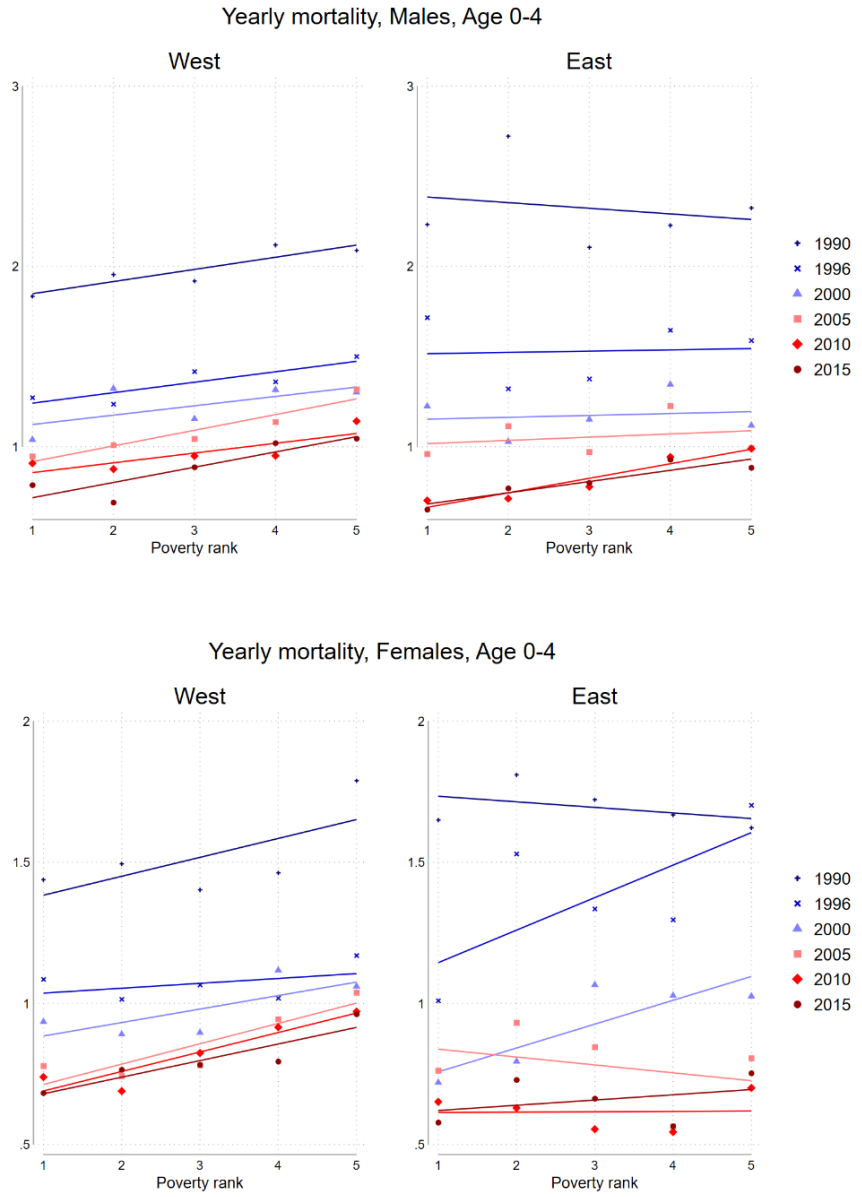


FIGURE A.2 (PANELS B)

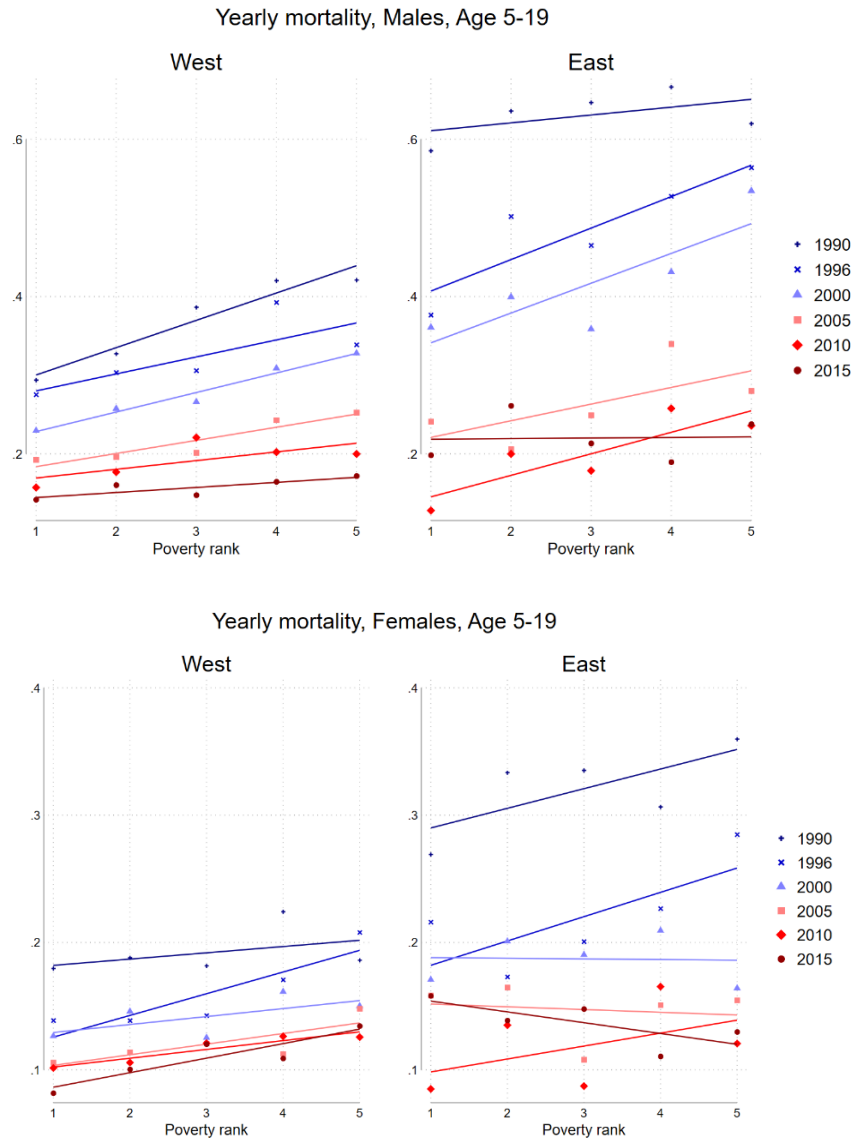


FIGURE A.2 (PANELS C)

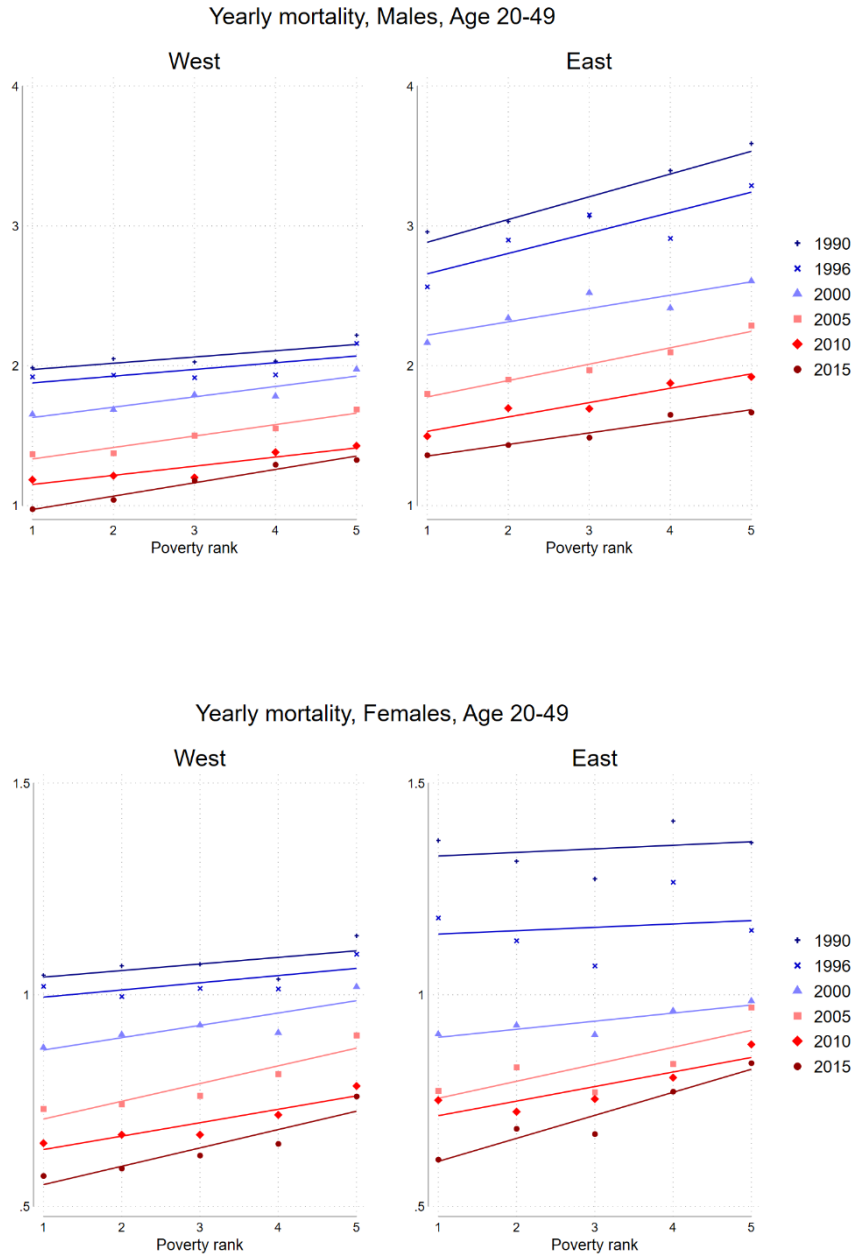


FIGURE A.2 (PANELS D)

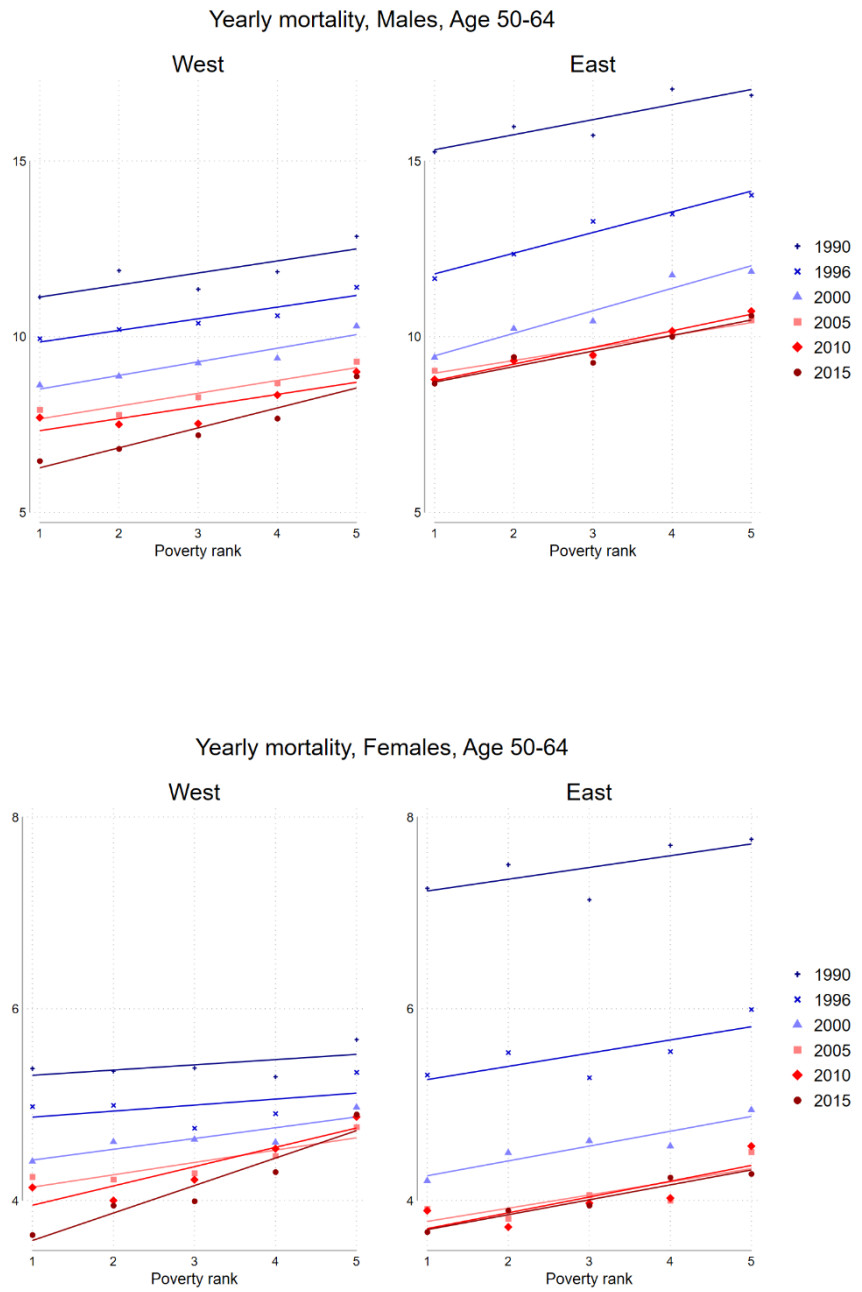


FIGURE A.2 (PANELS E)

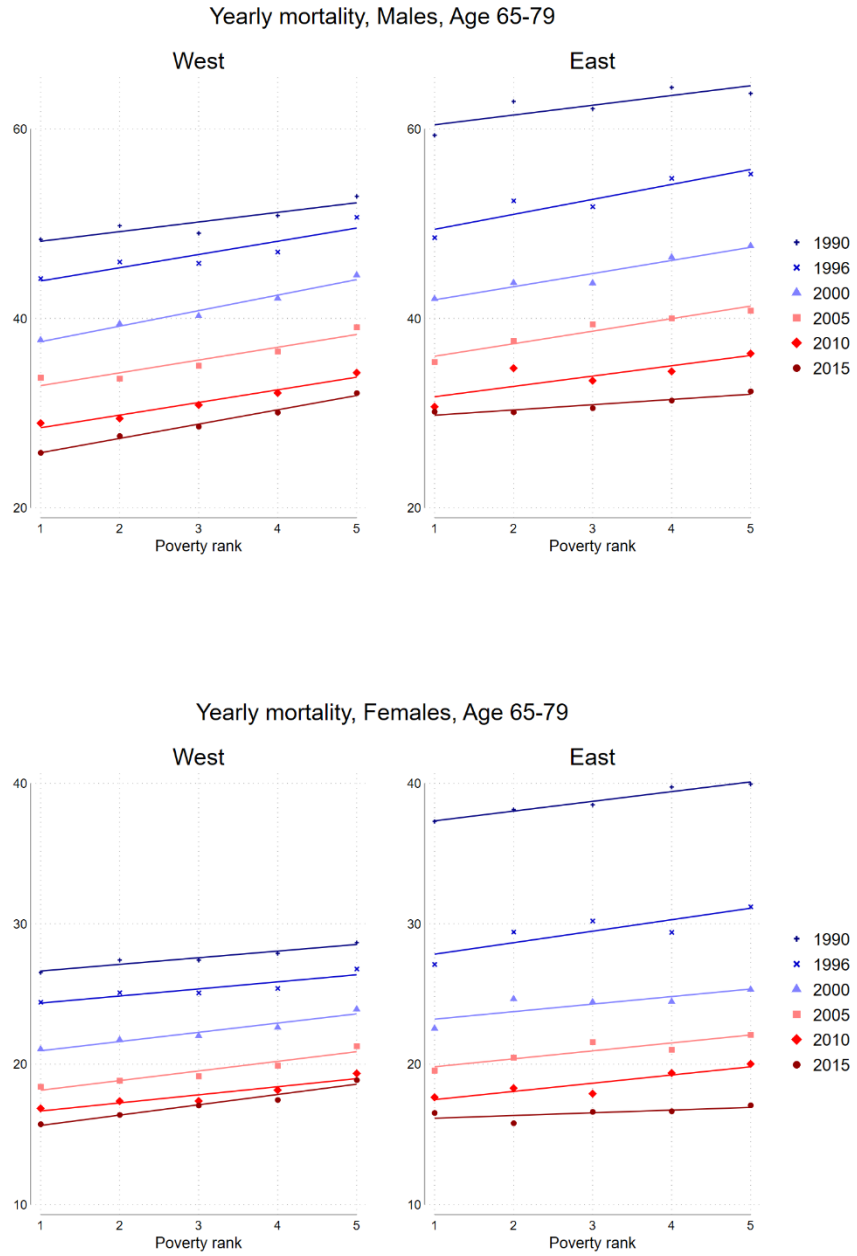
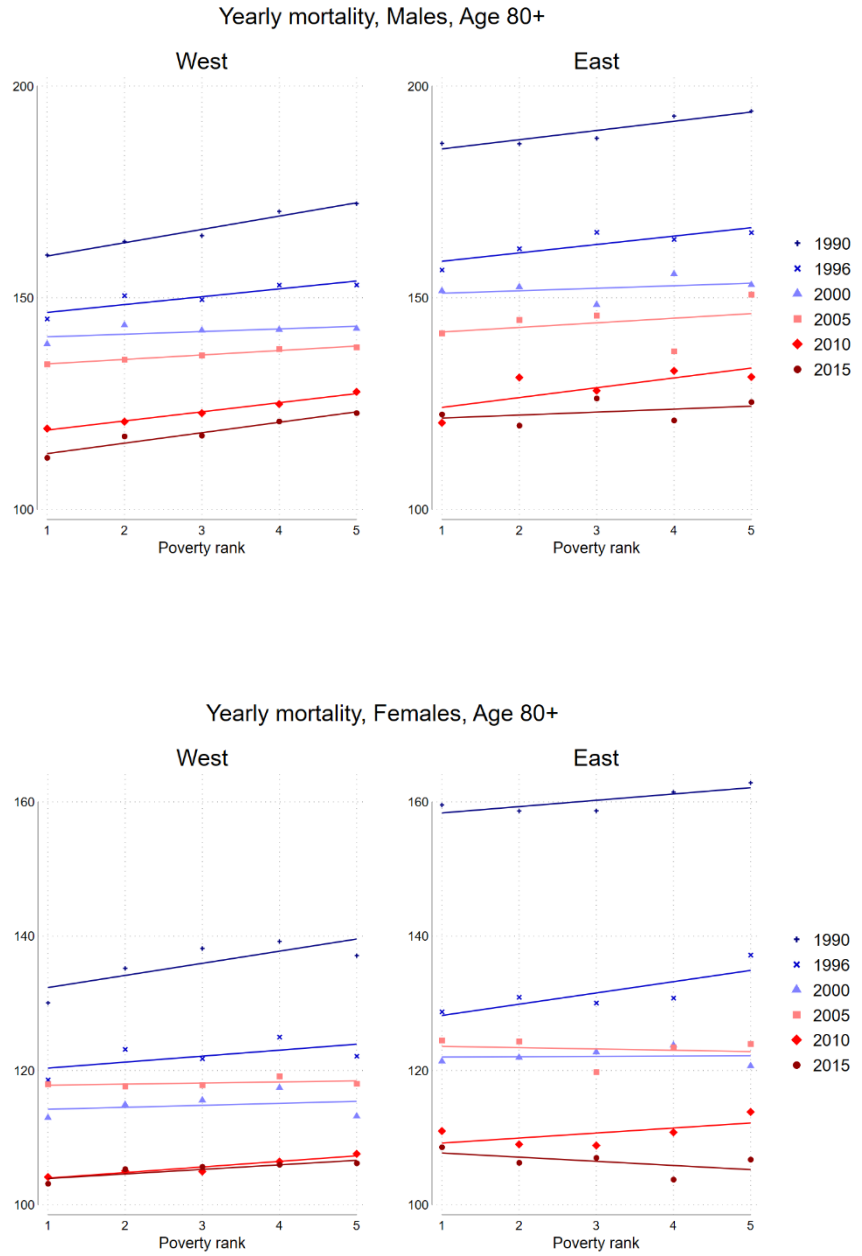


FIGURE A.2 (PANELS F)



Chapter 5

Inequality in mortality between Black and White Americans by age, place, and cause and in comparison to Europe, 1990 to 2018

This chapter is co-authored with Hannes Schwandt, Janet Currie, and twenty-four other co-authors. It was published in PNAS, Vol. 118, no. 40 (2021). I collected, processed, and analyzed the German mortality data used in the study. I contributed to the chapter's conceptualization, offered input during the review process, and also provided insights on the German context.

Inequality in mortality between Black and White Americans by age, place, and cause and in comparison to Europe, 1990 to 2018

Hannes Schwandt[a], Janet Currie[b,1], Marlies Bär[c], James Banks[d,e], Paola Bertoli[f], Aline Bütikofer[g], Sarah Cattan[d], Beatrice Zong-Ying Chao[a], Claudia Costa[h], Libertad González[i], Veronica Grembi[j], Kristiina Huttunen[k], René Karadakic[g], Lucy Kraftman[d], Sonya Krutikova[d], Stefano Lombardi[l], Peter Redler[m], Carlos Riumallo-Herl[n], Ana Rodríguez-González[o], Kjell G. Salvanes[g], Paula Santana[h], Josselin Thuilliez[p], Eddy van Doorslaer[n], Tom Van Ourti[n], Joachim K. Winter[m], Bram Wouterse[c], and Amelie Wuppermann[q]

[a]School of Education and Social Policy, Northwestern University, Evanston, IL 60208; [b]Department of Economics, Princeton University, Princeton, NJ 08540; [c]Erasmus School of Health Policy & Management, Erasmus University Rotterdam, 3000DR Rotterdam, The Netherlands; [d]Institute for Fiscal Studies, London WC1E 7AE, United Kingdom; [e]Department of Economics, University of Manchester, Manchester M13 9PL, United Kingdom; [f]Department of Economics, University of Verona, 37129 Verona, Italy; [g]Department of Economics, Norwegian School of Economics, Bergen, 5045, Norway; [h]Centre of Studies in Geography and Spatial Planning, University of Coimbra, 3004-531, Coimbra, Portugal; [i]Department of Economics and Business, Universitat Pompeu Fabra, 08005 Barcelona, Spain; [j]Department of Economics, Management, and Quantitative Methods, University of Milan, 20122 Milano, Italy; [k]Department of Economics, Aalto University School of Business, 02150 Espoo, Finland; [l]VATT Institute for Economic Research, 00100 Helsinki, Finland; [m]Department of Economics, University of Munich, 80539 Munich, Germany; [n]Erasmus School of Economics, Erasmus University Rotterdam, 3000DR Rotterdam, The Netherlands; [o]Department of Economics, Lund University, SE-220 07 Lund, Sweden; [p]CNRS, Centre d'économie de la Sorbonne, Université Paris 1, 75013 Paris, France; and [q]Department of Economics, University of Halle, 06108 Halle (Saale), Germany

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved August 15, 2021 (received for review March 18, 2021)

Although there is a large gap between Black and White American life expectancies, the gap fell 48.9% between 1990 and 2018, mainly due to mortality declines among Black Americans. We examine age-specific mortality trends and racial gaps in life expectancy in high- and low-income US areas and with reference to six European countries. Inequalities in life expectancy are starker in the United States than in Europe. In 1990, White Americans and Europeans in high-income areas had similar overall life expectancy, while life expectancy for White Americans in low-income areas was lower. However, since then, even high-income White Americans have lost ground relative to Europeans. Meanwhile, the gap in life expectancy between Black Americans and Europeans decreased by 8.3%. Black American life expectancy increased more than White American life expectancy in all US areas, but improvements in lower-income areas had the greatest impact on the racial life expectancy gap. The causes that contributed the most to Black Americans' mortality reductions included cancer, homicide, HIV, and causes originating in the fetal or infant period. Life expectancy for both Black and White Americans plateaued or slightly declined after 2012, but this stalling was most evident among Black Americans even prior to the COVID-19 pandemic. If improvements had continued at the 1990 to 2012 rate, the racial gap in life expectancy would have closed by 2036. European life expectancy also stalled after 2014. Still, the comparison with Europe suggests that mortality rates of both Black and White Americans could fall much further across all ages and in both high-income and low-income areas.

life expectancy | racial divide | area-level socioeconomic status | international comparison | age-specific mortality

Recent events, notably the Black Lives Matter movement and the disproportionate impact of the COVID-19 pandemic on the Black population, have highlighted the persistent gap in life expectancy between Black Americans and other Americans (1, 2). In 2018, the gap in life expectancy between Black and White Americans was 3.6 y. However, there have also been tremendous improvements in life expectancy among Black Americans relative to White Americans over time and especially since 1990 (3–7). Much of the highly publicized recent research investigating changes in inequality in life expectancy and mortality in the United States over the past 30 y highlights inequalities in adult mortality across educational and income groups (8–23).

This paper discusses the evolution of inequalities in mortality between Black and White Americans from 1990 to 2018 through the lens of place. There are two innovations: First, following several recent studies (1, 6, 24–31), we examine the evolution of mortality rates among Black and White Americans by age and county poverty rates. This analysis allows us to see whether racial gaps have evolved differently in higher- and lower-income parts of the United States. Trends in age-specific mortality rates provide insights into whether changes in life expectancy are specific to certain age groups, for example, people over 65 who qualify for Medicare, which in turn may provide additional insight into possible mechanisms.

Second, we benchmark these developments against trends in mortality inequality across high- and low-income places in a set

Significance

From 1990 to 2018, the Black–White American life expectancy gap fell 48.9% and mortality inequality decreased, although progress stalled after 2012 as life expectancy plateaued. Had improvements continued at the 1990 to 2012 rate, the racial gap in life expectancy would have closed by 2036. Despite decreasing mortality inequality, income-based life expectancy gaps remain starker in the United States than in European countries. At the same time, European mortality improved strongly and even those U.S. populations with the longest life spans–White Americans living in the highest-income areas–experience higher mortality at all ages than Europeans in high-income areas in 2018. Hence, mortality rates of both Black and White Americans could fall much further in both high-income and low-income areas.

Author contributions: H.S., J.C., J.B., S.C., S.K., and K.G.S. conceived the overall comparative project; H.S. and J.C. designed research; H.S., J.C., and B.Z.-Y.C. performed research; M.B., P.B., A.B., S.C., C.C., L.G., V.G., K.H., R.K., L.K., S.L., P.R., C.R.-H., A.R.-G., P.S., J.T., E.v.D., T.V.O., J.K.W., B.W., and A.W. provided European mortality data in a format similar to the US data; and H.S. and J.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: jcurrie@princeton.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2104684118/-/DCSupplemental.

Published September 28, 2021.

of six prosperous European countries. This comparison offers several potential insights, such as whether mortality in higher-income parts of the United States is more similar to that of European countries or whether both high- and low-income US places tend to lag behind. It also provides additional perspective on the gaps between Black and White Americans, allowing us to ask, for example, if only Black Americans fall short of a European benchmark or if the mechanisms driving lower life expectancy in the United States also affect White Americans.

Our analysis proceeds by first ranking counties by their poverty rate in each year and then grouping counties into clusters that each account for about 5% of the population (*Materials and Methods*). A key advantage of this approach is that we can examine all deaths, whereas information on income and completed education is not available for every person. We also avoid problems due to changes in the education distribution over time. For example, high school dropouts in the United States have become an increasingly small and more negatively selected group as high school completion and college attendance have become more normative (32–35). A limitation of our approach is that we cannot examine inequalities within small areas. Set against this limitation is evidence that low-income Americans live longer in high-income areas than in low-income ones (12), so that mortality across geographic areas is of independent interest. Our approach allows us to see whether changes in mortality occur in both high- and low-income areas or are driven largely by improvements in lower-income areas.

A second advantage of this geographical approach is that it can be easily adapted to examine mortality inequality in other countries using a similar framework. We examine trends in mortality inequality in six wealthy European countries using methods identical to the US analysis. Mortality inequality in these countries is of interest in its own right but also serves as a useful baseline for considering developments in the United States, contributing to a growing body of comparative literature on mortality differentials (6–39). Our main analysis focuses on six countries (England, France, Germany, the Netherlands, Norway, and Spain) for which consistent mortality data by geographic areas exist for the entire time period. All six are prosperous countries with well-developed health care and social welfare systems. The experience of these countries provides some insight into questions such as how low US mortality rates could fall given current medical standards; whether increasing gaps in life expectancy between the United States and Europe are driven only by lower-income areas or whether higher-income areas are also falling behind; and finally, whether mortality among Black Americans declined only relative to White Americans or whether it also declined relative to a European life expectancy benchmark.

Results

Our main results for age-specific mortality rates are shown in Figs. 1–4, representing the age groups 0 to 4, 5 to 19, 20 to 64, and 65 to 79. Each figure has three panels showing estimates for the years 1990, 2005, and 2018 and contains three heavy lines representing mortality rates for Black Americans, White Americans, and Europeans as defined in our study. Each marker on these figures represents a county group representing about 5% of a country's population. The lines drawn through the markers are simply least squares linear regression lines through the points. The data for each marker and the slopes of the regression lines are shown in *SI Appendix*, Tables S1–S4, along with *P* values for whether the slopes of the regression lines are equal to 0, whether the slopes for lines representing Black and White persons are equal, and whether the slopes of the White American and European lines are equal. The figures also show fainter gray lines corresponding to mortality rates in the individual European countries. Further details are provided in *Materials and Methods* and in *SI Appendix*.

The biggest takeaway from Fig. 1 is the huge gap in mortality between Black and White children aged 0 to 4 in 1990 and the equally stunning narrowing of the gap between Black and White child deaths over the subsequent decades. In 1990, 4.2 out of every 1,000 Black children aged 0 to 4 died compared to 1.82 per 1,000 White children. In 2018, the rates had fallen to 2.31 per 1,000 for young Black children and 1.13 per 1,000 for White children. Viewed as a percentage, the progress is less impressive—in 1990, 2.3 Black children died for every White child, while in 2018, 2.04 Black children died for every White child. However, the increase in the total number of lives saved was much greater for young Black children, resulting in much closer absolute mortality rates in 2018. Much of the improvement in Black child mortality rates happened between 1990 and 2005, with only slow progress from 2005 to 2018. Mortality rates for White children aged 0 to 4 also fell throughout the period, although at a slower rate.

Mortality improvements among young Black children occurred across the entire economic spectrum of US locations, although reductions were somewhat stronger in the highest-income areas, which led to an increase in mortality inequality for Black American children. Mortality inequality among White children decreased slightly. Hence, the strong reduction in overall inequality in mortality for young children aged 0 to 4 in high- and low-income areas that has been previously reported (6, 25–27) is due to the higher concentration of Black children in low-income areas combined with the large reduction in mortality rates among young Black children. Overall, despite strong improvements, mortality among young Black children remained substantially higher and more unequally distributed between high- and low-income places compared to White children.

Inequality in mortality among young children aged 0 to 4 in Europe was lower than in the United States in all years, and the European gradient between mortality rates and area poverty was almost entirely flat in 2018 (see *SI Appendix*, Table S1 for numerical values of the gradients). Fig. 1 shows that in 1990, the average US White mortality rate for the 0 to 4 age group was similar to the European rate, although deaths were more unequally distributed in the United States. Specifically, US death rates among White children aged 0 to 4 were lower than European rates in the highest-income areas and higher in the lowest-income areas in 1990. By 2005, mortality for White children aged 0 to 4 had pulled away from European levels and was uniformly higher than in Europe in both high- and low-income areas. This trend continued to 2018.

Fig. 2 tells a broadly similar story for children aged 5 to 19. The biggest difference is that even in 1990, White mortality rates for children aged 5 to 19 were higher than European rates in all but the highest-income US places. By 2005, the gap between European children and White American children had become wider than the gap between Black and White American children, which suggests that all American children in this age group suffered high levels of mortality relative to the lower potential mortality rates implied by the European baseline.

Fig. 3 shows trends in mortality for adults aged 20 to 64. (See *SI Appendix* for a further split into ages 20 to 49 and 50 to 64.) Focusing first on the US story, the three panels show that Black and White prime-age adult mortality (aged 20 to 64) converged sharply over time driven primarily by a rapid fall in Black mortality, especially in the lowest-income areas. In the highest-income areas of the United States, the gap in Black–White American death rates had fallen to 0.7 deaths per 1,000 by 2018, while in lower-income places it was still 1.47 (*SI Appendix*, Table S3). The comparison with Europe is striking. In 1990 and in 2005, White Americans in the highest-income area had mortality rates very similar to Europeans, while Black Americans suffered much higher mortality even in high-income areas. By 2018, European mortality rates were uniformly lower: the gap between Europeans and White Americans was generally larger than the gap between White and Black Americans. In large part this pattern is due to stagnation in US White mortality rates. The lower European mortality rates

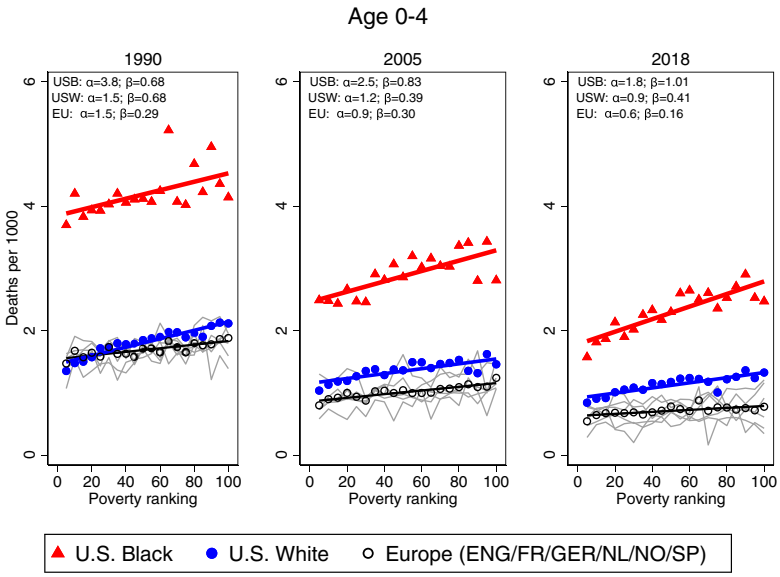


Fig. 1. One-year mortality for Black Americans, White Americans, and six European countries, ages 0 to 4, for (*Left*) 1990, (*Middle*) 2005, and (*Right*) 2018. Average 1-y mortality rates are plotted across poverty rate percentiles. For US Black (USB) and US White (USW) mortality, each bin represents a group of counties with about 5% of the overall population in the respective year. Black circles show population-weighted average mortality rates across England (ENG), France (FR), Germany (GER), the Netherlands (NL), Norway (NO), and Spain (SP), and each circle represents a group of municipalities or districts representing 5% of the overall population of each country in the respective year. Gray lines show mortality for each European country (see *SI Appendix*, Figs. S5–S9 for colorized figures with an extended set of European countries). Germany and Spain are included with 2016 data in *Right*. Straight lines provide linear fits. α and β refer to the fitted lines' intercepts and slopes, respectively. Additional numerical values are reported in *SI Appendix*, Table S1.

ECONOMIC SCIENCES

show the trajectory that might have been possible in a high-income country like the United States.

Fig. 4 shows trends for older adults (aged 65 to 79). This figure shows that mortality declined for both Black and White older adults in high- and low-income areas of the United States. Nevertheless, in the lowest-income areas, White American older adults went from having essentially the same mortality rates as Europeans in 1990 to having significantly higher rates in 2018: 27 per 1,000 in the United States compared to 20 per 1,000 in Europe. The mortality rate for Black American older adults aged 65 to 79 in low-income areas was even higher in 2018, at 32 per 1,000. We do not show mortality trends for adults older than age 80. For this group, we are lacking

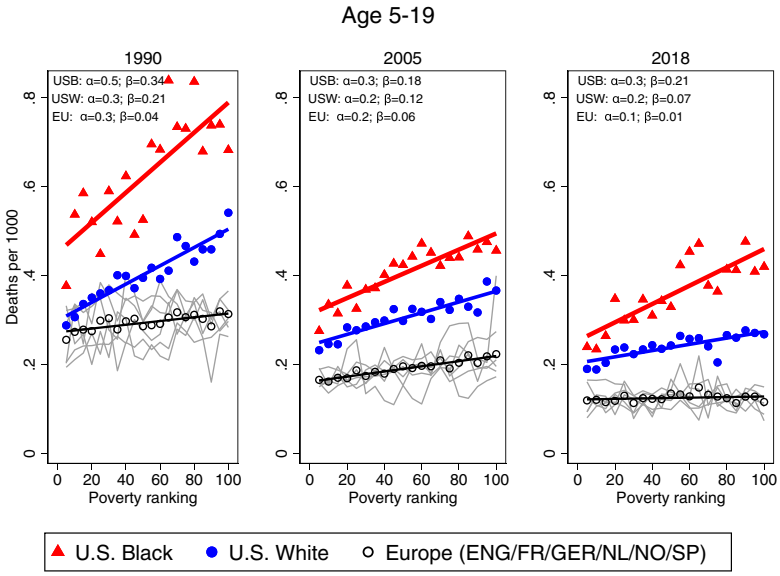


Fig. 2. One-year mortality for Black Americans, White Americans, and six European countries, ages 5 to 19, for (*Left*) 1990, (*Middle*) 2005, and (*Right*) 2018. Straight lines provide linear fits. α and β refer to the fitted lines' intercepts and slopes, respectively. For further notes, see Fig. 1. Numerical values and the slopes of fitted lines are reported in *SI Appendix*, Table S2.

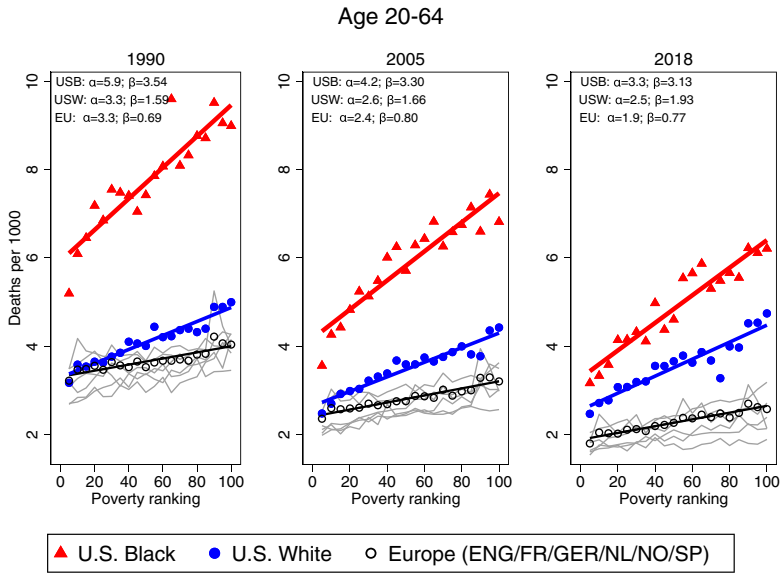


Fig. 3. One-year mortality for Black Americans, White Americans, and six European countries, ages 20 to 64, for (*Left*) 1990, (*Middle*) 2005, and (*Right*) 2018. Straight lines provide linear fits. α and β refer to the fitted lines' intercepts and slopes, respectively. For further notes, see Fig. 1. Numerical values and the slopes of fitted lines are reported in *SI Appendix*, Table S3.

the detailed data on death rates by single year of age to age-adjust the death rates, which is crucial to compare mortality across countries and over time.

Fig. 5 summarizes the trends in age-specific mortality by showing life expectancy at birth for Black and White persons for each year from 1990 to 2018. We have also drawn a trend line using data from 1990 through 2012. This figure highlights the strong convergence between Black and White American mortality rates between 1990 and 2012. Over this period, White American life expectancy continued to improve, but Black American life expectancy improved faster.

Fig. 5 shows that if mortality had continued to evolve at the same rate after 2012 as it did from 1990 to 2012, the gap in life expectancy between Black and White persons would have closed by 2036. However, improvements in life expectancy among both Black and White Americans faltered after about 2014. Both US White and Black American life expectancy plateaued and then fell between 2015 and 2018. Moreover, the decline in life expectancy

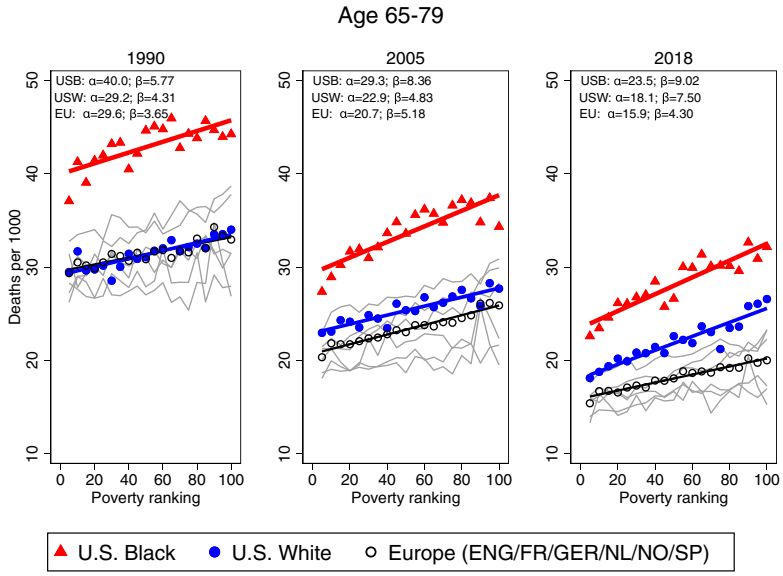


Fig. 4. One-year mortality for Black Americans, White Americans, and six European countries, ages 65 to 79, for (*Left*) 1990, (*Middle*) 2005, and (*Right*) 2018. Straight lines provide linear fits. α and β refer to the fitted lines' intercepts and slopes, respectively. For further notes, see Fig. 1. Numerical values and the slopes of fitted lines are reported in *SI Appendix*, Table S4.

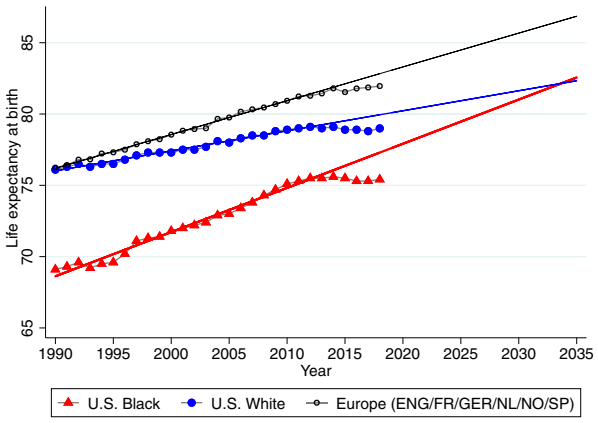


Fig. 5. Life expectancy for Black Americans, White Americans, and six European countries, extrapolated to 2035 fitting a linear trend through 1990 to 2012. Black American, White American, and European life expectancies are plotted over time and extrapolated to 2035 using a linear trend through 1990 to 2012. Black circles show the population weighted average life expectancy across England, France, Germany, the Netherlands, Norway, and Spain.

among Black Americans since 2015 appears to be even more severe than the decline among White Americans. Hence, although some observers have focused on the implications of the COVID-19 pandemic for Black/White American differences (40), even prepandemic, progress in improving the longevity of Black Americans and eliminating racial gaps in life expectancy had started to reverse.

A comparison with European mortality offers a useful perspective. In 1990, life expectancy among White Americans was the same as in the European benchmark countries. However, over the next 3 decades, White Americans increasingly fell behind Europeans. At the same time, life expectancy for Black Americans started far below both European and White American rates in 1990 but grew at a faster rate than European life expectancy.

Although European life expectancy remained above US life expectancy in 2018, European life expectancy also declined relative to the 1990 to 2012 trend after 2014, suggesting that there may be some element common to the United States and Europe that has moderated the growth of life expectancies in most high-income countries. It has been shown that the flattening of life expectancy in the United States was driven primarily by the plateauing of mortality improvements due to cardiovascular disease (41), and this may also be true in Europe.

Fig. 6 offers a closer look at the evolution of racial gaps in mortality by geographic area. As before, counties are sorted into population ventiles by overall poverty rate, but we focus on period life expectancy to summarize mortality rates across all ages. Each ventile represents approximately the same overall population, but race-specific populations are not balanced—in particular, Black people are overrepresented in the lower-income areas and underrepresented in the higher-income areas. This figure traces out the implications of that imbalance for the evolution of life expectancy. Fig. 6*A* shows the change in race-specific life expectancy between 1990 and 2018, calculated within each ventile. This panel confirms the evidence from Figs. 1–4 that between 1990 and 2018, Black American mortality declined faster than White mortality in all areas. The gap is larger in some ventiles than others but is sizeable in all but the highest-income 5% of counties. Fig. 6*B* confirms that Black people are overrepresented in lower-income counties and underrepresented in higher-income ones. Fig. 6*C* shows the contribution of each ventile group to overall life expectancy between 1990 and 2018 for Black and White Americans separately. In other words, Fig. 6*C* illustrates the impact on life expectancy if only mortality in a given ventile were allowed to change. Fig. 6*C* shows that improvements in the lowest-income counties made a disproportionate contribution to Black persons life expectancy gains.

Fig. 6*D* shows the contribution of each ventile to the reduction of the racial mortality gap. Life expectancy in the highest-income counties increased the racial life expectancy gap, not because Black Americans living in these places experienced smaller life expectancy gains than White Americans but because White Americans were overrepresented in high-income areas. In other words, mortality

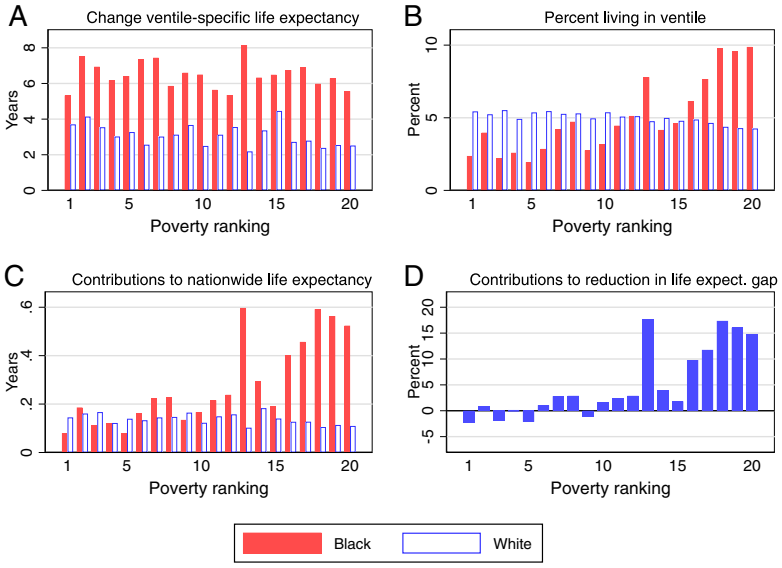


Fig. 6. Population distribution in 1990 and life expectancy contribution for 1990 to 2018, by ventile and race. (*A*) The change in race-specific life expectancy calculated within each ventile between 1990 and 2018. (*B*) The percent of the overall US Black and White population living in each ventile in 1990. (*C*) The contribution of the mortality changes in each ventile to the countrywide race-specific life expectancy. (*D*) The percent contribution of the mortality changes in each ventile to the reduction in the life expectancy gap between Black and White Americans.

improvements among Black Americans living in the highest-income counties had relatively little impact on overall life expectancy of Black people because fewer Black people enjoyed them, while the opposite was the case for White Americans. The result is an increase in the racial life expectancy gap. Similarly, the lowest-income areas contributed the most to life expectancy gains among Black people and to the closing of the racial life expectancy gap because Black Americans are overrepresented in these areas. These results suggest that given continuing overrepresentation of Black Americans in low-income places, improving life expectancy in these places is key to further reductions in racial life expectancy gaps.

SI Appendix, Fig. S1 offers a breakdown of the causes of death that were most responsible for the reduction in the Black–White American mortality gap. These results are subject to the usual caveats about the limitations of cause of death data, particularly when making comparisons over time (42). *SI Appendix*, Fig. S1*A* breaks out some of the most important contributors to changes in life expectancy separately for Black and White Americans. One can immediately see that cardiovascular disease and cancer are the most important individual contributors for both groups.

SI Appendix, Fig. S1*B* shows the percent contribution of each cause to the reduction of the racial difference in life expectancy. *SI Appendix*, Table S5 shows the corresponding numerical values. *SI Appendix*, Fig. S1*B* shows that cardiovascular mortality had the smallest impact on the gap across all causes because reductions in deaths from cardiovascular disease benefitted White and Black persons fairly similarly over this time period. The causes that contributed most to reductions in the gap were, in order of importance, so called "deaths of despair" (43) (16.18%; this category includes suicide, overdoses, and cirrhosis), cancer (15.96%), homicide (12.51%), deaths from causes originating in the fetal or infant period (11.05%), and HIV (9.89%). The importance of overdose deaths (44), homicide, and HIV in explaining racial differences in life expectancy has been previously noted (3–5, 45–48). However, it is notable in light of previous work that over the period we analyze, changes in mortality due to cardiovascular disease explain relatively little of the changing gap. This finding indicates that the role of cardiovascular mortality in closing the inequality gap has declined in the 2010s. Faster reductions in cancer deaths among Black Americans also seem to have had a larger impact than they did before 2010 (3–5).

Since the opioid epidemic is one of the most important causes of recent declines in US life expectancy relative to other countries (44), *SI Appendix*, Fig. S2*A* compares actual life expectancy with a counterfactual life expectancy computed by assuming that the rate of deaths due to drug overdoses had remained at its 1990 value. *SI Appendix*, Fig. S2*B* is similar but assumes that the broader category of deaths of despair, that is, deaths from drug overdose, suicides, and chronic liver disease, had not changed since 1990. *SI Appendix*, Fig. S2 shows that without drug overdoses, there would still have been a slight downturn in life expectancy around 2015 and 2016 but that life expectancy may well have continued upward after that, albeit with the flattening trend noted above. Comparing *SI Appendix*, Fig. S2 *A* and *B* shows that suicides and chronic liver disease are also important for White Americans, but for Black Americans, only overdoses have had a large impact in terms of life expectancy, and then only since about 2014.

SI Appendix, Fig. S3 and Table S7 provide a similar breakdown of which age groups contributed the most to the decline in the life expectancy gap between Black and White Americans. Consistent with the analysis by cause, the age groups that contributed the most are 0- to 4-y-olds and 20- to 64-y-olds, although this pattern varies somewhat over time (*SI Appendix*, Table S6). For example, 0- to 4-y-olds contributed 19.6% of the reduction between 1990 and 2000 and 9.83% between 2000 and 2018. The decomposition shows that the single years of age that contributed most to the closing of the gap were in infancy and among prime-age adults (aged 20 to 64). For Black Americans, contributions to improvements in life expectancy rose from age 20 to about age 65 and then declined. As has been noted in the literature (9–11, 49), this is strikingly not the case for White Americans, who experienced small declines in life expectancy from about ages 25 to 40, followed by only small gains to age 60.

Following Macinko and Elo (50), we also provide a breakdown of differences between Black and White Americans in preventable causes of death below age 65 in *SI Appendix*, Fig. S4 and Table S8 (all deaths are classified as in ref. 50, reported in *SI Appendix*, Table S9). *SI Appendix*, Fig. S4 indicates that Black Americans made major gains to life expectancy in terms of reductions in deaths from causes amenable to medical care and, to a lesser extent, from deaths amenable to intervention. These results indicate a continuation of the trends reported up to 2005 by Macinko and Elo (50), with the exception of ischemic heart disease, which decreased in relevance.

Discussion

We focus on the evolution of mortality in the 3 decades leading up to the COVID-19 pandemic in order to take stock of the improvements and remaining inequalities that were present in the United States before the pandemic struck. We view racial disparities through the lens of place, comparing gaps in the highest-income parts of the country to those in the lowest-income areas, and we use European mortality rates as a benchmark for assessing mortality differences and trends in those differences.

In 1990, there were remarkable mortality differences between Black and White Americans. For most age groups, Black Americans living in the highest-income US areas had substantially higher mortality rates than White Americans in the country's lowest-income areas. The mortality disadvantage of Black Americans in 1990 was even more pronounced when compared to European countries, while mortality rates of White Americans were fairly similar to those in Europe.

Since 1990, Black Americans experienced large mortality improvements across all age ranges and in both higher- and lower-income areas, although because Black Americans are more likely to live in lower-income counties, gains in these counties played an outsized role in reducing the racial life expectancy gap. These reductions in mortality were strong enough to reduce the racial mortality gap by 48.9%, despite the fact that White Americans also experienced mortality improvements. Between 1990 and 2018, the US Black–White American life expectancy gap decreased from 7.0 to 3.6 y, while the gap between the six European countries and Black Americans decreased from 7.1 to 6.5 y.

Mortality improvements among Black Americans and the closing of the racial mortality gap stalled after 2012. Moreover, despite mortality improvements since 1990, White Americans have increasingly lost ground compared to Europeans, with substantial gaps in mortality rates opening between Europeans and White Americans. The gap between Black Americans and the six European countries included decreased by 19% between 1990 and 2012 but only by 8.3% in the overall period from 1990 and 2018. Hence, the convergence in the US Black–White American mortality gap reflects real progress among Black Americans even relative to a non-US benchmark, but this progress has reversed after 2012.

The diverging mortality experience between the United States and Europe is especially evident when analyzing a larger set of nine European countries, although some of these countries are missing data for 1990 (*SI Appendix*, Figs. S5–S9). Despite strong differences in social and economic starting points across these European countries, by 2018 their mortality gradients fall into a narrow band. Even countries like Portugal, which was much lower income than the European average in 1990, or the Czech Republic, which experienced the fall of the Soviet Union, were able to catch up to higher-income and more stable European countries in terms of mortality rates by 2018. European mortality rates in 2018 lie below US White American mortality rates for each country and across high- and low-income areas. This US health disadvantage

even among economically advantaged groups in high-income US locations has also been shown for a broad set of health conditions (51, 52).

Another remarkable observation is how flat mortality gradients are at younger ages across all European countries. This pattern shows that health improvements among infants, children, and youth have been disseminated within European countries in a way that includes even the lowest-income areas. It suggests that there is great potential for disadvantaged infants, children, and youth living in lower-income US areas to catch up to European standards.

Focusing on the disparity between Black and White Americans, we show that improvements in Black Americans' relative to White Americans' life expectancy in the lowest-income counties had the greatest impact in narrowing the gap overall. In terms of mortality causes driving these improvements, greater reductions in Black relative to White American death rates due to cancer, homicide, HIV, and causes originating in the fetal or infant period had the largest impacts, while smaller increases in Black American compared to White American deaths of despair also closed the gap. Consistent with the importance of these causes, we find that rapid reductions in Black relative to White American deaths in early childhood and prime-aged adults (aged 20 to 64) accounted for the majority of the closure of the gap. Consistent once again with the importance of these causes (49), deaths due to causes amenable to medical care showed greater reductions for Black Americans relative to White Americans, greatly contributing to the closure of the gap in life expectancy. Reductions in causes amenable to intervention also played an important role, in line with prior research studying longitudinal racial disparities in more nuanced health indicators (53).

Many authors have commented on the role of systemic racism in shortening Black relative to White American life expectancies in the United States (54–56). Unpacking some of the dimensions of racism suggests that there are many possible reasons for these broad improvements in the health of Black Americans. The literature on the relationship between education and health suggests that improvements in the quantity and quality of education available to Black children and young adults over the past decades is one possible contributor to improved longevity and reduced gaps in life expectancy spread broadly over prime-aged adults (aged 20 to 64) (57, 58). Our results suggest that improvements in the availability of medical care are also likely to have been important in reducing racial disparities in mortality (59). Health care developments that may have been particularly important include expansions of the Medicaid program to cover pregnant women and children starting in the 1990s, which likely account for much of the improvement among infants (60), as well as improved access to treatment for cancer and HIV. Long-term health effects of access to Medicaid as well as other safety net programs such as food stamps and the Earned Income Tax Credit may also be an important contributor to mortality reductions (61). Reductions in pollution may have played a role given that Black Americans are more likely than White Americans to live in more-polluted areas (62–64).

Despite the strong mortality improvements among Black Americans over the past 3 decades, a dramatic gap remains, and this gap has increased again in recent years. It is important to understand which medical, social, and policy developments helped to increase the longevity of Black Americans through 2012 and how these positive changes can be reinforced over the coming decades with the ultimate goal of fully closing the racial longevity gap in the United States. Moreover, the comparisons with Europe suggest that mortality rates of both Black and White Americans could fall much further across all ages and across higher- and lower-income areas.

Materials and Methods

US Mortality. US Black and White American mortality rates are constructed by dividing death counts by population estimates for single years of age, county, and calendar year. Death counts come from the US Vital Statistics mortality files, while population estimates are provided by the National Center for Health Statistics (NCHS). The NCHS estimates are "bridged"; that is, they convert multiple race categories reported in the 2000 and 2010 Censuses back to single race categories comparable with those reported on the death certificates. Throughout the paper, "Black" populations include both non-Hispanic and Hispanic Black persons, while "White" populations include both non-Hispanic and Hispanic White persons. Neither race group includes American Indian and Alaska Native, Asian, or Native Hawaiian and Other Pacific Islander (the NCHS notes that these race categories "represent a social-political construct and are not anthropologically or biologically based") (65). Mortality rates spanning multiple years of age are age-adjusted using the 2015 US population. We age-adjust because in an age bracket such as ages 65 to 79, a group with more 79-y-olds would be expected to have higher mortality.

US Poverty Ranking. As in ref. 25, we rank counties by their poverty rate and place them into groups of fixed population size. This allows us to analyze trends in age-specific mortality across areas ranked by an area's poverty rate while taking into account population shifts across areas. We rank all US counties in 1990, 2005, and 2018 by their poverty level and then divide them into 20 groups, each representing roughly 5% of the overall US population. This way we can compare, for example, the 5% of the population living in the lowest-income counties in 1990 with the 5% of the population living in the highest-income counties in 1990 and analyze how the mortality differences between these groups change over time. We refer to the county groups with the highest (lowest) fractions of their populations in poverty as the lowest-income (highest-income) counties. Our approach reassigns county groups in 1990, 2005, and 2018 to adjust for changes in county ranking and population size. Poverty rates are taken from the 1990 and 2000 Censuses and the 2014 to 2018 American Community Survey 5-y sample and interpolated for intermediate years.

European Mortality. Data for nine European countries (Czech Republic, England, Finland, France, Germany, the Netherlands, Norway, Portugal, and Spain) come from the Institute for Fiscal Studies (IFS) project on Geographic Approaches to Inequalities in Mortality (66) and are treated similarly. Additional details on data sources are provided in *SI Appendix*.

Figs. 1–4 include the six European countries that provide consistent mortality data from 1990 onward (England, France, Germany, the Netherlands, Norway, and Spain). Figs. 1–4 include a mean Europe mortality rate, representing the population-weighted average of mortality rates across these countries in each ventile. *SI Appendix*, Figs. S5–S9 show analogous figures using all nine of the European countries in the IFS study. See *SI Appendix* for further information about these countries.

Mortality rates across all countries and years are age-adjusted using the 2015 US population, based on 5-y age groups. The following describes the respective area definitions, ranking measures, and available data years for each of the European countries included in Figs. 1–4: England, local authorities ranked by a deprivation index, for 1992 to 2017; France, départements ranked by poverty rate, for 1990 to 2018; Germany, districts ranked by per-capita income in 1990, 2005, and 2016 (1990 excludes East Germany because of the exceptionally high mortality in East Germany around the time of reunification); the Netherlands, municipalities ranked by poverty rate, 1995, 2005, and 2016; Norway, municipalities ranked by poverty rate, 1990 to 2018; and Spain, municipalities ranked by median income, 1990 to 2016. Further details on area definitions and the poverty or deprivation variables used for ranking areas can be found in ref. 51.

Life Expectancy Data. We construct US Black and White American life expectancy at birth based on 1-y mortality rates for the years 1990 to 2018. The Human Mortality Database (67) provides life expectancy estimates for the European countries in our study, while life expectancy estimates for England are provided by the United Kingdom Office for National Statistics.

Decompositions. We offer breakdowns of the contributions of location, age, and cause of death to Black and White American life expectancy and to the closing of the racial mortality gap. Each breakdown is based on asking how life expectancy would have changed if all other factors besides the one being considered had remained constant at their 1990 levels. For example, we ask how life expectancy would have changed between 1990 and 2018 if only the

ECONOMIC SCIENCES

homicide rate had fallen while all other causes remained at their 1990 values. In all cases, this hypothetical exercise is conducted separately for Black and White Americans. Letting only one mortality rate change while keeping all other rates constant understates the one rate's overall contribution to life expectancy if other rates also improved. The reason is that the life expectancy formula interacts mortality rates at all ages. For example, a higher survival rate in old age makes improvements in infant mortality more valuable and vice versa. Our hypothetical life expectancy measure ignores interaction effects because they cannot be assigned to a specific age. Hence, our results should be interpreted as the partial effect of a given factor, which is sometimes referred to as the exclusive life expectancy impact of an age-specific effect (68). For further details, see *SI Appendix*.

Data Availability. All datasets used in this study are publicly available online or can be obtained from the National Vital Statistics offices. See *SI Appendix*, section C, for additional details and download links.

ACKNOWLEDGMENTS. We thank J.B., S.K., and K.G.S. for organizing the IFS working group on geographical approaches to measuring inequality in mortality, with financial support from the Economic and Social Research Council Centre for the Microeconomic Analysis of Public Policy at IFS (Grant ES/T014334/1). C.C. received support from the Science and Technology Foundation (Fundação para a Ciência e a Tecnologia [FCT]), the European Social Fund, and the Centro Operational Programme (Grant SFRH/BD/132218/2017). P.S. received support from the Centre of Studies in Geography and Spatial Planning (Grant UIDB/04084/2020), through an FCT fund.

1. C. J. L. Murray *et al*., Eight Americas: Investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Med.* **3**, e260 (2006).
2. M. S. Shiels *et al*., Trends in premature mortality in the USA by sex, race, and ethnicity from 1999 to 2014: An analysis of death certificate data. *Lancet* **389**, 1043–1054 (2017).
3. S. Harper, J. Lynch, S. Burris, G. Davey Smith, Trends in the black-white life expectancy gap in the United States, 1983–2003. *JAMA* **297**, 1224–1232 (2007).
4. N. E. Adler, D. H. Rehkopf, U.S. disparities in health: Descriptions, causes, and mechanisms. *Annu. Rev. Public Health* **29**, 235–252 (2008).
5. S. Harper, D. Rushani, J. S. Kaufman, Trends in the black-white life expectancy gap, 2003-2008. *JAMA* **307**, 2257–2259 (2012).
6. J. Currie, H. Schwandt, Mortality inequality: The good news from a county-level approach. *J. Econ. Perspect.* **30**, 29–52 (2016).
7. A. Case, A. Deaton, Life expectancy in adulthood is falling for those without a BA degree, but as educational gaps have widened, racial gaps have narrowed. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024777118 (2021).
8. B. Bosworth, K. Zhang, Evidence of increasing differential mortality: A comparison of the HRS and SIPP. *SSRN* [Preprint] (2015). https://doi.org/10.2139/ssrn.2625792 (Accessed 14 September 2021).
9. A. Case, A. Deaton, Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15078–15083 (2015).
10. A. Case, A. Deaton, Mortality and morbidity in the 21st century. *Brookings Pap. Econ. Act.* **2017**, 397–476 (2017).
11. A. Case, A. Deaton, *Deaths of Despair and the Future of Capitalism* (Princeton University Press, 2020).
12. R. Chetty *et al*., The association between income and life expectancy in the United States, 2001–2014. *JAMA* **315**, 1750–1766 (2016).
13. D. M. Cutler, F. Lange, E. Meara, S. Richards-Shubik, C. J. Ruhm, Rising educational gradients in mortality: The role of behavioral risk factors. *J. Health Econ.* **30**, 1174–1187 (2011).
14. E. R. Meara, S. Richards, D. M. Cutler, The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Aff. (Millwood)* **27**, 350–360 (2008).
15. J. K. Montez, L. F. Berkman, Trends in the educational gradient of mortality among US adults aged 45 to 84 years: Bringing regional context into the explanation. *Am. J. Public Health* **104**, e82–e90 (2014).
16. J. K. Montez, A. Zajacova, Explaining the widening education gap in mortality among U.S. white women. *J. Health Soc. Behav.* **54**, 166–182 (2013).
17. National Academies of Sciences, Engineering, and Medicine *et al*., *The Growing Gap in Life Expectancy by Income: Implications for Federal Programs and Policy Responses* (National Academies Press, 2015).
18. S. J. Olshansky *et al*., Differences in life expectancy due to race and educational differences are widening, and many may not catch up. *Health Aff. (Millwood)* **31**, 1803–1813 (2012).
19. G. Pappas, S. Queen, W. Hadden, G. Fisher, The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N. Engl. J. Med.* **329**, 103–109 (1993).
20. J. Pijoan-Mas, J.-V. Ríos-Rull, Heterogeneity in expected longevities. *Demography* **51**, 2075–2102 (2014).
21. S. H. Preston, I. T. Elo, Are educational differentials in adult mortality increasing in the United States? *J. Aging Health* **7**, 476–496 (1995).
22. H. Waldron, Mortality differentials by lifetime earnings decile: Implications for evaluations of proposed Social Security law changes. *Soc. Secur. Bull.* **73**, 1–37 (2013).
23. A. S. Venkataramani, R. O'Brien, A. C. Tsai, Declining life expectancy in the United States: The need for social policy as health policy. *JAMA* **325**, 621–622 (2021).
24. M. Baker, J. Currie, H. Schwandt, Mortality inequality in Canada and the United States: Divergent or convergent trends? *J. Labor Econ.* **37**, S325–S353 (2019).
25. J. Currie, H. Schwandt, Inequality in mortality decreased among the young while increasing for older adults, 1990–2010. *Science* **352**, 708–712 (2016).
26. J. Currie, H. Schwandt, J. Thuilliez, Pauvreté, Egalité, Mortalité: Mortality (in)equality in France and the United States. *J. Popul. Econ.* **33**, 197–231 (2020).
27. M. Ezzati, A. B. Friedman, S. C. Kulkarni, C. J. L. Murray, The reversal of fortunes: Trends in county mortality and cross-county mortality disparities in the United States. *PLoS Med.* **5**, e66 (2008).
28. N. Krieger *et al*., The fall and rise of US inequities in premature mortality: 1960–2002. *PLoS Med.* **5**, e46 (2008).
29. H. Wang, A. E. Schumacher, C. E. Levitz, A. H. Mokdad, C. J. Murray, Left behind: Widening disparities for males and females in US county life expectancy, 1985–2010. *Popul. Health Metr.* **11**, 8 (2013).
30. G. K. Singh, M. Siahpush, Widening socioeconomic inequalities in US life expectancy, 1980-2000. *Int. J. Epidemiol.* **35**, 969–979 (2006).
31. Y. C. Vierboom, S. H. Preston, A. S. Hendi, Rising geographic inequality in mortality in the United States. *SSM Popul. Health* **9**, 100478 (2019).
32. J. B. Dowd, A. Hamoudi, Is life expectancy really falling for groups of low socioeconomic status? Lagged selection bias and artefactual trends in mortality. *Int. J. Epidemiol.* **43**, 983–988 (2014).
33. A. S. Hendi, Trends in U.S. life expectancy gradients: The role of changing educational composition. *Int. J. Epidemiol.* **44**, 946–955 (2015).
34. T. Goldring, F. Lange, S. Richards-Shubik, Testing for changes in the SES-mortality gradient when the distribution of education changes too. *J. Health Econ.* **46**, 120–130 (2016).
35. A. S. Hendi, I. T. Elo, P. Martikainen, The implications of changing education distributions for life expectancy gradients. *Soc. Sci. Med.* **272**, 113712 (2021).
36. J. R. Wilmoth, C. Boe, M. Barbieri, "Geographic differences in life expectancy at age 50 in the United States compared with other high-income countries" in *International Differences in Mortality at Older Ages: Dimensions and Sources*, E. M. Crimmins, S. H. Preston, B. Cohen, Eds. (The National Academies Press, 2010), pp. 333–366.
37. J. P. Mackenbach *et al*., Socioeconomic inequalities in health in 22 European countries. *N. Engl. J. Med.* **358**, 2468–2481 (2008).
38. J. P. Mackenbach *et al*., Changes in mortality inequalities over two decades: Register based study of European countries. *BMJ* **353**, i1732 (2016).
39. J. Y. Ho, A. S. Hendi, Recent trends in life expectancy across high income countries: Retrospective observational study. *BMJ* **362**, k2562 (2018).
40. T. Andrasfay, N. Goldman, Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014746118 (2021).
41. N. K. Mehta, L. R. Abrams, M. Myrskylä, US life expectancy stalls due to cardiovascular disease, not drug deaths. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 6998–7000 (2020).
42. I. T. Elo, G. L. Drevenstedt, Cause-specific contributions to black-white differences in male mortality from 1960 to 1995. *Demogr. Res.* **2** (Special), 255–276 (2004).
43. A. Case, A. Deaton, Mortality and morbidity in the 21st century. *Brookings Pap. Econ. Act.* **2017**, 397–476 (2017).
44. S. Harper, C. A. Riddell, N. B. King, Declining life expectancy in the United States: Missing the trees for the forest. *Annu. Rev. Public Health* **42**, 381–403 (2021).
45. K. D. Kochanek, J. D. Maurer, H. M. Rosenberg, Why did black life expectancy decline from 1984 through 1989 in the United States? *Am. J. Public Health* **84**, 938–944 (1994).
46. M. Roberts, E. N. Reither, S. Lim, Contributors to the black-white life expectancy gap in Washington D.C. *Sci. Rep.* **10**, 13416 (2020).
47. P. Sharkey, M. Friedson, The impact of the homicide decline on life expectancy of African American males. *Demography* **56**, 645–663 (2019).
48. A. Leive, C. Ruhm, *Education Gradients in Mortality Trends by Gender and Race* (National Bureau of Economic Research, 2021).
49. I. T. Elo, A. S. Hendi, J. Y. Ho, Y. C. Vierboom, S. H. Preston, Trends in Non-Hispanic White mortality in the United States by metropolitan-nonmetropolitan status and region, 1990–2016. *Popul. Dev. Rev.* **45**, 549–583 (2019).
50. J. Macinko, I. T. Elo, Black-white differences in avoidable mortality in the USA, 1980-2005. *J. Epidemiol. Community Health* **63**, 715–721 (2009).
51. E. J. Emanuel *et al*., Comparing health outcomes of privileged US Citizens with those of average residents of other developed countries. *JAMA Intern. Med.* **181**, 339–344 (2021).
52. J. Banks, M. Marmot, Z. Oldfield, J. P. Smith, Disease and disadvantage in the United States and in England. *JAMA* **295**, 2037–2045 (2006).
53. K. M. Harris, P. Gordon-Larsen, K. Chantala, J. R. Udry, Longitudinal trends in race/ethnic disparities in leading health indicators from adolescence to young adulthood. *Arch. Pediatr. Adolesc. Med.* **160**, 74–81 (2006).
54. D. R. Williams, C. Collins, US socioeconomic and racial differences in health: Patterns and explanations. *Annu. Rev. Sociol.* **21**, 349–386 (1995).
55. D. R. Williams, P. B. Jackson, Social sources of racial disparities in health. *Health Aff. (Millwood)* **24**, 325–334 (2005).
56. J. C. Phelan, B. G. Link, Is racism a fundamental cause of inequalities in health? *Annu. Rev. Sociol.* **41**, 311–330 (2015).
57. I. T. Elo, S. H. Preston, Educational differentials in mortality: United States, 1979-85. *Soc. Sci. Med.* **42**, 47–57 (1996).

58. D. Frisvold, E. Golberstein, The effect of school quality on black-white health differences: Evidence from segregated southern schools. *Demography* **50**, 1989–2012 (2013).
59. A. Nelson, Unequal treatment: Confronting racial and ethnic disparities in health care. *J. Natl. Med. Assoc.* **94**, 666–668 (2002).
60. A. D. Racine, R. Kaestner, T. J. Joyce, G. J. Colman, Differential impact of recent Medicaid expansions by race and ethnicity. *Pediatrics* **108**, 1135–1142 (2001).
61. H. Hoynes, D. W. Schanzenbach, D. Almond, Long-run impacts of childhood access to the safety net. *Am. Econ. Rev.* **106**, 903–934 (2016).
62. J. Currie, J. Voorheis, R. Walker, *What Caused Racial Disparities in Particulate Exposure to Fall?: New Evidence from the Clean Air Act and Satellite-based Measures of Air Quality* (National Bureau of Economic Research, 2020).
63. S. Banzhaf, L. Ma, C. Timmins, Environmental justice: The economics of race, place, and pollution. *J. Econ. Perspect.* **33**, 185–208 (2019).
64. J. Currie, Inequality at birth: Some causes and consequences. *Am. Econ. Rev.* **101**, 1–22 (2011).
65. Centers for Disease Control and Prevention (CDC), Documentation for Bridged-Race Vintage 2006 (July 1, 2000 - July 1, 2006) Postcensal Population Estimates for Calculating Vital Rates. https://wonder.cdc.gov/wonder/help/populations/bridged-race/Estimates2000-06.html. Accessed 14 September 2021.
66. J. Banks, J. Currie, S. Krutikova, K. G. Salvanes, H. Schwandt, The evolution of mortality inequality in 11 OECD countries: Introduction. *Fisc. Stud.* **42**, 9–23 (2021).
67. M. Barbieri *et al.*, Data resource profile: The human mortality database (HMD). *Int. J. Epidemiol.* **44**, 1549–1556 (2015).
68. E. E. Arriaga, Measuring and explaining the change in life expectancies. *Demography* **21**, 83–96 (1984).

Appendix to Chapter 5

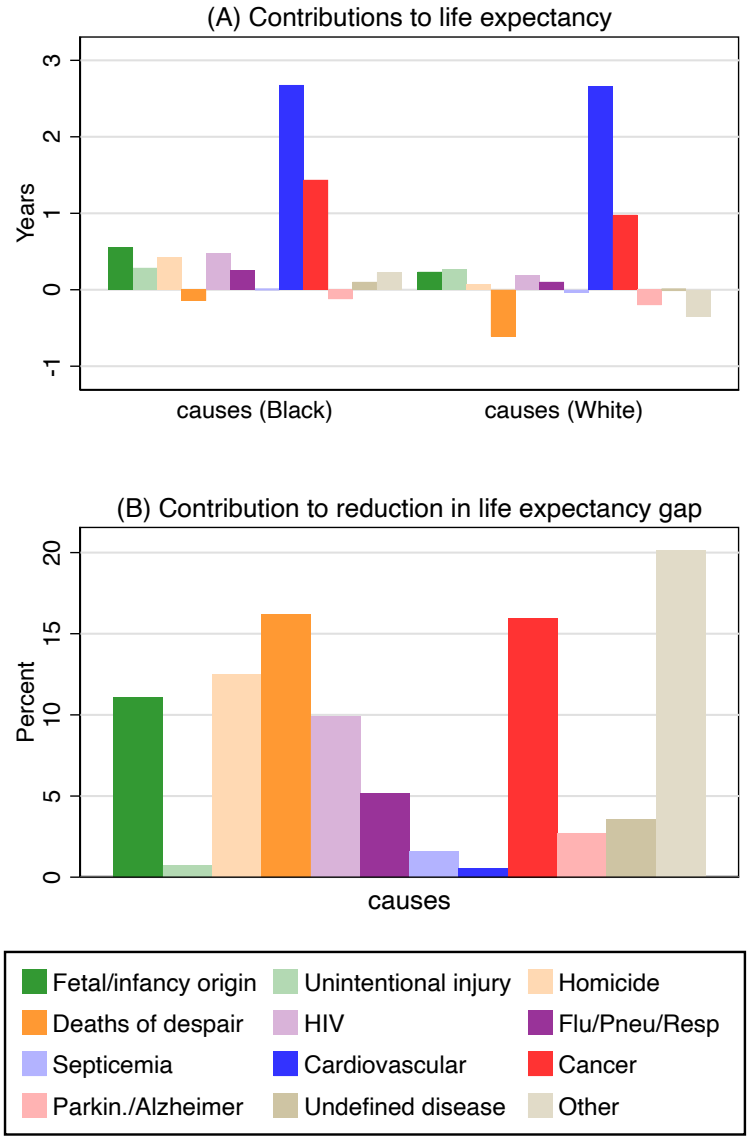
Supplementary information

Supplementary information for:

Schwandt, Hannes, Janet Currie, Marlies Bär, James Banks, Paola Bertoli, Aline Bütikofer, Sarah Cattan, Beatrice Zong-Ying Chao, Claudia Costa, Libertad González, Veronica Grembi, Kristiina Huttunen, René Karadakic, Lucy Kraftman, Sonya Krutikova, Stefano Lombardi, Peter Redler, Carlos Riumallo-Herl, Ana Rodríguez-González, Kjell Salvanes, Paula Santana, Josselin Thuilliez, Eddy van Doorslaer, Tom Van Ourti, Joachim Winter, Bram Wouterse, and Amelie Wuppermann.
"Inequality in Mortality between Black and White Americans by Age, Place, and Cause, and in Comparison to Europe, 1990--2018" (2021)

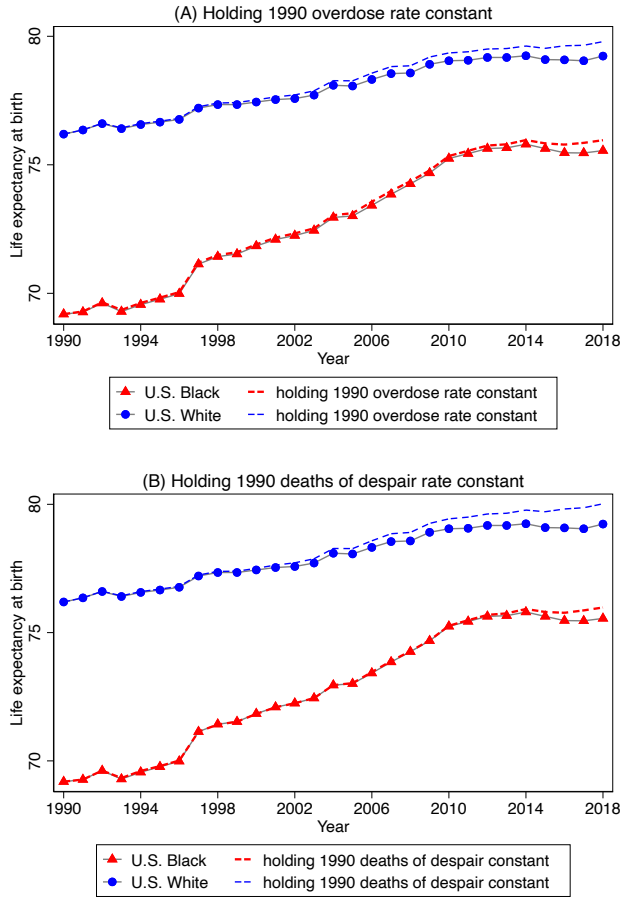
A Supplementary figures and tables

Figure S1: Cause-specific contributions to life expectancy gains and to the reduction of the Black-White life expectancy gap, 1990-2018



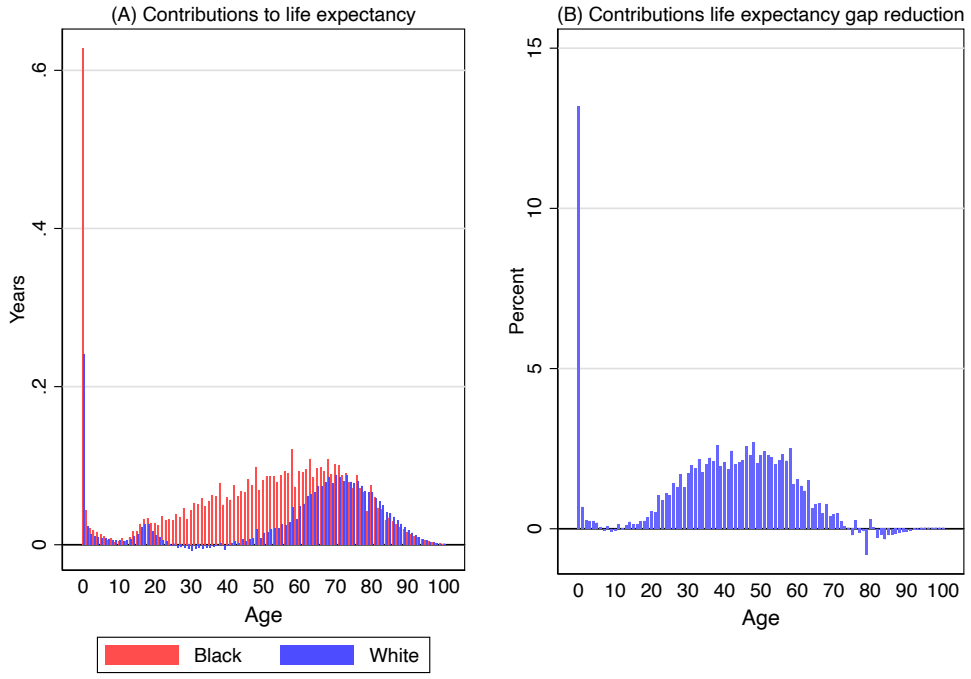
Notes: The top panel (A) shows the change in hypothetical race-specific life expectancy between 1990 and 2018 if only mortality from a given cause had changed between 1990 and 2018. The bottom panel (B) shows the percent contribution to the reduction in the Black-White life expectancy gap between 1990 and 2018. Numerical values and the list of included causes are provided in Appendix Tables S5 and S6.

Figure S2: Black Americans and White Americans actual and counterfactual life expectancy, holding drug overdose deaths and deaths of despair constant



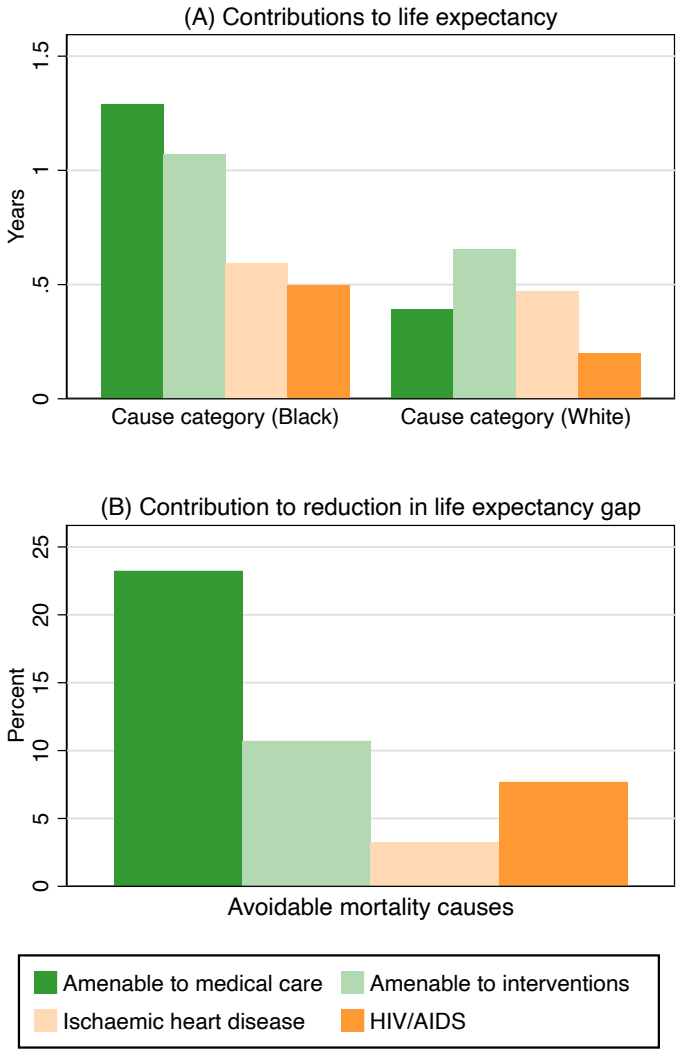
Notes: The red triangles and blue circles show actual life expectancy for Black Americans and White persons, respectively, over time. The dashed lines show counterfactual life expectancy estimates assuming constant death rates for drug overdose deaths (top panel) and constant death rates for deaths of despair (bottom panel), respectively, at their 1990 level. Source: Authors' calculations based on Vital Statistics mortality data.

Figure S3: Age-specific contributions to life expectancy gains and to the reduction of the Black-White life expectancy gap, 1990-2018



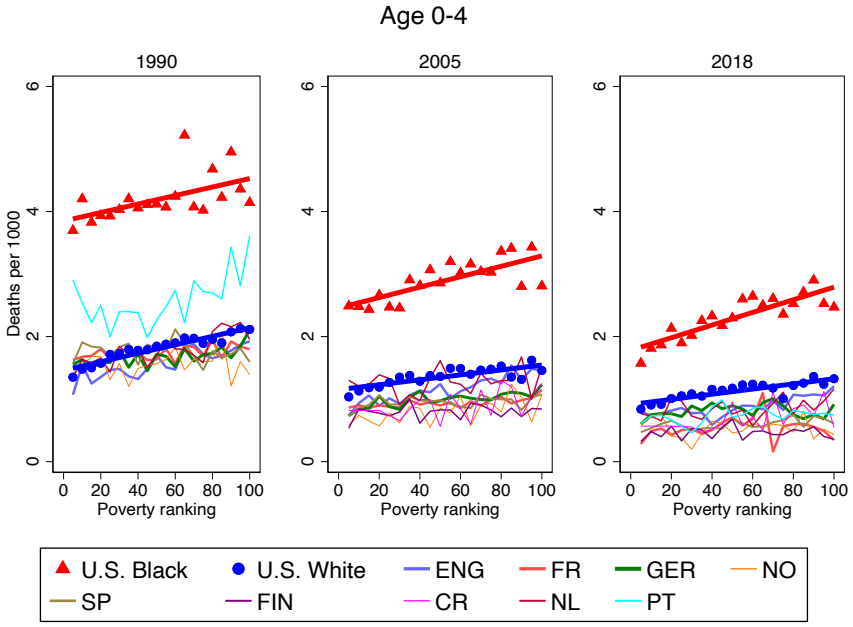
Notes: The top panel (A) shows the change in hypothetical race-specific life expectancy between 1990 and 2018 if only mortality at a given age had changed between 1990 and 2018. The bottom panel (B) shows the percent contribution to the reduction in the Black-White life expectancy gap between 1990 and 2018. Numerical values and contributions by subperiod are provided in Appendix Tables S7.

Figure S4: Contributions of avoidable mortality below age 65 (defined following Macinko and Elo (2009)) to life expectancy gains and to the reduction of the Black--White life expectancy gap, 1990-2018



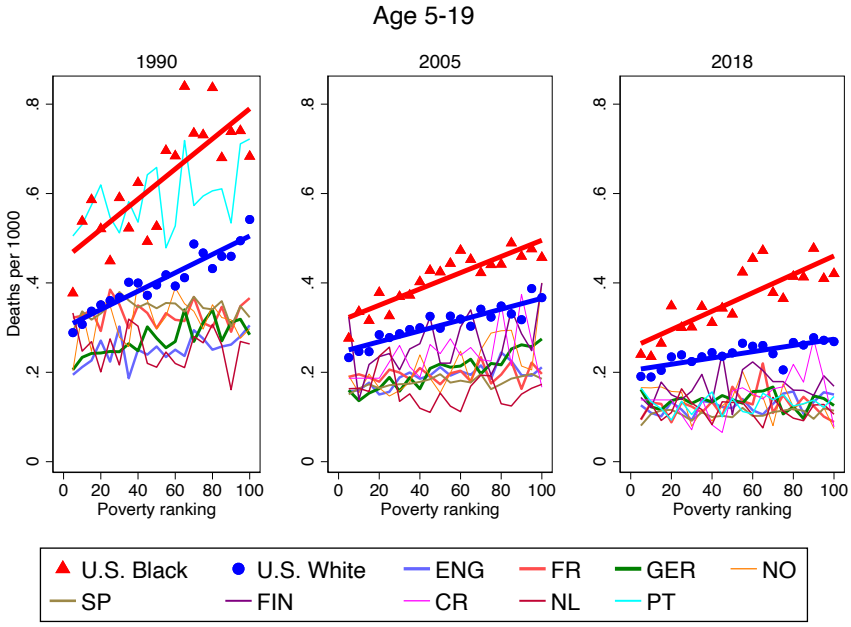
Notes: The top panel (A) shows the change in hypothetical race-specific life expectancy between 1990 and 2018 if only mortality in a given cause category, and below age 65, had changed between 1990 and 2018. The bottom panel (B) shows the percent contribution to the reduction in the Black-White life expectancy gap between 1990 and 2018. Mortality categories and age restrictions are defined following Macinko and Elo (2009), and causes in each category are listed in Table S8.

Figure S5: 1-year mortality for Black Americans, White Americans, and nine European countries, age 0-4.



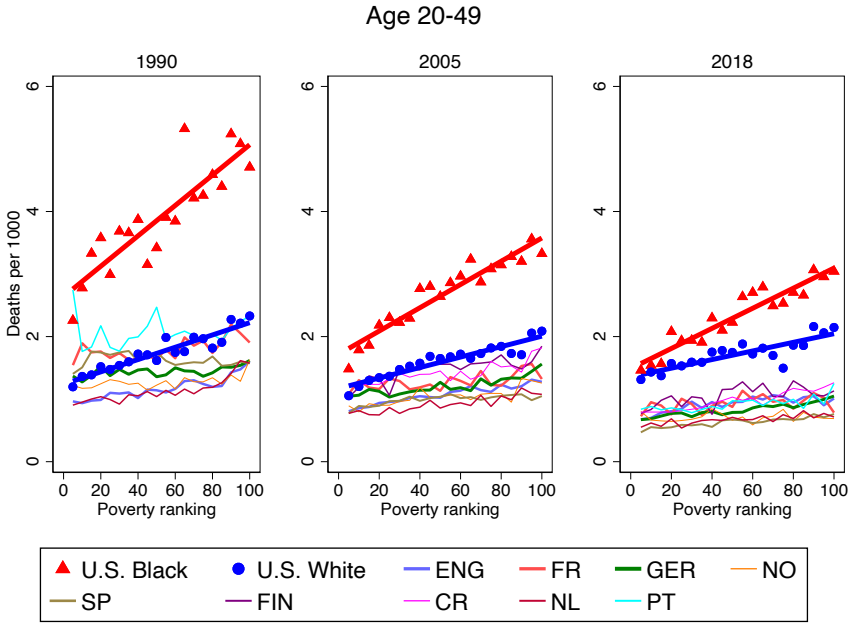
Notes: Average 1-year mortality rates are plotted across poverty rate percentiles for Black Americans, White Americans, Finland, France, the Netherlands, and Norway; across median income percentiles (higher percentiles refers to lower income) for Czech Republic, Germany, and Spain; across deprivation percentiles for England and Portugal. Each bin represents a group of counties, districts, or municipalities with about 5% of the overall population in the respective year. Germany excludes East Germany in 1990. The 2018 subpanel shows 2018 rates for the U.S., France, Norway, and Portugal; 2017 rates for England; 2016 rates for Czech Republic, Germany, Netherlands, and Spain; 2015 rates for Finland. More details provided in the Materials and Methods Section. Straight lines provide linear fits for Black Americans and White Americans mortality.

Figure S6: 1-year mortality for Black Americans, White Americans, and nine European countries, age 5-19.



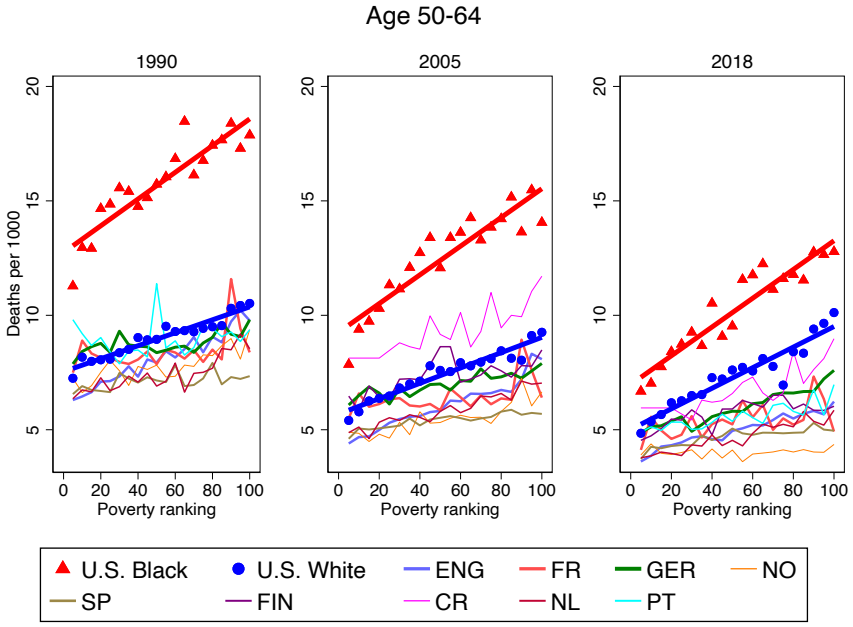
Notes: Average 1-year mortality rates are plotted across poverty rate percentiles for Black Americans, White Americans, Finland, France, the Netherlands, and Norway; across median income percentiles (higher percentiles refers to lower income) for Czech Republic, Germany, and Spain; across deprivation percentiles for England and Portugal. Each bin represents a group of counties, districts, or municipalities with about 5% of the overall population in the respective year. Germany excludes East Germany in 1990. The 2018 subpanel shows 2018 rates for the U.S., France, Norway, and Portugal; 2017 rates for England; 2016 rates for Czech Republic, Germany, Netherlands, and Spain; 2015 rates for Finland. More details provided in the Materials and Methods Section. Straight lines provide linear fits for Black Americans and White Americans mortality.

Figure S7: 1-year mortality for Black Americans, White Americans, and nine European countries, age 20-49.



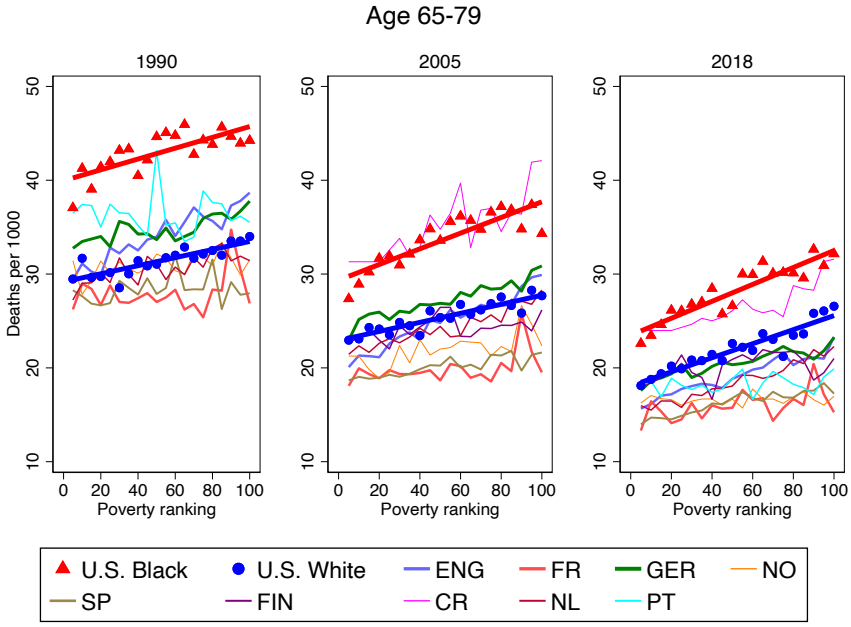
Notes: Average 1-year mortality rates are plotted across poverty rate percentiles for Black Americans, White Americans, Finland, France, the Netherlands, and Norway; across median income percentiles (higher percentiles refers to lower income) for Czech Republic, Germany, and Spain; across deprivation percentiles for England and Portugal. Each bin represents a group of counties, districts, or municipalities with about 5% of the overall population in the respective year. Germany excludes East Germany in 1990. The 2018 subpanel shows 2018 rates for the U.S., France, Norway, and Portugal; 2017 rates for England; 2016 rates for Czech Republic, Germany, Netherlands, and Spain; 2015 rates for Finland. More details provided in the Materials and Methods Section. Straight lines provide linear fits for Black Americans and White Americans mortality.

Figure S8: 1-year mortality for Black Americans, White Americans, and nine European countries, age 50-64.



Notes: Average 1-year mortality rates are plotted across poverty rate percentiles for Black Americans, White Americans, Finland, France, the Netherlands, and Norway; across median income percentiles (higher percentiles refers to lower income) for Czech Republic, Germany, and Spain; across deprivation percentiles for England and Portugal. Each bin represents a group of counties, districts, or municipalities with about 5% of the overall population in the respective year. Germany excludes East Germany in 1990. The 2018 subpanel shows 2018 rates for the U.S., France, Norway, and Portugal; 2017 rates for England; 2016 rates for Czech Republic, Germany, Netherlands, and Spain; 2015 rates for Finland. More details provided in the Materials and Methods Section. Straight lines provide linear fits for Black Americans and White Americans mortality.

Figure S9: 1-year mortality for Black Americans, White Americans, and nine European countries, age 65-79.



Notes: Average 1-year mortality rates are plotted across poverty rate percentiles for Black Americans, White Americans, Finland, France, the Netherlands, and Norway; across median income percentiles (higher percentiles refers to lower income) for Czech Republic, Germany, and Spain; across deprivation percentiles for England and Portugal. Each bin represents a group of counties, districts, or municipalities with about 5% of the overall population in the respective year. Germany excludes East Germany in 1990. The 2018 subpanel shows 2018 rates for the U.S., France, Norway, and Portugal; 2017 rates for England; 2016 rates for Czech Republic, Germany, Netherlands, and Spain; 2015 rates for Finland. More details provided in the Materials and Methods Section. Straight lines provide linear fits for Black Americans and White Americans mortality.

Table S1: Mortality rates per 1,000 across poverty ventiles, age 0-4.

	1990			2005			2018		
	U.S. Black	U.S. White	Europe	U.S. Black	U.S. White	Europe	U.S. Black	U.S. White	Europe
Overall	4.20	1.82	1.70	2.90	1.36	1.02	2.31	1.13	0.71
Percentile									
5	3.70	1.35	1.47	2.49	1.04	0.80	1.57	0.84	0.54
10	4.20	1.48	1.68	2.48	1.13	0.90	1.81	0.91	0.64
15	3.83	1.50	1.57	2.43	1.18	0.92	1.87	0.92	0.67
20	3.93	1.57	1.64	2.66	1.19	1.00	2.13	1.01	0.68
25	3.93	1.72	1.58	2.47	1.27	0.94	1.90	1.05	0.68
30	4.03	1.73	1.74	2.46	1.35	0.87	2.02	1.08	0.69
35	4.20	1.80	1.63	2.91	1.38	1.03	2.26	1.05	0.65
40	4.06	1.78	1.63	2.82	1.29	1.03	2.33	1.16	0.69
45	4.11	1.79	1.58	3.07	1.38	1.00	2.17	1.14	0.71
50	4.12	1.85	1.71	2.86	1.36	1.04	2.30	1.18	0.78
55	4.07	1.88	1.71	3.20	1.49	1.00	2.60	1.23	0.74
60	4.24	1.90	1.65	3.02	1.49	1.00	2.65	1.24	0.71
65	5.22	1.98	1.83	3.16	1.40	1.00	2.50	1.22	0.88
70	4.07	1.97	1.73	3.04	1.46	1.06	2.61	1.18	0.71
75	4.02	1.89	1.65	3.03	1.48	1.07	2.36	1.01	0.77
80	4.68	1.96	1.80	3.36	1.53	1.09	2.53	1.22	0.76
85	4.23	1.90	1.76	3.41	1.35	1.14	2.71	1.26	0.73
90	4.95	2.08	1.78	2.80	1.32	1.09	2.90	1.36	0.76
95	4.36	2.13	1.86	3.43	1.62	1.10	2.53	1.24	0.72
100	4.14	2.12	1.88	2.81	1.46	1.24	2.47	1.33	0.78
slope (x100)	0.680	0.679	0.289	0.832	0.393	0.296	1.011	0.413	0.159
p-values									
slope $= 0$	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
US_B=US_W		0.997			0.054			0.001	
US_W=EU		0.000			0.298			0.001	

Notes: This table shows numerical values of average mortality rates plotted in Figure 1, along with the value of the slopes of the fitted lines shown in Figure 1 and p-values of differences between slopes. Europe refers to population weighted average mortality rates across England, France, Germany, Netherlands, Norway, and Spain. For additional comments see the note of Figure 1 and the Materials and Methods section in the main paper.

Table S2: Mortality rates per 1,000 across poverty ventiles, age 5-19.

	1990			2005			2018		
	U.S. Black	U.S. White	Europe	U.S. Black	U.S. White	Europe	U.S. Black	U.S. White	Europe
Overall	0.63	0.41	0.30	0.41	0.31	0.19	0.36	0.24	0.13
Percentile									
5	0.38	0.29	0.26	0.28	0.23	0.17	0.24	0.19	0.12
10	0.54	0.31	0.27	0.34	0.25	0.16	0.24	0.19	0.12
15	0.59	0.34	0.28	0.32	0.25	0.17	0.26	0.20	0.12
20	0.52	0.35	0.28	0.38	0.28	0.17	0.35	0.23	0.12
25	0.45	0.36	0.30	0.33	0.28	0.19	0.30	0.24	0.13
30	0.59	0.37	0.30	0.37	0.29	0.18	0.30	0.22	0.11
35	0.52	0.40	0.28	0.37	0.30	0.18	0.35	0.23	0.13
40	0.62	0.40	0.30	0.40	0.30	0.18	0.31	0.24	0.12
45	0.49	0.37	0.30	0.43	0.33	0.19	0.34	0.24	0.12
50	0.53	0.40	0.29	0.42	0.30	0.20	0.33	0.24	0.14
55	0.70	0.42	0.29	0.44	0.33	0.19	0.42	0.27	0.13
60	0.68	0.39	0.29	0.47	0.32	0.20	0.45	0.26	0.13
65	0.84	0.41	0.31	0.45	0.30	0.20	0.47	0.26	0.15
70	0.73	0.49	0.32	0.42	0.34	0.21	0.38	0.24	0.13
75	0.73	0.47	0.31	0.44	0.32	0.19	0.36	0.21	0.13
80	0.84	0.43	0.31	0.44	0.35	0.20	0.41	0.27	0.12
85	0.68	0.46	0.30	0.49	0.33	0.22	0.41	0.26	0.11
90	0.74	0.46	0.29	0.46	0.32	0.20	0.48	0.28	0.13
95	0.74	0.49	0.32	0.48	0.39	0.22	0.41	0.27	0.13
100	0.68	0.54	0.31	0.46	0.37	0.22	0.42	0.27	0.12
slope (x100)	0.337	0.206	0.042	0.182	0.121	0.058	0.206	0.070	0.007
p-values									
slope $= 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.316
US_B=US_W		0.041			0.031			0.000	
US_W=EU		0.000			0.000			0.000	

Notes: This table shows numerical values of average mortality rates plotted in Figure 2, along with the value of the slopes of the fitted lines shown in Figure 2 and p-values of differences between slopes. Europe refers to population weighted average mortality rates across England, France, Germany, Netherlands, Norway, and Spain. For additional comments see the note of Figure 2 and the Materials and Methods section in the main paper.

Table S3: Mortality rates per 1,000 across poverty ventiles, age 20-64.

	1990			2005			2018		
	U.S. Black	U.S. White	Europe	U.S. Black	U.S. White	Europe	U.S. Black	U.S. White	Europe
Overall	7.79	4.12	3.67	5.90	3.51	2.82	4.92	3.56	2.28
Percentile									
5	5.20	3.17	3.22	3.56	2.47	2.35	3.16	2.46	1.79
10	6.09	3.58	3.47	4.26	2.69	2.59	3.33	2.71	2.04
15	6.45	3.54	3.48	4.43	2.91	2.57	3.58	2.77	2.02
20	7.19	3.65	3.55	4.83	2.98	2.59	4.14	3.07	2.02
25	6.85	3.63	3.46	5.24	3.03	2.60	4.15	3.07	2.10
30	7.55	3.76	3.64	5.13	3.22	2.70	4.32	3.19	2.12
35	7.48	3.85	3.56	5.48	3.31	2.66	4.11	3.20	2.07
40	7.41	4.10	3.52	6.01	3.38	2.68	4.98	3.55	2.19
45	7.05	4.06	3.65	6.25	3.67	2.75	4.37	3.55	2.21
50	7.43	4.00	3.52	5.71	3.59	2.75	4.61	3.66	2.26
55	7.86	4.44	3.63	6.29	3.59	2.87	5.54	3.78	2.37
60	8.08	4.20	3.67	6.44	3.74	2.86	5.66	3.63	2.36
65	9.60	4.23	3.67	6.83	3.65	2.83	5.87	3.86	2.45
70	8.09	4.36	3.70	6.26	3.76	3.01	5.30	3.67	2.39
75	8.33	4.40	3.67	6.59	3.86	2.88	5.49	3.27	2.47
80	8.77	4.32	3.81	6.75	3.99	2.97	5.66	3.99	2.38
85	8.72	4.40	3.83	7.15	3.81	3.00	5.55	3.97	2.48
90	9.52	4.89	4.22	6.60	3.77	3.28	6.23	4.52	2.70
95	9.06	4.89	4.06	7.44	4.35	3.29	6.11	4.53	2.62
100	9.00	5.00	4.04	6.82	4.42	3.20	6.21	4.74	2.57
slope (x100)	3.539	1.586	0.694	3.297	1.655	0.802	3.130	1.929	0.768
p-values									
slope $= 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
US_B=US_W		0.000			0.000			0.000	
US_W=EU		0.000			0.000			0.000	

Notes: This table shows numerical values of average mortality rates plotted in Figure 3, along with the value of the slopes of the fitted lines shown in Figure 3 and p-values of differences between slopes. Europe refers to population weighted average mortality rates across England, France, Germany, Netherlands, Norway, and Spain. For additional comments see the note of Figure 3 and the Materials and Methods section in the main paper.

Table S4: Mortality rates per 1,000 across poverty ventiles, age 65-79.

	1990			2005			2018		
	U.S. Black	U.S. White	Europe	U.S. Black	U.S. White	Europe	U.S. Black	U.S. White	Europe
Overall	42.99	31.43	31.47	33.73	25.45	23.42	28.22	22.02	18.17
Percentile									
5	37.07	29.47	29.35	27.36	22.95	20.34	22.60	18.12	15.41
10	41.26	31.68	30.50	28.92	23.09	21.83	23.46	18.77	16.71
15	39.05	29.63	30.18	30.22	24.30	21.74	24.60	19.38	16.73
20	41.41	29.74	29.85	31.68	24.15	21.73	26.15	20.19	16.56
25	41.97	30.15	30.50	31.93	23.54	22.06	26.08	19.90	17.10
30	43.19	28.54	31.41	30.97	24.85	22.36	26.78	20.83	17.31
35	43.34	30.02	31.19	32.13	24.49	22.45	27.02	20.77	17.11
40	40.48	31.42	30.67	33.64	23.46	22.80	28.46	21.44	17.84
45	42.16	30.87	31.54	34.81	26.08	23.24	25.75	20.78	17.82
50	44.64	31.06	30.84	33.59	25.38	23.02	26.62	22.60	18.08
55	45.07	31.74	31.67	35.59	25.29	23.78	30.00	22.20	18.84
60	44.77	32.01	31.82	36.18	26.76	23.78	29.93	21.86	18.66
65	45.94	32.88	30.98	35.72	25.71	23.64	31.35	23.65	18.82
70	42.75	31.69	31.72	34.77	26.18	24.09	30.09	23.04	18.69
75	44.29	32.14	31.58	36.61	26.85	24.03	30.17	21.22	19.28
80	43.82	32.51	33.05	37.19	27.55	24.54	30.11	23.48	19.20
85	45.66	32.01	32.05	36.86	26.65	24.82	29.58	23.62	19.21
90	44.68	33.52	34.28	34.80	25.84	26.08	32.65	25.82	20.22
95	43.95	33.52	33.33	37.39	28.27	26.15	30.90	26.07	19.75
100	44.24	34.00	32.94	34.32	27.69	25.89	32.15	26.57	20.01
slope (x100)	5.766	4.309	3.654	8.357	4.827	5.183	9.023	7.496	4.298
p-values									
slope $= 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
US_B=US_W		0.332			0.026			0.138	
US_W=EU		0.445			0.571			0.000	

Notes: This table shows numerical values of average mortality rates plotted in Figure 4, along with the value of the slopes of the fitted lines shown in Figure 4 and p-values of differences between slopes. Europe refers to population weighted average mortality rates across England, France, Germany, Netherlands, Norway, and Spain. For additional comments see the note of Figure 4 and the Materials and Methods section in the main paper.

Table S5: Cause-specific contributions to life expectancy gains and to the reduction of the Black-White life expectancy gap, 1990-2018

	Contributions to life LE (in yrs)		% contribution to
Cause	Black	White	LE gap reduction
Fetal/infancy origin	0.552	0.234	11.05
Unintentional injury	0.287	0.267	0.71
Homicide	0.426	0.067	12.51
Deaths of despair	-0.144	-0.608	16.18
HIV	0.475	0.191	9.89
Flu/Pneu/Resp	0.252	0.103	5.17
Septicemia	0.008	-0.037	1.59
Cardiovascular	2.678	2.662	0.54
Cancer	1.438	0.979	15.96
Parkin./Alzheimer	-0.124	-0.202	2.72
Undefined disease	0.103	0.001	3.57
Other	0.224	-0.353	20.11

Notes: This table shows the hypothetical change in Black and White life expectancy between 1990 and 2018 had only the selected group of cause obtained its 2018 mortality rate. ICD codes for all causes included in each group are listed in Table S6. Note that these contributions only reflect the causes individual contributions and not the interaction terms with other causes, which are positive in most cases (hence, the contributions tend to understate the overall contributions). The third column calculates the percent of the reduction in the Black-White life expectancy gap contributed to each cause.

Table S6: ICD-9 and ICD-10 codes included in different groups of causes of death

Group	ICD 9 (1990)	ICD 10 (2018)
Fetal/infancy origin:		
	Congenital anomalies (740-759), Certain conditions originating in the perinatal period (760-779), Sudden infant death syndrome (798)	Certain conditions originating in the perinatal period (P00-P96), Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99), Sudden infant death syndrome (R95)
Unintentional injury:		
	Motor vehicle accidents (E810-E825), All other accidents and adverse effects (E800-E807, E826-E949)	Motor vehicle accidents (V02-V04, V09.0, V12-V14, V19.0-V19.2, V19.4-V19.6, V20-V79, V80.3-V80.5, V81.0-V81.1, V82.0-V82.1, V83-V86, V87.0-V87.8, V88.0-V88.8, V89.0, V89.2), All other accidents and adverse effects (V01, V05-V06, V09.1, V09.3-V09.9, V10-V11, V15-V18, V19.3, V19.8-V19.9, V80.0-V80.2, V80.6-V80.9, V81.2-V81.9, V82.2-V82.9, V87.9, V88.9, V89.1, V89.3, V89.9, V90-X39, X46-X59, Y40-Y86, Y88)
Homicide:		
	Homicide and legal intervention (E960-E978)	Assault (homicide) (*U01-*U02,X85-Y09,Y87.1)
Deaths of despair:		
	Liver disease (571), Drug Poisoning (E850-E860, E980), Suicide (E950 -E959)	Chronic liver disease and cirrhosis (K70,K73-K74), Drug Poisoning (X40-X45), Intentional self-harm (suicide) (X60-X84,Y87.0)
HIV:		
	Other viral diseases (042-044)	Human immunodeficiency virus (HIV) disease (B20-B24)
Flu/Pneu/Resp:		
	Acute upper respiratory infections (460-465), Bronchitis and bronchiolitis (466, 490-491), Pneumonia and influenza (480-487), Pneumonia (480-486), Influenza (487), Remainder of diseases of respiratory system (470-478, 492-519)	Influenza and pneumonia (J10-J18), Other acute lower respiratory infections (J20-J22), Chronic lower respiratory diseases (J40-J47), Pneumoconioses and chemical effects (J60-J66,J68), Pneumonitis due to solids and liquids (J69), Other diseases of respiratory system (J00-J06,J30-J39,J67,J70-J98)
Septicemia:		
	Septicemia (038)	Septicemia (A40-A41)
Cardiovascular:		
	Major cardiovascular diseases (390-448)	Major cardiovascular diseases (I00-I78)
Cancer:		
	Malignant neoplasms, including neoplasms of lymphatic and hematopoietic tissues (140-208)	Malignant neoplasms (C00-C97)
Parkinson / Alzheimer:		
	Parkinson's disease (332), Alzheimer's Disease (331)	Parkinson's disease (G20-G21), Alzheimer's disease (G30)
Undefined disease:		
	Symptoms, signs, and ill-defined conditions (780-797, 799)	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (excluding Sudden infant death syndrome) (R00-R94,R96-R99)

Table S7: Percent contributions to the reduction of the Black-White life expectancy gap, by age, 1990-2018.

	1990-2000	2000-2018	overall 1990-2018
Age 0-4	19.60	9.83	14.61
Age 5-19	5.87	-1.40	1.71
Age 20-49	59.06	46.04	53.50
Age 50-64	18.71	34.85	27.92
Age 65-79	-1.98	10.76	3.54
Age 80 plus	-1.25	-0.08	-1.30
Sum	100	100	100

Notes: This table shows the contributions to the reduction of the Black-White life expectancy gap due to mortality improvements at different ages, in percent of the total gap reduction. These percent contributions are calculated by constructing race-specific hypothetical life expectancies that only let mortality rates in the given age group change between 1990 and 2018. Additional information is provided in Supplementary Material.

Table S8: Contributions of avoidable mortality to life expectancy gains and to the reduction of the Black-White life expectancy gap, 1990-2018

	Contributions to life LE (in yrs)		% contribution to
Cause	Black	White	LE gap reduction
Amenable to Medical Care	1.291	0.390	23.25
Amenable to Interventions	1.069	0.653	10.72
Ischaemic Heart Disease	0.595	0.469	3.26
HIV/AIDS	0.496	0.198	7.69

Notes: This table shows the hypothetical change in Black and White life expectancy between 1990 and 2018 had only the selected type of avoidable mortality obtained its 2018 mortality rate. ICD codes for all causes included in each avoidable mortality group are listed in Table S9. Note that these contributions only reflect the causes individual contributions and not the interaction terms with other causes, which are positive in most cases (hence, the contributions tend to understate the overall contributions). The third column calculates the percent of the reduction in the Black-White life expectancy gap contributed to each cause.

Table S9: Avoidable mortality, replicated from Macinko and Elo (44), Table A1

ICD 9 (1990)	ICD 10 (2018)
Amenable to Medical Care Treatment and Prevention: Causes: Intestinal infections, Tuberculosis, Other infectious (Diphtheria, Tetanus, Poliomyelitis, Septicaemia, Whooping Cough), Measles, Malignant neoplasm of colon and rectum, skin, breast; cervix uteri, testis; Hodgkinas disease,Leukaemia, Diseases of the thyroid, Diabetes mellitus, Epilepsy, Chronic rheumatic heartdisease, Hypertensive disease, Cerebrovascular disease, All respiratory diseases (excluding pneumonia/influenza),Influenza and Pneumonia, Peptic ulcer, Appendicitis, Abdominal hernia, Cholelithiasis and cholecystitis, Nephritis and nephrosis, Benign prostatic hyperplasia, Maternal deaths, Congenital cardiovascular anomalies, Perinatal deaths (excl. stillbirths), Misadventures to patients during surgical/medical care.	
001-009, 010--018, 137, 032, 033, 037, 038, 045, 55, 153--154, 173, 174, 180, 186, 201, 204--208, 240--246, 250, 345, 393--398, 401--405, 430--438, 460--479, 488--519, 487, 480--486, 531--533, 540--543, 550--553, 571, 574--575.1, 580--589, 600, 630--676, 745--747, 760--779, E870--E876, E878--E879	A00--A09, A15--A19, B90, A36, A35, A37, A40--A41, A80, B05, C18--C21, C44, C50, C53, C62, C81, C91--C95, E00--E07, E10--E14, G40--G41, I05--I09, I10--I13, I15, I60--I69, J00--J09,J20--J99, J10--J11, J12--J18, K25--K27, K35--K38, K40--K46, K80--K81, N00--N07, N17--N19, N25--N27, N40, O00--O99, Q20--Q28, P00--P96, A33, Y60--Y69, Y83--Y84
Amenable to Policy and Behavioural Interventions: Causes: Malignant neoplasm of the trachea, bronchus, and lung; Road traffic injuries; Homicide	
162, E810--E819, E960--E969	C33, V01, V03, V06, V09, V13, V15, V19, V20, V25--V29, V40--49, V80, V82, V87--89, X85--Y09, K70
Ischaemic Heart Disease (IHD): Causes considered amenable to both policy/behavioural interventions and medical care treatment	
410--414, 429.2	I20--I25
HIV/AIDS: Deaths from early infections considered not avoidable; amenable to policy/behavioural interventions until 1996, and amenable to medical care and policy/behaviour post-HAART (1996--2005).	
042--044	B20--B24

B Additional life expectancy decomposition details

Calculating age, cause, and ventile-specific contributions to Black and White life expectancy. We start with the description of the calculation of age contributions before turning to contributions by cause and ventile (the calculation of contributions by cause or ventile is equivalent and we will explain the method using cause contributions as the example). The age contribution calculation analyzes how much life expectancy was gained (or lost) due to mortality changes solely at a single year of age, keeping mortality at all other ages at their 1990 value. Specifically, we calculate a hypothetical life expectancy at age a using the 2018 mortality rate at age a while all other ages enter with their 1990 mortality rates. The difference between the resulting hypothetical life expectancy for age a and the actual 1990 life expectancy is then the life expectancy change contributed to the mortality change at age a. This procedure is repeated at each individual year of age.[1]

For cause (or ventile) contributions, we calculate hypothetical life expectancies the same way, letting only one cause category change while mortality rates for all other categories are kept constant. However, this calculation is somewhat complicated by the fact that the construction of life expectancy requires mortality rates at each single year of age while cause- or ventile-specific mortality encompasses deaths at all ages. We therefore translate cause-specific mortality rates in 2018 at each single year of age to hypothetical cause-by-age mortality counts in 1990 (multiplying the 2018 cause-by-age rate with the 1990 population at each age), then replace the actual 1990 cause-by-age mortality counts with these hypothetical counts across all ages, before finally collapsing these hypothetical age-specific mortality rates into hypothetical life expectancy. The

[1] Since we do not observe mortality rate above age 84, we use a Gompertz extrapolation for mortality at age 85-110. In particular, we fit a linear regression line through the logarithm of mortality between age 60 and 84 and then predict this line forward to age 110. Fitting the regression line over a longer or shorter age window (e.g. age 75-84 instead of age 60-84) only marginally changes life expectancy estimates.

difference between the resulting hypothetical life expectancy for cause c and the actual 1990 life expectancy is then the life expectancy change contributed to the mortality change in cause c.

Calculating the age, cause, and ventile-specific percent contributions to the reduction in the Black-White life expectancy gap. We use the age-, cause-, and ventile-specific contributions to changes in Black and White life expectancy described above to calculate each categories contribution to the reduction in the Black-White life expectancy gap the following way. First, we subtract the contributions to White life expectancy from the contributions to Black life expectancy for each category, which gives us a measure of the change in the life expectancy gap due to this category. We then express this change as the percent relative to the sum of changes across all categories for a given factor.

C Data Availability

Czech Republic

Years: 1994-2018
Area type: District
Area ranking measure: Mean income
Death data source: Mortality register data can be purchased from the Czech Statistical Office. Data must be requested following the instructions at this link https://www.czso.cz/csu/czso/zadosti-o-poskytnuti-pristupu-k-duvernym-statistickym-udajum-pro-ucely-vedeckeho-vyzkumu
Population data source: Population counts by gender and age can be requested and purchased from the Czech Statistical Office following the procedure at this link https://www.czso.cz/csu/czso/zadosti-o-poskytnuti-pristupu-k-duvernym-statistickym-udajum-pro-ucely-vedeckeho-vyzkumu
Ranking measure data source: Data taken from the annual publication titled "Okresy Ceske republiky v roce" published by the Czech Statistical Office (available as hard copies for the least recent years) and own calculation using data at the regional level for most recent years (Data available from https://vdb.czso.cz/vdbvo2/faces/en/index.jsf?page=statistiky&katalog=31799

England

Years: 1992-2017
Area type: Local authority Area ranking measure: Index of multiple deprivation
Death data source: ONS Vital Statistics (lookup necessary to convert from LSOA to Local Authority)
https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/009628birthsanddeathsbylowersuperoutputarealsoaenglandandwales1991to1992to2016to2017
Population data source: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland
Ranking measure data source: English indices of deprivation from GOV.UK
https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

Finland

Years: 1990-2016
Area type: Municipality
Area ranking measure: Poverty rate
Death data source: Cause-of-death records, Statistics Finland
https://www.stat.fi/meta/til/ksyyt_en.html
Population data source: FLEED and FOLK registers, Statistics Finland

https://www.stat.fi/tup/mikroaineistot/aineistot_en.html
Ranking measure data source: FLEED and FOLK registers (demographics, family links and labor market information of all Finnish residents from 1988-2016/2018) Statistics Finland
https://www.stat.fi/tup/mikroaineistot/aineistot_en.html

France

Years: 1990-2018
Area type: Departement
Area ranking measure: Poverty rate
Death data source: https://www.cepidc.inserm.fr/
Population data source: https://www.insee.fr/fr/accueil
Ranking measure data source: https://www.insee.fr/fr/accueil

Germany

Years: 1990, 1995-2016
Area type: District (Kreise/Landkreise)
Area ranking measure: Per capita disposable income
Death data source: Federal Statistical Office (account needed, then code 12613-93-01-4).
Data for 1990 not publicly available, requiring data requested
www.regionalstatistik.de
Population data source: Federal Statistical Office (account needed, then code 12411-02-03-4).
Data for 1990 not publicly available, requiring data requested
www.regionalstatistik.de
Population counts pre-2011 adjusted according to estimates by Kluesener et al. (2018)
data appendix: https://comparativepopulationstudies.de/index.php/CPoS
/article/view/251
Ranking measure data source: Federal Statistical Office (and Statistical Offices of the Laender
https://www.statistikportal.de/de/vgrdl/ergebnisse-kreisebene/einkommen-kreise

Netherlands

Years: 1990-2018
Area type: Municipality
Area ranking measure: Poverty rate
Death data source: 1990-1994 death counts obtained from Statline
https://opendata.cbs.nl/statline/#/CBS/nl/dataset/03747/table?ts=1613379024732
1995 onwards are based on own calculations using non-public microdata on death registries from Statistics Netherlands. For information on data applications, visit
https://www.cbs.nl/en-gb/onze-diensten/customised-services-microdata/microdata-conducting-your-own-research

Population data source: Non-public microdata on personal records database from Statistics Netherlands. For information on data applications, visit https://www.cbs.nl/en-gb/onze-diensten/customised-services-microdata/microdata-conducting-your-own-research.
Ranking measure data source: Poverty rates for 1990 can be accessed at https://opendata.cbs.nl/#/CBS/nl/dataset/70050ned/table?dl=4C0D7
For the years 2005 & 2016, poverty measure based on regional income distribution data using (non-public) household level tax registries data from Statistics Netherlands. For information on data applications, visit https://www.cbs.nl/en-gb/onze-diensten/customised-services-microdata/microdata-conducting-your-own-research.

Norway

Years: 1990-2018
Area type: Municipality
Area ranking measure: Poverty rate
Death data source: Cause of Death Registry
https://www.fhi.no/en/hn/health-registries/cause-of-death-registry/
and medical Birth Registry
https://www.fhi.no/en/hn/health-registries/medical-birth-registry-of-norway/
Death data access: Both registries are maintained by the Norwegian Public Health Institute (FHI) and are restricted to researchers with data use agreements only. Interested researchers can apply for data access from the maintainer by following their data application guide at https://www.fhi.no/en/more/access-to-data/applying-for-access-to-data/.
Population data source: Population Registry at Statistics Norway (SSB)
https://www.ssb.no/en/omssb/tjenester-og-verktoy/data-til-forskning/befolkning
Population data access: Data are restricted to researchers with data use agreements only. Interested researchers outside of Norway can apply for microdata access following the steps described at the SSB website
https://www.ssb.no/en/omssb/tjenester-og-verktoy/data-til-forskning.
Ranking measure data source: Statistics Norway
https://www.ssb.no/en/omssb/tjenester-og-verktoy/data-til-forskning/inntekt
Ranking measure data access: Data are restricted to researchers with data use agreements only. Interested researchers from outside of Norway can apply for microdata access following the steps described at the SSB website
https://www.ssb.no/en/omssb/tjenester-og-verktoy/data-til-forskning.

Portugal

Years: 1990-2018
Area type: Municipality
Area ranking measure: Deprivation index based on illiteracy, unemployment and hous-

ing conditions (three indicators are normalized and added) Death data source: Instituto Nacional de Estatistica -Mortality data
https://ine.pt/xportal/xmain?xpid=INE&xpgid=ine_main
Population data source: Instituto Nacional de Estatistica -Census and Estimated data
https://ine.pt/xportal/xmain?xpid=INE&xpgid=ine_main
Ranking measure data source: Instituto Nacional de Estatistica -Census data
https://ine.pt/xportal/xmain?xpid=INE&xpgid=ine_main

Spain

Years: 1990-2018
Area type: Municipality
Area ranking measure: Mean income
Death data source: Death Statistics microdata (INE)
https://ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736177008&menu=ultiDatos&idp=1254735573002
Population data source: INE Continuous Register Statistics, results by municipalities (1996-2018)
https://ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736177012&menu=ultiDatos&idp=1254734710990
and INE 1991 Population Census
https://ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736176992&menu=ultiDatos&idp=1254735572981
Ranking measure data source: Household Income Distribution Map
https://www.ine.es/en/experimental/atlas/experimental_atlas_en.htm

United States

Years: 1990-2018
Area type: county
Area ranking measure: poverty rate
Death data source: US Vital Statistics
https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm
Population data source: US Census bridged race
https://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm#vintage2019
Ranking measure data source: US Census and ACS
accessed at www.socialexplorer.org