Cream Skimming by Health Care Providers and Inequality in Health Care Access: Evidence from a Randomized Field Experiment*

Anna Werbeck[a] **Ansgar Wübker**[b] **Nicolas R. Ziebarth**[c]
RWI Essen RWI Essen Cornell University

DRAFT – December 2018

Abstract

We show in a randomized field experiment that specialists cream-skim patients by their profitability. In the German two-tier system, reimbursement rates for both the publicly and privately insured are centrally determined and fixed but more than twice as high for the privately insured. In addition, the privately insured are on average healthier, have higher incomes, and their reimbursement is 100% fee-for-service. The same hypothetical patient called almost one thousand outpatient private practices in 36 representative German counties and attempted to make appointments for allergy tests, hearing tests and gastroscopies. Each telephone call followed a specific protocol where we randomized and explicitly stated the insurance status of the caller and called each practice twice in time intervals of at least two weeks. Our findings show, first, that privately insured patients are a statistically significant 10 percent more likely to being offered an appointment. Second, we find statistically significant differences in wait times, conditional on being offered an appointment: public insured patients had to wait more than twice as long, or 13 more weekdays for an appointment. The findings provide randomized evidence from the field showing that specialists cream-skim patients based on expected profitability. This implies that systematic insurance-related differences in reimbursement rates lead to inequality in health care access.

JEL Codes: I10, I11, I18

Key words: reimbursement rate differences, cherry picking, discrimination, health inequality, gastroscopy, audiometry, allergy test, allergist, otorhinolaryngologist, gastroenterologist

*The research reported in this paper is not the result of a for-pay consulting relationship. Our employers do not have a financial interest in the topic of the paper which might constitute a conflict of interest. The field experiment was carried out at RWI Essen, a German publicly funded research institute, which does not have IRB approvals. All ethics guidelines by the German Science Foundation were followed; the guidelines for good scientific practice are here http://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/. The protocols used minimized the time need to inquire about appointments and no actual appointment was made at the end of each call.

[a]RWI – Leibniz-Institut für Wirtschaftsforschung, Hohenzollernstraße 1-3, 45128 Essen, Germany, Tel.: (0201) 81 49-515, e-mail: anna.werbeck@rwi-essen.de
[b]RWI – Leibniz-Institut für Wirtschaftsforschung, Hohenzollernstraße 1-3, 45128 Essen, Leibniz Science Campus Ruhr and RUB, Germany, Tel.: (0201) 81 49-242, Fax: (0201) 81 49-200, Germany, e-mail: ansgar.wuebker@rwi-essen.de
[c]Cornell University, Policy Analysis and Management (PAM), 426 Kennedy Hall, Ithaca, NY 14853, DIW Berlin, IZA Bonn, and RWI Essen, e-mail: nrz2@cornell.edu, Phone: +1-(607) 255-1180, Fax: +1-(607) 255-4071.

1. Introduction

Access barriers to health care are a major performance indicator in comparative health care system analyses (Siciliani and Hurst 2005; Sicilinani and Verzulli 2009; Jones et al. 2011, Viberg et al. 2013). The nonpartisan *Commonwealth Fund* uses wait times as the main measure of health care access in their “Timeliness to Care” category, where the United States ranks 5th among 12 countries in a 2014 survey (Commonwealth Fund 2014). At the same time, wait times have long been cited by critics as proof of mediocre outcomes of single-payer systems (cf. Mackillop et al. 1995). Indeed, wait times for specialists are significantly longer in Canada as compared to the largely private system in the United States. In Canada, 30% of patients have to wait more than 2 months for a specialist appointment as compared to just 6% in the United States (Commonwealth Fund 2017).

Another major performance dimension to rate health care systems is *equity* in access to health care (e.g. van Doorslaer et al. 2000)—a dimension in which the United States has consistently ranked last among the 12 OECD countries benchmarked by the *Commonwealth Fund* (2014, 2017). Critics of the largely private U.S. system have been pointing to the large inequalities within the U.S. system, even among those who have insurance. The reason is that thousands of private managed care insurers individually negotiate reimbursement rates with networks of providers. What’s more, the public Medicaid system for the poor pays significantly lower rates than private insurers or the single-payer Medicare system for the old (CMS, 2018a). One consequence of this fragmented system with major differences in provider reimbursement rates would be that providers would cream-skim and discriminate against the poor and sick although those population subgroups needed health care the most (Reinhardt, 2011). Although plenty of anecdotal and descriptive evidence exists, it is difficult to prove in a causal framework that health care providers discriminate against Medicaid enrollees and cherry-pick the privately insured *because* they are more profitable.

This paper uses a randomized field experiment in a well-suited institutional setting outside the U.S. to show that specialists cream-skim the more profitable privately insured patients and discriminate against the publicly insured because of lower reimbursement rates. Germany has a multi-payer two-tier system where the majority is mandatorily insured under the public system and one of the 110 non-profit “sickness funds” (Schmitz and Ziebarth, 2017). In the public system, provider reimbursement rates are centrally negotiated and do not vary across sickness funds. Moreover, cost-sharing is standardized and invariant across sickness funds, while provider networks are non-existent and enrollees can freely choose their provider (Buennings et al., 2018). The situation is very similar for the 9 million privately insured patients: reimbursement rates are uniform across the 44 private insurers and provider networks do not exist; insurers mostly process claims (Atal et al., 2018). However, reimbursement rates for the privately insured are on average twice as high than for publicly insured (Walendzik et al., 2008).

In our field experiment, we selected a total of 36 representative counties (both urban and rural) and called a total of 991 outpatient specialists to ask for wait times and make appointments for elective medical treatments. We called each practice *twice*, once as a fictitious privately insured new patient and once as a fictitious publicly insured new patient, randomizing the insurance status over the two calls. In other words, the same test person called each private outpatient practice twice following the exact same protocol, thereby ensuring balanced covariates by construction. This allows us to carry out straightforward statistical tests to assess whether extensive and intensive access barriers to health care differ significantly by insurance status.

Our findings show that access to the health care system differs significantly between the privately and publicly insured, both on the extensive and intensive margin. The likelihood of being offered an appointment is a highly significant 10 percent larger for privately insured patients. Moreover, conditional on being offered an appointment, the wait times for publicly insured patients are more than twice as long, on average 13 weekdays longer.

This paper makes important contributions to the literature. Although the literature on physician behavior and treatment styles is rich and traditional in economics (e.g., Clemens and Gottlieb 2014, also see Section 2), the causal effects literature on how providers discriminate against less profitable patients is less diverse. We contribute to a better understanding of the role of varying reimbursement rates in achieving more equitable access to the health care system for disadvantaged population groups. For example, for the Unites States, Cooper et al. (2018) document that reimbursement rates just among the privately insured could vary by a factor of 10 within cities and more than 20 across the U.S. Moreover, ours is one of the first real-world studies that leverage a randomized field experiment, calls almost one thousand providers twice, and randomizes the insurance status in a non-emergency outpatient setting. Unlike the few existing studies outside the field of economics, to minimize selection concerns, our caller routinely informs providers about the insurance status and inquires wait times without further framing. Moreover, not conditioning on emergency care but routine specialist visits for gastroscopies and

Our findings yield important insights into the driving forces of inequality in health care access. They suggest that uniform reimbursement rates (or reimbursements rates that are higher for disadvantaged population groups) could help mitigate inequality in health care access and align economic incentives with medical needs and priorities.

The next section describes the literature on this topic, followed by a discussion of the institutional setting in Germany. Section 4 explains the setup of our field experiment and Section 5 the data. After that we outline the statistical approach of this study before discussing the findings. Section 8 concludes.

2. Previous literature

This paper relates to various rich literature strands in economics. However, maybe surprisingly, while there exist many descriptive papers on socio-economic difference in health care access, the causal effects literature on discrimination in the health care sector is comparatively thin.

In contrast, the economics literature has a long tradition of investigating theoretically and empirically the role of physicians as (imperfect) agents of their patients, see McGuire, 2000 for an excellent overview. In his seminal paper, Arrow (1963) was one of the first to discuss the role of uncertainty and physician behavior in health care markets.

Moreover, a robust research strand in economics has investigated theoretically and empirically how physician behavior and productivity changes in response to the reimbursement method, in outpatient as well as in inpatient settings (Ellis and McGuire, 1986; Nicholson et al., 2008). For ethical reasons, field experiments are almost impossible to implement to study actual treatment behavior, which is why researchers have studied hypothetical physician behavior in the lab; see, for example, Brosig-Koch et al. (2017) for lab experiments in Germany. In one of the few real-world causal effects studies leveraging relative price changes in the Medicare outpatient market, Clemens and Gottlieb (2014) demonstrate that higher relative reimbursement rates increase treatments, especially for elective procedures.

Absent price variation in single-payer markets, implicit rationing of medical care through wait times is another popular topic of inquiry for economists (e.g. Lindsay and Feigenbaum, 1984). Cullis et al. (2000) provide a comprehensive overview of the topic. In addition to theoretical analyses (Siciliani, 2006; Gravelle and Siciliani, 2008; Felder, 2008), especially correlations between wait times and socio-economic status have drawn researchers' interest. For example, Monstad et al. (2014) find a negative statistical association between income and wait times as well as education and wait times in Norway. Laudicella et al. (2012) show that the same correlations exist in England and that they hold up over the entire wait time distribution.

The impact of insurance status on wait times is a topic of inquiry that is highly relevant in countries with co-existing insurance systems that pay providers differently, such as the U.S., Switzerland or Germany. In the U.S., it is highly policy relevant as the sickest, poorest and least educated members of society are enrolled in Medicaid, which pays by far the lowest reimbursement rates of all insurance systems. Several papers have studied the association between insurance status and wait times of patients (Roll et al., 2012; Sundmacher and Kopetsch, 2013; Ramos et al., 2018). All of them find that patients whose insurer pays lower rates have to wait longer for an appointment. However, because enrollment in, e.g. Medicaid, is correlated with specific socio-demographics as well as Managed Care elements such as gatekeeping or capitation, it remains challenging to identify *causal effects* of insurance status on discrimination through providers. Similar arguments hold for the case of Germany.

To our knowledge, there exist three studies (two outside the field of economics) which are similar in design to ours and called providers at least twice with the insurance status randomized. First, in 2002/2003, Asplin et al. (2005) called ambulatory clinics in 9 U.S. cities twice and randomized the insurances status of the caller. They find that a higher share of privately insured patients was offered an urgent ambulatory follow-up visit within a week (i.e., they only requested appointments within a week). Second, Kuchinke et al. (2009) requested appointments at acute care hospitals in Germany and find that the privately insured get appointments 1.6 days faster than the publicly insured.[1] However, this difference in wait times was only estimated conditional on the hospital *inquiring* about the insurance status (only 25% did).[2] Third, Heinrich et al. (2018) evaluate a 2015 reform that intended to reduce wait times for the publicly insured in Germany. They compare data from 2014 to data from 2016 but do not find evidence that the reform reduced wait times.

In constrast to these studies, ours calls each practice twice and randomizes the insurance status of the caller. Moreover, the same person called all practices, followed a strict protocol, and always indicated the insurance status during the call. In addition, our randomzied

[1] In a third study outside of economics, Lüngen at al. (2008) also find that privately insured patients were offered faster appointments than publicly insured patients. However, they only called each practice *once* and the "inclusion rates" differed by insurance status; the authors do not show whether practice covariates were balanced and whether the randomization was successful.

[2] In a spin-off paper, Schwierz et al. (2011) investigate effect heterogeneity and differentiate the findings by the financial soundness of the hospital.

experiment uses a contemporaneous setting and took place over the course of one year between 2017 and 2018. Finally, we called almost one thousand practices, more than any other study, and focus on elective care among outpatient specialists. The focus on ambulatory care in a nonurgent setting reflects the regular day-to-day access barriers and inequalities in health care access much better than eliciting access barriers in medical emergency situations.

3. The German Health Care System

Germany has a two-tier health insurance system with a co-existing multi-payer public and an individual private market. Ninety percent of the population are covered by the public tier and one of the 110 non-profit sickness funds (GKV Spitzenverband 2018). They pay income-dependent contribution rates for a standardized benefit package with very little cost-sharing. However, for historical reasons, selected population subgroups have the right to leave the public system permanently and fully insure their health risks on an individual long-term health insurance market with relatively little regulation. As a result, applicants can choose between thousands of plans but also face experience rating (when signing their first individual contract). Schmitz and Ziebarth (2017), Pilny et al. (2017), and Buennings et al. (2018) provide more details on the overall structure of the German health insurance market. Atal et al. (2018) provide additional specific details of the private market.

Reimbursement Rates in Statutory Health Insurance (SHI)

In SHI, in the outpatient sector, primary care physicians and specialists are members of and sign contracts with the state-level Regional Association of Statutory Health Insurance Physicians, ASHIP (KBV 2018a). There are 17 ASHIPs, who are responsible for the provision of health care services in their region. These ASHIPs all have contracts with the 110 sickness funds who pay out a "total reimbursement sum'' (*Gesamtvergütung*) to each of these 17 ASHIPs who, in turn, reimburse their member physicians on a quarterly basis.

In SHI, the so-called "Unified Quantification Scale'' (*Einheitlicher Bewertungsmaßstab, EBM*) lists services that can be reimbursed. The existence of the EBM is justified by the German Social Insurance Law (§ 87f. SGB V, KBV 2018b). The EBM assigns a point value for each health care service, similar to the Relative Value Units (RVU) to outpatient providers in Medicare in the U.S. (CMS, 2018a). The relative point values intend to represent resource use for each service to guarantee according compensation.

As in Medicare, the point values are then converted into monetary reimbursement amounts by defining annual values per point. For example, in 2018, the value per point is 10.654 euro cent (BMG, 2018)[3]. For a colonoscopy for preventive reasons, including visits to prepare and inform the patient, the EBM lists 1945 points under their schedule of fees position ("Gebührenordnungsposition") 01741 (KBV, 2018b

[3] Geographic adjustment factors take differences in regional living costs into account.

); then the total basic compensation for such as colonoscopy would be €207.23.[4] In comparison to this, for the state-level Medicaid insurance for low-income populations, Halpern et al. (2014) report reimbursement rates between $83.94 in New York and $598.20 in Alaska for a colonoscopy.

However, the SHI system can also impose budget caps. (Since 2012, they can but do not have to be imposed by the sickness funds in cooperation with the ASHIPs, see Simon, 2017.[5]) When physicians provide more services than allocated by the "standardized service volume'' (*Regelleistungsvolumen" – RLV*), which is defined by the sum of last quarter's services and average in the field, the point value decreases (§87b SGB V). Primarily regions with lack of physicians do not impose these budget caps.

Reimbursement Rates in Private Health Insurance (PHI)

In PHI, the physician has a contractual relationship with the patient. Patients have to pay providers first (after receiving an invoice), and then submit their claim to the insurer to get reimbursed.

In PHI, the "fee schedule for physicians" (*Gebührenordnung für Ärzte, GOÄ*) lists all services that can be reimbursed along with their baseline prices. As in the case of the SHI, each medical service has a specific number and point value; the latter expresses the relative resource

---

[4] Interestingly, the reimbursement rates for colonoscopy in the U.S. under Medicare are similar. Under CPT code 45380 "Colonoscopy and biopsy", the Medicare feel schedule lists a reimbursement of $212.70 (CMS, 2018b).

[5] E.g. North Rhine still imposes budget caps (KVNO 2018).

utilization for the treatment. Point values are multiplied with a fixed value of 5.82873 euro cent to obtain the baseline reimbursement rate.

Depending on the complexity of the treatment and the time spent on its provision, the physician has then the freedom to multiply the baseline rate with “leverage factors” between 1.15 for laboratory services and 2.3 for personal services. In specific cases, a leverage factor of 3.5 can be applied for personal services (§5 II-IV GOÄ). Also, the physician can perform and charge treatments not listed in the GOÄ, taking the prices of similar treatments as a reference (Simon, 2017). Overall, the GOÄ is a classic fee-for-service schedule without any budget caps or cost containment elements. For example, a standard colonoscopy is listed as number 687 with 1500 points and a baseline value of €87.43 (GOÄ, 2018).

Comparison of the SHI and PHI Reimbursement

A direct comparison of the SHI and the PHI reimbursement scheme is difficult. First, the services listed usually do not exactly correspond. Second, the SHI schedule is closer to a bundled payment schedule and reimbursement rates include consultations and follow-up visits. In the PHI, physicians typically charge every single service separately under a pure fee-for-service schedule. Third, importantly, the GOÄ has no budget cap or cost containment element whatsoever. Moreover, the EBM has been constantly developed and changed, whereas the GOÄ has not changed since 1996.

Walendzik et al. (2008) analyze and compare differences in the billing amounts for the same treatments under SHI and PHI. They compare data from the largest German sickness fund with more than 10million enrollees (“Techniker Krankenkasse” – TK). For the same services, they find that providers charge 2.28 times higher reimbursement rates for privately as compared to publicly insured patients. When researching the amounts that providers can charge for the medical examinations that we inquire (KBV 2017, 2018b, GOÄ 2018b-d), we find SHI/PHI reimbursement rates of €48.80 vs. €183.60 for allergy tests, €15.66 vs. €23.02 for hearing tests and €88.06 vs. €163.20 for an upper gastrointestinal endoscopy, implying that reimbursement rates for privately insured are 1.34 to 3.8 times higher.

4. The Experiment

Selection of Counties

Before determining which outpatient providers to include in the experiment, we selected a set of counties that are jointly approximately representative for Germany. In choosing the treatment counties among the 401 existing German counties, we considered the following three indicators: household income per capita, area in km^2, and the population (BBSR, 2018; Destatis, 2018a, b). Appendix B describes in detail how we selected the counties.

[Insert Figure 1 about here]

Figure 1 shows Germany with its 401 counties; we included the dark gray shaded counties in the field experiment. As seen, the geographic distribution of all 36 counties is relatively even across all 16 German states as well as East and West Germany.

Comparing the monthly household income per capita of the 36 counties to the monthly household income for the whole of Germany, we only find a minor difference of €30 (€1,723 vs. €1,753). Also, the physician density per 100,000 population is almost identical when comparing the 36 counties to Germany as a whole (174 vs. 168 physicians per 100,000 population, see Versorgungsatlas, 2018).

Selection of Outpatient Specialists and Treatments

Next, for the 36 counties, we selected the outpatient specialists and requested treatments. We used google maps along with the websites of the three major telephone books in Germany: "The Telephone Book", "Yellow Pages" and "The Local" to identify operating outpatient specialists in each of the 36 counties (Das Telefonbuch, 2018).

In a pre-test, we called specialists anonymously and requested appointments for six different non-urgent medical examinations in the cities of Berlin, Cologne, Bonn, Leverkusen, Hamburg and Munich. The treatments that we asked for were an allergy test, a hearing test, an eye examination, a gastroscopy, a magnet-resonance-therapy of the right knee, and a pulmonary function test.

After this pre-test, in the remaining 30 counties, using the exact same protocol as in the pre-test, we called gastroenterologists, otorhinolaryngologists, and allergists to request appointments for the following three examinations: (a) an upper gastrointestinal endoscopy, (b) an audiometry, and (c) an allergy test. We chose these three (out of six) examinations because they turned out to be the most popular, non-urgent routine examinations which were relatively easy to request and schedule.

Study Design

In total, we called 991 private practices to request appointments by telephone. The same test person (the "caller") made the calls over the course of one calendar year, between

April 6, 2017 and May 3, 2018. Importantly, the test person called each practice *twice* and clearly indicated the insurance status of the hypothetical patient. We randomized whether the caller would pretend to be privately or publicly insured.[6] Moreover, we made the two calls in time intervals of at least two weeks to not trigger any suspicion about being part of a field experiment. (The administrative front desk staff who took the calls did not know that they were part of a scientific field experiment.)

During each call, we followed a pre-determined standardized protocol on how to start and end the call and what answers to give in response to the most frequently asked questions. All calls were made between Monday and Friday during the regular office hours of each practice.[7] During the call, the caller mentioned that she had a referral by her Primary Care Physician (PCP). When asked for the name of the PCP, the caller gave a fictional name and indicated that the practice would be located in her hometown. Finally, the caller ended all calls without fixing the suggested appointment to not occupy a slot that could be used for a real treatment. Also, recall that all requests were for elective non-urgent treatments.

As mentioned, we called 991 unique private practices in the 36 German counties displayed in Figure 1. Figure A1 in the Appendix shows the distribution of the contacted practices across the 36 counties. The number of contacted practices varies between 1 in two

[6] When asked about the name of the insurance company, a real name was given. Because reimbursement rates are centrally determined for publicly and privately insured (Section 3) and not negotiated individually between insurer and provider, the actual insurer is not crucial in the German setting.

[7] If special office hours were indicated on voicemail, the following calls were made during these times. When nobody answered the phone, the practice was market as "not available" after three unsuccessful attempts. When the line was busy in one of these three attempts, the maximal number of attempts was raised to six.

very small and low populated counties and 126 in one big German city. The mean number of practices contacted was 26 per county. In the majority of counties, all three specialists were available.

5. Data

Sample Selection

First, we entirely excluded certain practices from our study for the following reasons: (i) the specialist was not active anymore (17, 1.7%), (ii) the practice offered only treatments for privately insured patients[8] (31, 3.1%), and (iii) other reasons[9] (27, 2.7%). These reasons reduced the number of unique practices in our study by 75 from 991 to 916.

Second, there were other reasons why practices were unresponsive and we could (structurally) not make appointments; (1) the practice was closed for at least one week, for example during vacations (18 practices, 2%) or (2) the practice was not reachable after several unsuccessful attempts (282, 28%, see footnote 7). In these cases, the practice was only included once under one randomized insurance status, for example, when the vacations were over. In other words, for all eligible practices that were not entirely excluded due to reasons (i) and (ii) above, we either tried to make an appointment during the first time we called under insurance

[8] Practices have the option to entirely opt out of treating publicly insured patients and declare themselves a practice for the privately insured or people who pay entirely out-of-pocket. These practices, however, are then prohibited from treating any publicly insured patient and charge the social insurance, even when demand from private patients is low. We do not consider these practices as relevant to the experiment.

[9] E.g. practices for children only and misleading telephone numbers.

status A, during the second time we called under insurance status B, or in both cases.[10] We call this unbalanced sample "Sample A;" it has 1552 caller-appointment observations. Table 1 shows the descriptive statistic for this full sample.

In contrast, our "Sample B" is a sample where we only include caller-appointment observations where the practice offered an appointment to *both* hypothetical patients, the publicly and the privately insured. This sample is balanced, includes 504 unique private practices, and 1,008 caller-appointment observations.

Main Outcome Variables

We generate two main outcome variables, both of which measure access to the health care system. The first variable is binary and called *apptm.* It indicates whether the successfully contacted practice was willing to offer an appointment to the hypothetical patient. As seen in Table 1, in 81% of all cases, the caller was offered an appointment.

` The second variable is continuous and called *dayswait*. It counts the number of workdays from the calling date to the offered appointment.[11] It only has valid values for the 81% of cases where a specific appointment was offered. Figure A2 show the distribution of *dayswait* and Table 1 shows the summary statistic. As seen, the minimum wait time is an

[10] Practices provided several reasons for why no appointment could be offered, some of which may have been true and others excuses. For example, office staff said that the practice would not make fixed appointments or the practice temporarily would not accept new patients. We remain agnostic about why specifically no appointment was offered but test below whether, overall, the insurance status had an impact on the likelihood to receive an appointment.

[11] This means that we excluded weekends (Saturday, Sunday) as well as public holidays. In a robustness checks, we also excluded work days between a public holiday and weekends ("Brückentage") as many Germans take vacation days during these days to have an extended weekend off. We call this variable *dayswait II* (see Table 1).

immediate appointment, when patients could be seen on the same day. The maximum wait time is 171 weekdays and the average wait time is 19 weekdays (almost 4 weeks). Figure A2 shows a left-skewed distribution with a long right tail.

Main Control Variables

The main variable of interest is *privately insured*. Even in the unbalanced sample with 1,522 caller-appointment observations, this variable is almost perfectly balanced with 49.5% of all observations representing a caller who said she would be privately insured (Table 1).

Other important control variables indicate the day of the week, the exact calendar date, the time of the day the call took place, whether the randomized insurance status was privately or publicly insured during the first call, and the specialty of the practice.

County-Level Control Variables

The final panel in Table 1 lists the county-level control variables. These have been provided by the *Federal Statistical Office* (Destatis, 2018a, b) and by the *Federal Institute for Construction, Urban and Space Research*, "Bundesinstitut für Bau-, Stadt- und Raumforschung" (BBSR, 2018). As seen, the average age of all residents in the 36 counties is 43 years, the average net income is €1,786 and the average unemployment rate is 7.9%.

6. Statistical Methods

Most important for causal inference in this setting is the fact that we set up a field experiment in which a test person called each private practice twice with a randomized public or private insurance status. Calling each practice twice and randomizing the insurance status guarantees balanced covariates by design.

Because the same test person called all 991 practices and strictly followed a pre-determined protocol, simple descriptive statistics and nonparametric bivariate tests should yield first reliable evidence about access differences between the two insurance groups.

As our main statistical tests, we run OLS and count data regression models which routinely control for the calendar date, the day of the week and the time during the day of the call - in addition to practice-level and county-level controls. In particular, we run the following model as our first model:

$$a_{ip} = \alpha + \beta * PHI_i + X_p'\tau + Z_c'\theta + \gamma DOW_{ip} + \delta_t + \rho_p + e_{ip} \quad (1)$$

where a_{ip} stands for our first outcome variable *apptm*, which is binary and indicates—using the unbalanced Sample A—whether practice p offered hypothetical patient i an appointment or not. The main variable of interest is PHI_i and indicates whether the caller indicated to be publicly or privately insured. The model also controls for a set of practice-level controls X_p', in particular the specialty group, as well as a set of county-level controls Z_c' such as the county-level unemployment rate or the physician density (see Table 1). In addition, the model routinely controls for the day-of-the-week when the caller called a practice (DOW_{ip}) and the time of the day of the call. In the saturated specifications, we include calendar date fixed effects δ_t. Similarly, we replace the practice-level controls with practice fixed effects ρ_p in some specifications. We routinely cluster the standard errors e_{ip} at the practice level and estimate linear probability models using OLS. (However, we also test the robustness of the coefficients using probit models and calculating marginal effects which are available upon request.)

Our second model uses our balanced Sample B and can be written as:

$$\ln(w_{ip}) = \alpha + \beta PHI_i + X_p'\tau + Z_c'\theta + \gamma DOW_{ip} + \delta_t + \rho_p + e_{ip} \qquad (2)$$

where w_{ip} stands for our second outcome variable *dayswait,* and measures the wait time in weekdays for hypothetical patient i in practice *p.* It is continuous but skewed to the left (Figure A1), which is one reason why we replace 0s with 0.01 and take the logarithm. The coefficient estimates of the main variable of interest PHI_i then approximate the wait time differential in percent. The other control variables are defined as above. We also test the robustness of the results by using the plain w_{ip} variable and running negative binomial count data models that consider excess zeros and overdispersion.

In extended specifications, we test for effect heterogeneity by interacting PHI_i with regional and other variables and add these interaction terms to the model.

7. Results

Nonparametric Findings

We start by plotting nonparametric results. In a perfectly randomized setting, they should well approximate the parametric findings that additionally control for date, day-of-week, time-of-day and practice fixed effects.

Figure 2 plots bar diagrams of the first outcome variable *apptm* along with 95% confidence intervals. As can be easily spotted with bare eyes, the share of privately insured who were offered an appointment (85%) is larger than the share of publicly insured (77%) who were offered an appointment. The confidence intervals do not overlap, suggesting that the eight percentage point difference is statistically significant at the 5% level. A formal t-test has a t-value of 3.7 and is statistically significant at the 0.1% level.

[Figure 2 about here]

Next, Figure 3 plots the distribution of the second outcome variable *dayswait* separately for the privately and publicly insured using the balanced Sample B. Again, it is easy to see that the wait time distribution for the privately insured is much more left-skewed than the wait time distribution for the publicly insured. The former has a lot more mass over the 0 to 20 day region and the latter has a much longer right tail and mass exceeding 100 weekdays wait time.

[Figures 3 and 4 about here]

Figure 4 illustrates mean differences in wait times between the publicly and privately insured by showing bar diagrams along with 95% confidence intervals, also using Sample B. As seen, the mean wait time for publicly insured is almost twice as large and 25 days long, whereas the wait time for privately insured is just 11.6 days, on average. The confidence intervals clearly do not overlap indicating a highly significant difference in wait times, depending on the insurance status. This prior is confirmed by a formal t-test which is significant at the 0.1% level with a t-value of 9.8.

Parametric Findings

Now, we move on to our parametric findings and multivariate regression models. Table 2 shows the findings from our first model in equation (1), which uses the binary *apptm* measure as outcome variable. These models assess the impact of the insurance status on the likelihood to be offered an appointment for non-urgent treatments. Each column in Table 2 represents one model. The models only differ by the inclusion of different sets of covariates as indicated in the bottom panel of the Table.

[Table 2 about here]

Table 2 shows the following: First, we find that the insurance status of the caller has a highly significant impact on being offered an appointment. All four coefficient estimates are significant at the 5% or even 1% level. Being privately insured increases the likelihood for an appointment by 7 to 8 percentage points or by about 10 percent relative to the mean of 0.81. Second, the estimates are very robust across model specifications. The inclusion of week-of-year fixed effects, county fixed effects, and even practice fixed effects barely alter the size of the coefficients. Third, the findings are also robust to running probit models and calculating marginal effects (available upon request).

Table 3 follows the same setup as Table 2 but estimates our second model and equation (2). It is basically identical to equation (1) but uses the second continuous outcome variables *dayswait*, which counts the waiting time in weekdays. The coefficient estimates then indicate the impact of being privately insured on the mean wait time in weekdays, relative to being publicly insured. Because we take the logarithm of the dependent variable, all coefficient estimates are then approximately differences in percent. In the Appendix, in Table A1, we replicate Table 3 but do not take the logarithm. All six models in Table 3 and Table A1 use our balanced Sample B and only include the 504 unique practices that offered specific appointments to both callers, the publicly and the privately insured.

[Table 3 about here]

Table 3 and A1 show the following: All six models show coefficients that are highly significant at the 1% level. Moreover, the estimates are very robust to the sets of covariates included, reinforcing that our randomization was very successful. It also implies the absence of structural differences in terms of the week of the year or the county of residence. Moreover, because the differences are very close to the differences of simple t-tests, it suggests the absence of structural differences in terms of the day of the week or the time of the day when the call was made.

In terms of content, the results show that privately insured patients have to wait on average 13 fewer weekdays for an appointment, conditional on being offered one. In other words, publicly insured patients have to wait more than twice as long for an appointment; the mean wait time for publicly insured people is 24.9 weekdays (or about 5 weeks on average), whereas the mean wait time for privately insured patients is only 11.6 weekdays (or a little more than 2 weeks on average).

Finally, we test for effect heterogeneity in the inequality in health care access. Technically, we interact our variable of interest *Privately Insured* with the stratifying county-level covariate. Then we add the interaction term along with the two variables in levels to the models as in equations (1) and (2). Panel A of Table 4 shows the results for *apptm* and Panel B of Table 4 shows the results for *dayswait*.

[Table 4 about here]

As seen, few of the interaction terms (which indicate the differences in insurance status by the stratifying covariate) are statistically significant. The findings for East Germany are relatively large and the sign of the effects consistent with the notion that the differences in East Germany are smaller than in West Germany. However, the two interaction terms in column (4) of Table 4 are only significant at the 20% level and rather suggestive.

Second, the findings in Panel B suggest that a lower physician density is associated with a decrease in the discrimination of publicly insured patients. However, this effect is at least partly driven by the lower physician density in the East German states (except Berlin).

Finally, the finding in column (2) of Panel B suggest that a higher population density, e.g. in cities as compared to more rural counties, is associated with an increase in discrimination and inequality in access.

8. Conclusion

The main objective of this research was to implement a field experiment to cleanly assess the impact of insurance-related differences in reimbursement rates on health care access. The findings also allow us to draw conclusions about whether private practices discriminate against patients based on their profitability.

We use the German institutional setting for the field experiment because it is particularly well suited as a clean testing ground. The reason is that Germany has coexisting public and private insurance systems. The reimbursement rates for the privately insured are about 2.5 times higher and are classic fee-for-service schemes without caps or bundled payments. Importantly, reimbursement rates for both regimes are standardized and centrally set, not by individual negotiations between insurers and providers. Likewise, provider networks do not exist in Germany.

Our test person called almost one thousand private outpatient practices over the course of one calendar year. A strict protocol was followed and the randomized insurance status of the hypothetical patients was clearly revealed during the call. Each practice was called twice, once the caller pretended to be publicly insured and once the caller pretended to be privately insured. In each case, an appointment for a non-urgent medical treatment was requested. The treatments requested were a gastroscopy with a gastroenterologists, a hearing test with an otorhinolaryngologists, and an allergy test with an allergist.

Our findings show clear evidence that inequalities in reimbursement rates create inequalities in health care access and higher access barriers for less profitable patients, both on the extensive and intensive margin. Publicly insured patients were 10 percent less likely to be offered an appointment. Moreover, when offered appointments, publicly insured patients had to wait 13 weekdays, or almost 3 weeks, longer (and more than twice as long) compared to private patients. When stratifying the findings by county characteristics, we find that inequality in wait times is larger when the physician density in the county is larger. The same relationship holds for the population density. Moreover, we find suggestive evidence that inequality in access is smaller in East as compared to West Germany, possibly indicating a long-term effect of socialist norms (Alesina and Fuchs-Schündeln 2007; Rainer and Siedler, 2009). For example, in a recent survey, PWC (2017) finds that a 10 percentage point higher share of West as compared to East Germans have a positive attitude towards more competition in the health care sector (59% vs. 49%, see).

The policy implications of our findings suggest that uniform reimbursement rates would minimize inequalities in health care access. Because, in the U.S. and Germany, healthier and wealthier individuals tend to have private insurance with higher reimbursements rates, such a system exacerbates inequalities in health care access and population health.

References

Alesina, A. ,Fuchs-Schündeln,N., 2007. Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences. American Economic Review, 97 (4), pp. 1507-1528.

Arrow, K. J., 1963. Uncertainty and the Welfare Economics of Medical Care. American Economic Review. 53(5), pp. 941-973.

Asplin, B. R., Rhodes, K. V., Levy, H., Lurie, N., Crain, A. L., Carlin, B. P., , Kellermann, A. L., 2005. Insurance status and access to urgent ambulatory care follow-up appointments. Journal of the American Medical Association, 294(10), pp. 1248-54.

Atal, J., H. Fang, M. Karlsson, Ziebarth, N. R. 2018. Exit, voice or loyalty? an investigation into mandated portability of front-loaded private health plans. Journal of Risk and Insurance, forthcoming.

BBSR - Bundesinstitut für Bau-, Stadt- und Raumforschung, 2018. Inkar. Indikatoren und Karten zur Raum- und Stadtentwicklung. http://www.inkar.de (date of access: June 10, 2018).

BMG - Bundesministerium für Gesundheit, 2018. BMG. https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/e/einheitlicher-bewertungsmassstab-ebm.html (date of access: June 10, 2018).

Brosig-Koch, J., Kairies-Schwarz, N., Kokot, J., 2017. Sorting into payment schemes and medical treatment: a laboratory experiment. Health Economics, 26(S3), pp. 52-65.

Buennings, C., H. Schmitz, H. Tauchmann, and N. R. Ziebarth (2018). How health plan enrollees value prices relative to supplemental benefits and service quality. Journal of Risk and Insurance forthcoming.

CMS – Centers for Medicaire and Medicaid Services 2018a. Physician Fee Schedule, https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/index.html, (date of access: September 10, 2018).

CMS – Centers for Medicaire and Medicaid Services 2018b. CY 2018 Medicare Fee Schedule Changes, https://gi.org/wp-content/uploads/2018/01/MPFS-Final-2018-v-2017-Rates.pdf, (date of access: September 10, 2018).

Clemens, J., Gottlieb, J.D., 2014. Do physician's financial incentives affect medical treatment and patient health? American Economic Review, 104(4), pp. 1320-1349.

Commonwealth Fund, 2014. Mirror, Mirror on the Wall, 2014 Update: How the U.S. Health Care System Compares Internationally, by Karen Davis, Kristof Stremikis, David Squires, and Cathy Schoen, https://www.commonwealthfund.org/publications/fund-reports/2014/jun/mirror-mirror-wall-2014-update-how-us-health-care-system (date of access: October 18, 2018).

Commonwealth Fund, 2017. Mirror, Mirror 2017: International Comparison Reflects Flaws and Opportunities for Better U.S. Health Care, by Eric C. Schneider, Dana O. Sarnak, David Squires, Arnav Shah, and Michelle M. Doty, https://www.commonwealthfund.org/publications/fund-reports/2017/jul/mirror-mirror-2017-international-comparison-reflects-flaws-and (date of access: October 18, 2018).

Cooper, C., Craig, S.V., Gaynor, M., Reenen, J.V. 2018. The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured. The Quarterly Journal of Economics, forthcoming.

Cullis, J. G., Jones, P. R., Propper, C., 2000. Chapter 23 Waiting lists and medical treatment: Analysis and policies. Handbook of Health Economics, 1(PART B), pp. 1201-1249.

Das Telefonbuch, 2018. https://telefonbuch.t-online.de/ (date of access: June 10, 2018).

Destatis - Federal Statistical Office Germany, 2018a. Daten aus dem Gemeindeverzeichnis - Kreisfreie Städte und Landkreise nach Fläche und Bevölkerung auf Grundlage des Zensus 2011 und Bevölkerungsdichte. https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Gemeindeverzeichnis/Administrativ/Aktuell/04Kreise.html (date of access: June 10, 2018).

Destatis - Federal Statistical Office Germany, 2018b. Daten aus dem Gemeindeverzeichnis - Bundesländer mit Hauptstädten nach Fläche und Bevölkerung auf Grundlage des Zensus 2011 und Bevölkerungsdichte.

Ellis, R. P., McGuire, T. G. (1986). Provider behavior under prospective reimbursement. Journal of Health Economics, 5(2), pp. 129-151.

Felder, S. 2008. To wait or to pay for medical treatment? Restraining ex-post moral hazard in health insurance, Journal of Health Economics,27(6), pp. 1418-1422.

GKV Spitzenverband, 2018. Statutory Health Insurance. https://www.gkv-spitzenverband.de/english/statutory_health_insurance/statutory_health_insurance.jsp, (date of access: September 10, 2018).

GOÄ - Gebührenordnung für Ärzte 2018a. https://diego.one/go%C3%A4/687 (date of access: September 10, 2018).

GOÄ - Gebührenordnung für Ärzte 2018b. https://diego.one/go%C3%A4/676 (date of access: November 17, 2018).

GOÄ - Gebührenordnung für Ärzte 2018c. https://diego.one/go%C3%A4/385 (date of access: November 17, 2018).

GOÄ - Gebührenordnung für Ärzte 2018d. https://diego.one/go%C3%A4/1403 (date of access: November 17, 2018).

Gravelle, H., Siciliani, L. 2008. Optimal quality, waits and charges in health insurance, Journal of Health Economics, 27(3), pp. 663-674.

Halpern, M. T., Romaire, M. A., Haber, S. G., Tangka, F. K., Sabatino, S. A., Howard, D. H. 2014. Impact of state-specific Medicaid reimbursement and eligibility policies on receipt of cancer screening, Cancer, 120(19), pp. 3016-3024.

Heinrich, N., A. Wübker und C. Wuckel (2018), Waiting times for outpatient treatment in Germany: New experimental evidence from primary data. Jahrbücher für Nationalökonomie und Statistik 238 (5), pp. 375-394.

Jones, A.M., Rice, N., Robone, S., Dias, P.R. 2011. Inequality and polarisation in health systems' responsiveness: A cross-country analysis, Journal of Health Economics, 30(4), pp. 616-625.

KBV - Kassenärztliche Bundesvereinigung, 2017. Einheitlicher Bewertungsmaßstab (EBM). http://www.kbv.de/media/sp/EBM_Gesamt___Stand_1._Quartal_2017.pdf (date of access: June 10, 2018).

KBV - Kassenärztliche Bundesvereinigung, 2018a. Aktuelles zu den Honorarverhandlungen 2019, http://www.kbv.de/html/2054.php (date of access: September 10, 2018).

KBV - Kassenärztliche Bundesvereinigung, 2018b. Einheitlicher Bewertungsmaßstab (EBM). http://www.kbv.de/media/sp/EBM_Gesamt___Stand_2._Quartal_2018.pdf (date of access: December 09, 2018).

Kuchinke, B., Sauerland, D., Wübker, A., 2009. The influence of insurance status on waiting times in German acute care hospitals: an empirical analysis of new data. International Journal for Equity in Health, 8(44).

KVNO - Kassenärztliche Vereinigung Nordrhein, 2018. Honorarverteilungsmaßstab der Kassenärztlichen Vereinigung Nordrhein. https://www.kvno.de/downloads/honorar/hvm1801.pdf (date of access: June 10, 2018).

Laudicella, M., Siciliani, M., Cookson, R., 2012. Waiting times and socioeconomic status: evidence from England. Social Science & Medicine, 74(9), pp. 1331-1341.

Lindsay, C.M., Feigenbaum, B., 1984. Rationing by Waiting Lists, The American Economic Review, 74(3), pp. 404-417.

Lüngen, M., Stollenwerk, B., Messner, P., Lauterbach, K. W., Gerber, A., 2008. Waiting times for elective treatments according to insurance type: A randomized empirical study in Germany. International Journal for Equity in Health, 7.

Ramos, A.L., Hoffmann, F. and Spreckelsen, O., 2018. Waiting times in primary care depending on insurance scheme in Germany, BMC Health Services Research, 18(1), pp. 191-199.

McGuire, T. G. (2000). Chapter 9 -Physician agency. In: Handbook of Health Economics, Eds. Culyer, A. J., Newhouse J. P., 1(Part A), pp. 461-536. Amsterdam: North Holland.

Mackillop, William J., Zhou, Y., Quirt, C.F. 1995. A comparison of delays in the treatment of cancer with radiation in Canada and the United States, International Journal of Radiation Oncology, 32(2), pp. 531 – 539.

Monstad, K., Birger, L., Ãeter, E., Espehaug, B., 2014. Waiting time and socioeconomicstatus an individual-level analysis. Health Economics, 23(4), pp. 446-461.

Nicholson S., Pauly M.V., Wu A.Y.J., Murray J.F., Teutsch S.M., Berger M.L., 2018. Getting real performance out of pay-for-performance, The Milbank Quarterly, 86(3), pp. 435-457.

Pilny, A., Wübker, A., Ziebarth, N. R., 2017.Introducing risk adjustment and free health plan choice in employer-based health insurance: Evidence from Germany, Journal of Health Economics, Elsevier, 56(C), pp. 330-351.

PWC 2017. Bevoelkerungsbefragung: Wuensche and das Gesundheitssystem, September 2017, https://www.pwc.de/de/gesundheitswesen-und-pharma/berichtsband-gesundheitssystem.pdf (date of access: October 10, 2018).

Rainer, H. Siedler, T., 2019. Does democracy foster trust? Journal of Comparative Economics, 37(2), pp. 251-269.

Reinhardt, U. E., 2011. The many different prices paid to providers and the flawed theory of cost shifting: is it time for a more rational all-payer system? Health Affairs, 30(11), pp. 2125-2133.

Roll, K., Stargardt, T., Schreyögg, J., 2012. Effect of type of insurance and income on waiting time for outpatient care. The Geneva Papers on Risk and Insurance Issues and Practice, 37(S4), pp. 609-632.

Schmitz, H., Ziebarth, N. R. 2017). Does framing prices affect the consumer price sensitivity of health plan choice? The Journal of Human Resources, 52(1), pp. 89–128.

Schwierz, C., Wübker, A., Wübker, A., Kuchinke, B. A., 2011. Discrimination in waiting times by insurance type and financial soundness of German acute care hospitals. European Journal of Health Economics, 12(5), pp. 405-416.

Siciliani, L., Hurst, J., 2005. Tackling excessive waiting times for elective surgery: A comparative analysis of policies in 12 OECD countries. Health Policy, 72(2), pp. 201-215.

Simon, M. 2017. Das Gesundheitssystem in Deutschland: Eine Einführung in Struktur und Funktionsweise. Bern: Hogrefe Verlag.

Siciliani, L., 2006. A dynamic model of supply of elective surgery in the presence of waiting times and waiting lists, Journal of Health Economics, 25(5), pp. 891-907.

Sundmacher, L., Kopetsch, T., 2013. Waiting times in the ambulatory sector--the case of chronically ill patients. International Journal for Equity in Health, 12(1), pp. 77-87.

van Doorslaer, E., Wagstaff, A., van der Burg, H., Christiansen, T., De Graeve, D., Duchesne,I., Gerdtham, U.-G., Gerfin, M., Geurts, J., Gross, L., Häkkinen, U., John, J., Klavus, J., Leu, R. E., Nolan, B., O'Donnell, O., Propper, C., Puffer, F., Schellhorn, M., Sundberg, G., Winkelhake, O., 2000. Equity in the delivery of health care in Europe and the US, Journal of Health Economics,19(5), pp. 553-583.

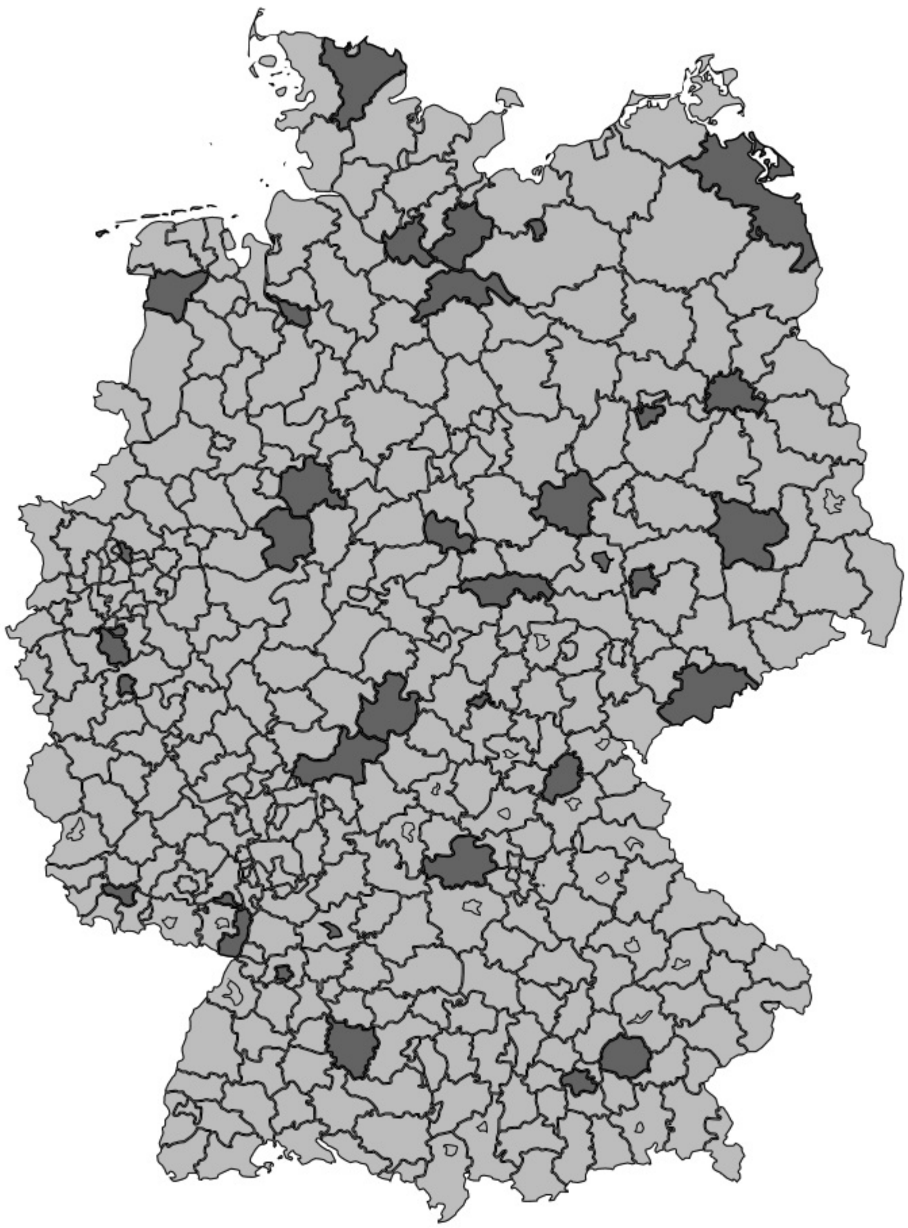
Versorgungsatlas, 2018. Versorgungsstrukturen. https://www.versorgungsatlas.de/themen/versorgungsstrukturen/ (date of access: June 10, 2018).

Viberg, N., Forsberg, B. C., Borowitz, M., Molin, R., 2013. International comparisons of waiting times in health care - Limitations and prospects. Health Policy, 112(1–2), pp. 53-61.

Walendzik, A., Greß, S., Manouguian, M., Wasem, J., 2008. Vergütungsunterschiede im ärztlichen Bereich zwischen PKV und GKV auf Basis des standardisierten Leistungsniveaus der GKV und Modelle der Vergütungsangleichung. Diskussionbeitrag Universität Duisburg-Essen 165.

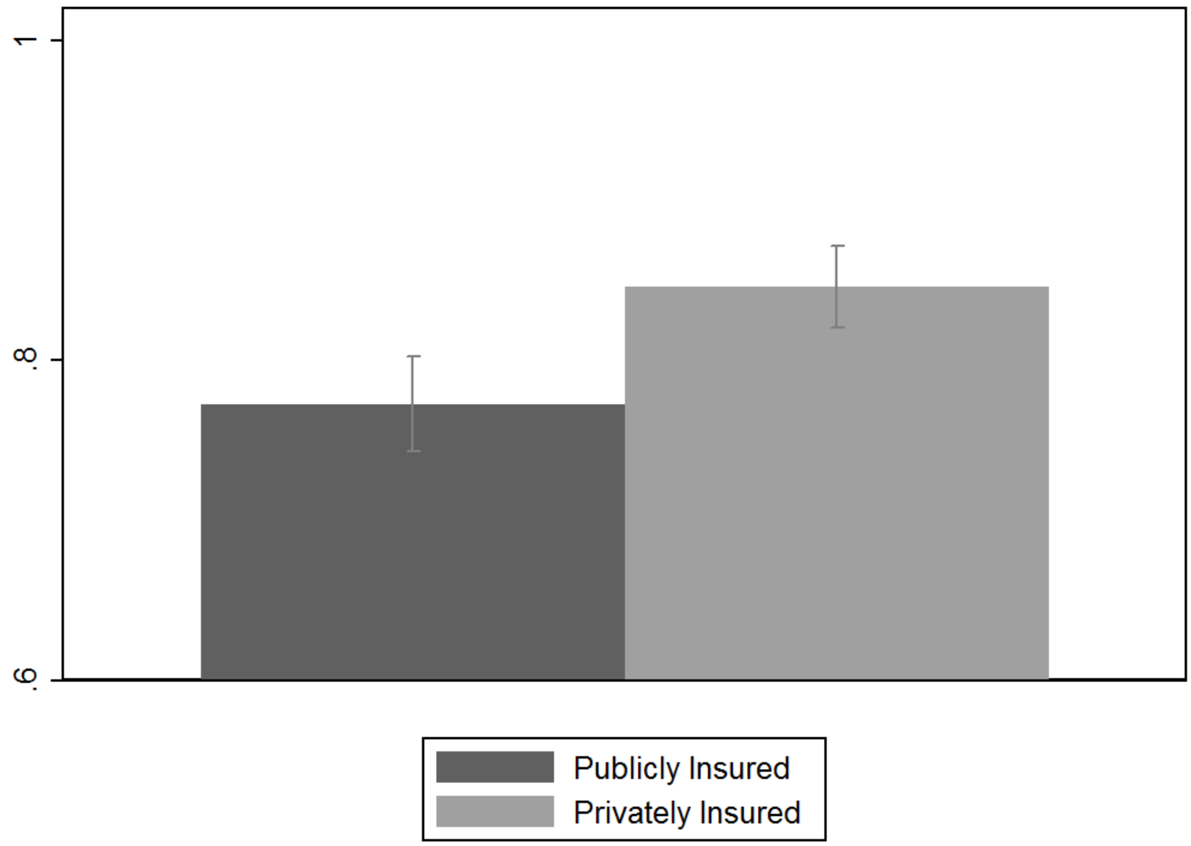
Figures and Tables

Figure 1: Selected Counties for Field Experiment



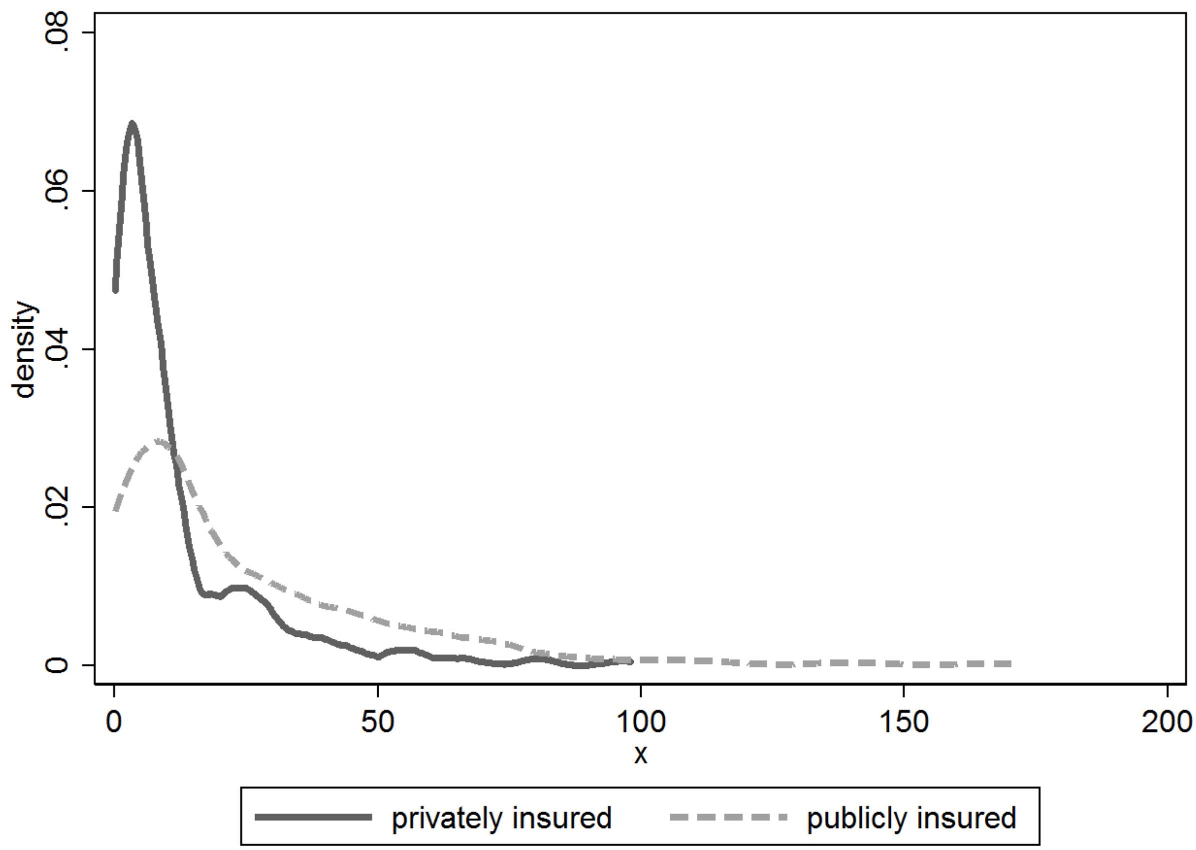
Source: Own illustration. The 36 selected counties that we selected for the field experiment are dark gray. Between one and four counties in each of the 16 federal states were selected (see Appendix B).

Figure 2: Likelihood to be Offered Appointment by Insurance Status



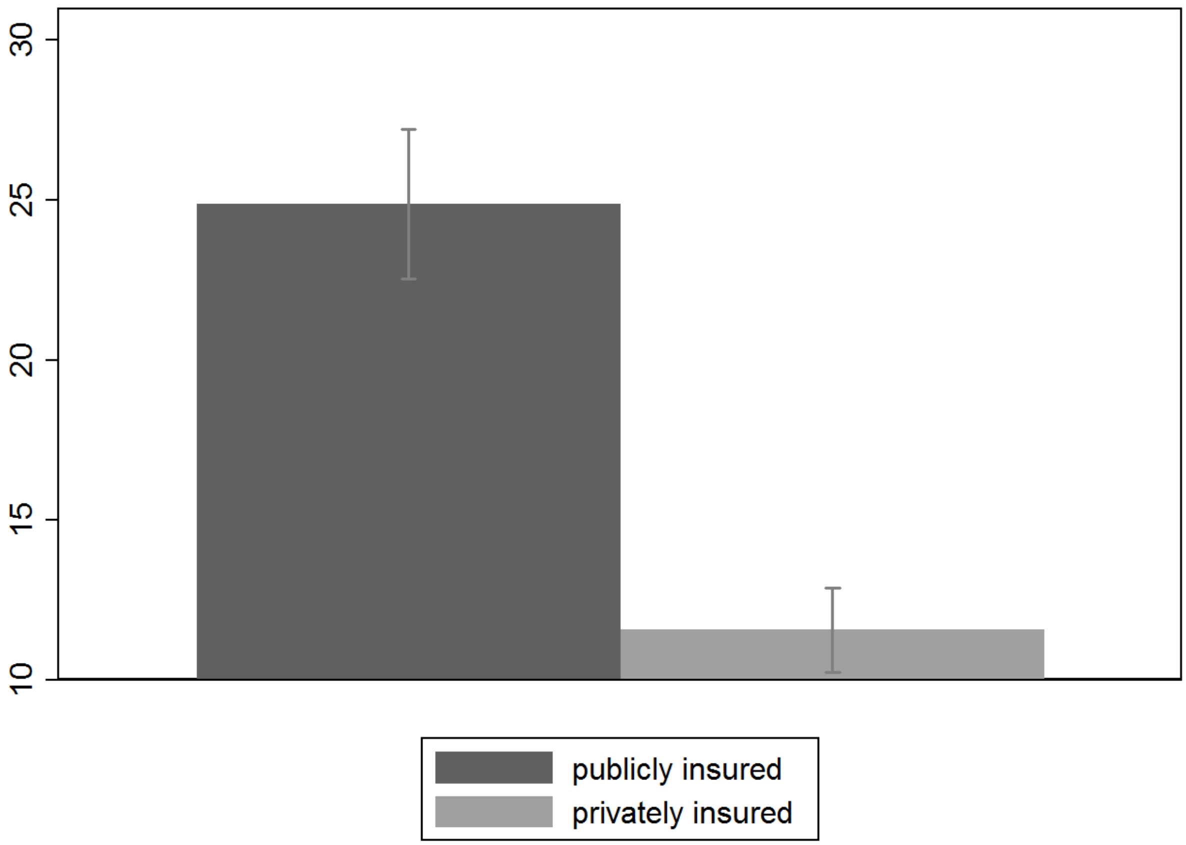
Source: Own calculation, own illustration. Graph uses Sample A. The bars show 95% confidence intervals.

Figure 3: Distribution of Wait Times in Week Days by Insurance Status



Source: Own calculation, own illustration. Graph uses Sample B. X-axis shows the number of weekdays, counting from the day of the call until an appointment was offered.

Figure 4: Average Wait Times by Insurance Status



Source: Own calculation, own illustration. Graph uses Sample B. The bars show 95% confidence intervals.

Table 1: Descriptive Statistic

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) N
Dependent Variables					
Apptm	0.809	0.393	0	1	1,522
Dayswait	18.92	23.11	0	171	1231
Dayswait II	18.27	22.65	0	171	1231
Main Independent Variables					
Privately Insured	0.495	0.500	0	1	1,522
Allergy test	0.229	0.421	0	1	1,522
Hearing test	0.343	0.475	0	1	1,522
Gastroscopy	0.175	0.380	0	1	1,522
First call privately insured	0.493	0.500	0	1	1,522
County-Level Controls					
Unemployment rate	7.933	2.643	2.100	13.60	1,522
Living space in m^2	41.82	4.071	38	52.60	1,522
Employment rate per 100 population in working age	79.67	2.650	76.10	86.70	1,522
Average age	42.95	2.079	41.20	49.60	1,522
Share of population 65+	19.95	2.907	16.20	28.90	1,522
Employees w/ academic degree	10.21	3.870	3.800	16.80	1,522
Household inc. per capita, €	1,786	319.6	1,362	3,451	1,522
Share recreational area	48.36	35.05	18.80	225.4	1,522
Residents per doctor	484.5	148.0	303	853.1	1,522
Population	1.088e+06	1.068e+06	35,608	3.575e+06	1,522
Area in km^2	715.9	648.1	78.87	3,946	1,522
Residents per km^2	2,129	1,576	55	4,668	1,522
Driving distance to city center in minutes	3.487	5.435	0	16	1,522
Eastern federal state	0.140	0.347	0	1	1,522
Northern federal state	0.627	0.484	0	1	1,522
Rural district	0.283	0.450	0	1	1,522
Income class (1-5)	2.783	1.569	1	5	1,522
Residents per doctor in categories (1-5)	2.551	1.376	1	5	1,522

Sources: Own data collection, own illustration. See Section 4 and 5 for details. County-level controls are taken from BBSR (2018) and Destatis (2018a, b).

Table 2: Impact of Insurance Status on Likelihood to be Offered Appointment

	(1)	(2)	(4)	(3)
	apptm	*apptm*	*apptm*	*apptm*
Privately Insured	**0.0690******	**0.0771******	**0.0823******	**0.0734*****
	(0.0213)	**(0.0210)**	**(0.0210)**	**(0.0297)**
Day-of-week FE	X	X	X	X
Month-of-year FE	X			
Week-of-year FE		X	X	X
Calling time of day	X	X	X	X
Specialty controls		X	X	X
Practice FE				X
County FE			X	
Observations	1,522	1,522	1,522	1,522
R-squared	0.0374	0.0938	0.1256	(3)

Sources: Own data collection, own illustration. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column is one model as in equation (1) using Sample A, see Section 5 for details. The mean of the dependent binary variable *apptm* indicating whether an appointment was offered is 0.81 (see Table 1).

Table 3: Impact of Insurance Status on Wait Times

	(1) *Log(dayswait)*	(2) *Log(dayswait)*	(4) *Log(dayswait)*	(3) *Log(dayswait)*
Privately Insured	**-1.1196***** **(0.0893)**	**-1.1273***** **(0.0898)**	**-1.1368***** **(0.0899)**	**-1.1381***** **(0.1279)**
Day-of-week FE	X	X	X	X
Month-of-year FE	X			
Week-of-year FE		X	X	X
Calling time of day	X	X	X	X
Specialty controls		X	X	X
Practice FE				X
County FE			X	
Observations	1,008	1,008	1,008	1,008
R-squared	0.1540	0.2445	0.3056	0.7316

Sources: Own data collection, own illustration. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column is one model as in equation (2) using Sample B, see Section 5 for details. The mean of the dependent binary variable *daywait* indicating the number of weekdays until the offered appointment is 24.89 for publicly and 11.57 days privately insured patients. The overall mean is 18.23 (all values for Sample B, not shown in Table 1). All models use the logarithm of *dayswait* where values of 0 have been replaced with 0.01.

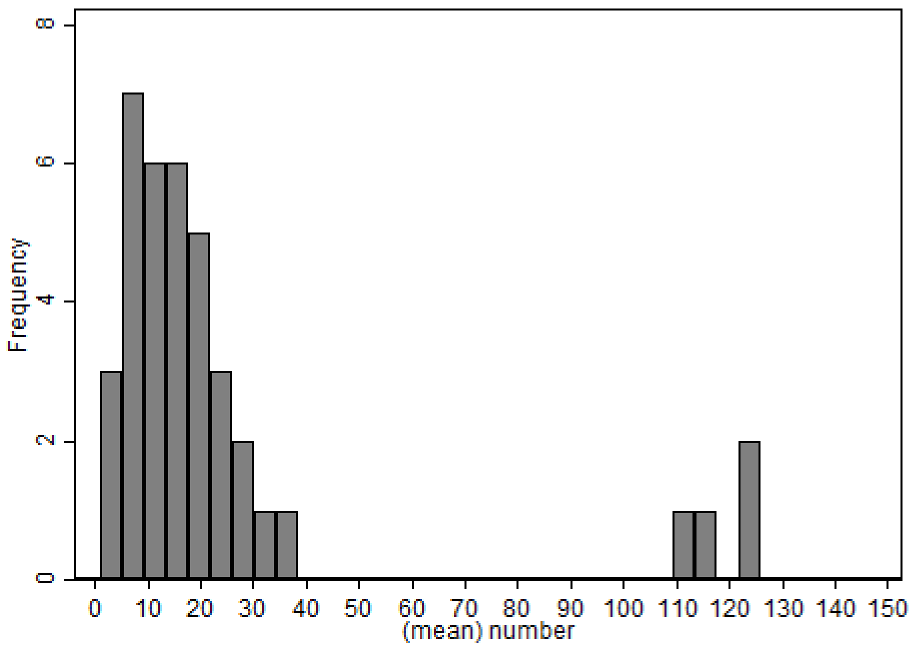
Table 4: Effect Heterogeneity

	(1) Resident per Physician	(2) Resident per km2	(3) Household income	(4) East Germany
Panel A: apptm				
Privately Insured *[column]	**0.0002***	**-0.0000**	**0.0000**	**-0.0956**
	(0.0001)	**(0.0000)**	**(0.0001)**	**(0.0694)**
Privately Insured	-0.0351	0.0969***	0.0240	0.0928***
	(0.0697)	(0.0362)	(0.1295)	(0.0218)
Column	-0.0000	0.0000*	-0.0001	-0.0451
	(0.0002)	(0.0000)	(0.0001)	(0.0707)
Panel B: dayswait				
Privately Insured *[column]	**0.0010***	**-0.0001****	**-0.0004**	**0.3419**
	(0.0006)	**(0.0001)**	**(0.0003)**	**(0.2815)**
Privately Insured	-1.6347***	-0.8553***	-0.4909	-1.1831***
	(0.3040)	(0.1472)	(0.5315)	(0.0957)
Column	-0.0033***	-0.0002*	-0.0001	0.3292
	(0.0010)	(0.0001)	(0.0003)	(0.3116)

Sources: Own data collection, own illustration. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Each column in each panel is one model. Panel A runs models as in equation (1) with 1,522 observations. Panel B runs models as in equation (2) with 1,003 observations. The column header indicates the stratifying variable and [column] represents these variables.

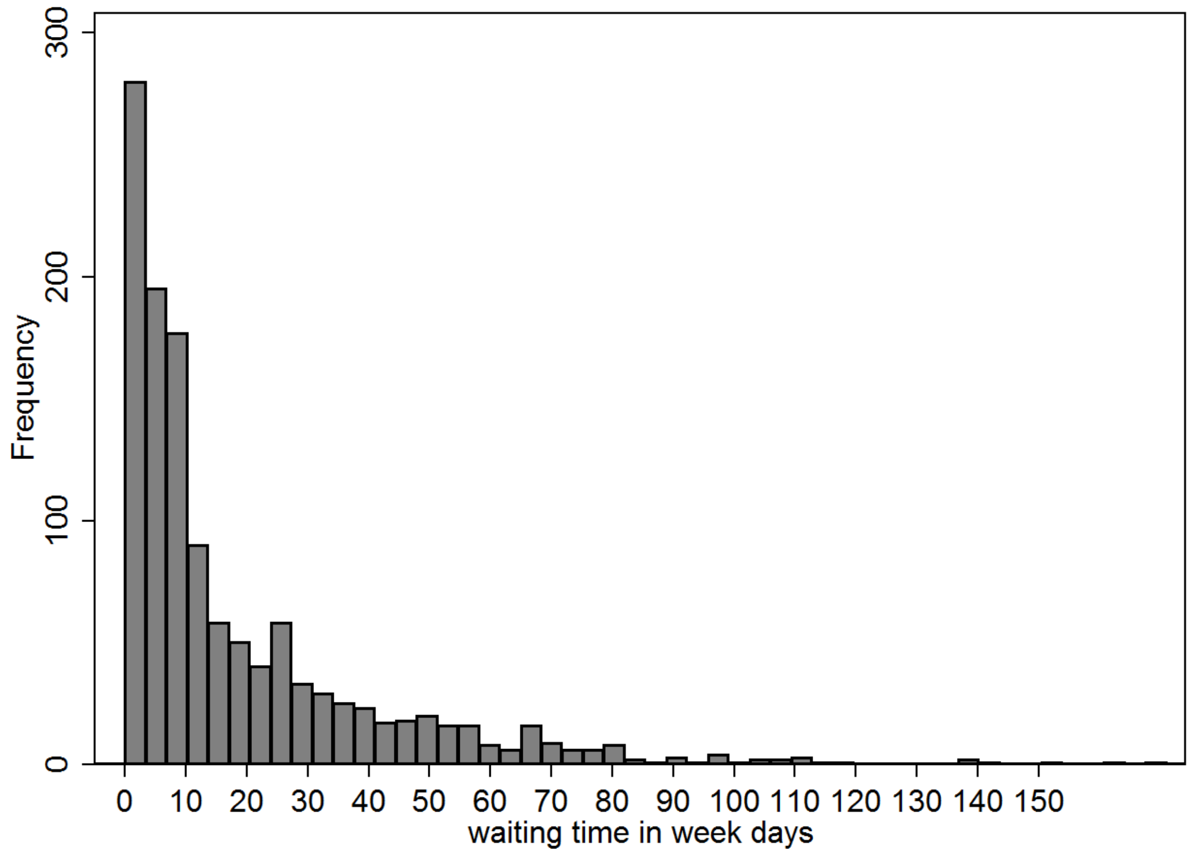
Appendix A

Figure A1: Distribution of Number of Contacted Specialists by 36 Counties



Note: The histogram displays the number of contacted specialists by county. In total, 36 representative counties were included in the experiment (see Section 4).

Figure A2: Distribution of Wait Time in Weekdays



Note: The histogram displays the wait time in weekdays for all successfully contacted practices that offered an appointment, i.e., the 81% of practices in Sample A.

Table A1: Impact of Insurance Status on Wait Times

	(1) *dayswait*	(2) *dayswait*	(4) *dayswait*	(3) *dayswait*
Privately Insured	**-13.1280*** (1.1197)**	**-12.7656*** (1.1044)**	**-13.0267*** (1.1367)**	**-13.1197*** (1.6365)**
Day-of-week FE	X	X	X	X
Month-of-year FE	X			
Week-of-year FE		X	X	X
Calling time of day	X	X	X	X
Specialty controls		X	X	X
Practice FE				X
County FE			X	
Observations	1,008	1,008	1,008	1,008
R-squared	0.1540	0.2445	0.3056	0.7316

Sources: Own data collection, own illustration. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column is one model as in equation (2) using Sample B, see Section 5 for details. The mean of the dependent binary variable *dayswait* indicating the number of weekdays until the offered appointment is 24.89 for publicly and 11.57 days privately insured patients. The overall mean is 18.23 (all values for Sample B, not shown in Table 1). In contrast to Table 3, this table does not take the logarithm of the dependent variable.

Appendix B

Selection of Treatment Counties

We selected the 36 treatment counties using the following procedure based on official data from (BBSR, 2018; Destatis, 2018a, b):

1. Within the 16 federal German states, we chose the number of counties to include based on the population and the geographic size of the counties, such that at least one but at most four counties per federal state were included.

2. We ranked all 16 states by their population and their area in km^2. Then we build four categories based on these two rankings. The four categories then determined whether we included 1, 2, 3, or 4 counties of this state in the field experiment. For example, Bavaria is the largest German state in terms of size (70,542 km^2 or 27,236 $miles^2$). It is the second largest German state in terms of its population (12,930,751 residents in 2017). Hence we included four Bavarian counties in the experiment.

3. Within a state, we decided which specific counties to include based on the average household income in the counties. First, we assigned all counties to one of five income categories.[12] Then, we counted the number of counties in each of the five income categories. For example, Bavaria is a very prosperous state. None of the 70 counties is in the lowest income category, 6 are in the second lowest, 13 in the third lowest, 26 in the second highest and 25 in the highest. Because (2) determined to select four Bavarian

[12] (1) €16.274 - €19.148, (2) €19.149 - €20.928, (3) €20.929 - €22.058, (4) €22.059 - €23.443, (5) €23.444 - €25.663.

counties we chose one from each income category. As another example, Brandenburg (a state in East Germany) is not very populous and prosperous. It has 15 counties in the lowest income category, 2 in the second lowest and 1 in the third lowest. Because (2) determined to select only one county from Brandenburg due to the relatively low number of residents (2,494,648 in 2017), we included a county from the lowest income category.

4. In the last step, we randomly selected the specific county to be included within the income category. For example, steps (2) and (3) determined to choose one of the 15 Brandenburg counties in the lowest income category. We selected the final county to be included randomly. It is gray shaded in Figure 1.