Do Health Plans Risk-Select?
An Audit Study on Germany's Social Health Insurance

Sebastian Bauhoff*

Abstract

This paper evaluates whether health plans in Germany's Social Health Insurance select on an easily observable predictor of risk: geography. To identify plan behavior separately from concurrent demand-side adverse selection, I implement a double-blind audit study in which plans are contacted by fictitious applicants from different locations. I find that plans are less likely to respond and follow-up with applicants from higher-cost regions, such as West Germany. The results suggest that supply-side selection may emerge even in heavily regulated insurance markets. The prospect of risk selection by firms has implications for studies of demand-side selection and regulatory policy in these settings.

Published in:

Journal of Public Economics, Oct 2012, 96 (9-10): 750-759,
http://dx.doi.org/10.1016/j.jpubeco.2012.05.011

Keywords: health insurance, risk selection, risk adjustment, geographic variations, Germany

JEL classification: H5, I18

---

*RAND Corporation, 1200 South Hayes Street, Arlington, VA 22202, sbauhoff@rand.org. Many thanks to Martin Andersen, Eliana Carranza, Amitabh Chandra, Michael Chernew, David Cutler, Mark Duggan, Dirk Göpffarth, Jens Hainmüller, Holger Kern, Miriam Krieger, James O'Malley, Thomas McGuire, Nolan Miller, Joseph Newhouse, Lindsay Sabik, Wynand van de Ven, Amelie Wuppermann, Alan Zaslavsky, two anonymous referees and participants of the Harvard Health Economics workshop for helpful comments. I am grateful to my helpers in Germany and the Harvard Institute for Quantitative Social Sciences for generous financial support. This study was exempt by the Harvard University IRB F17130-101.

1 Introduction

In competitive markets with regulated premiums, health plans have economic incentives to exploit predictable, unpriced heterogeneity in risk by selecting individuals who are low-cost within a premium group. This "cream-skimming" is inequitable and inefficient, and limits payers' ability to leverage high-powered payment systems to encourage efficiency in production (Newhouse, 1996; van de Ven and Ellis, 2000). In addition to the policy concern, the possibility of supply-side selection also has implications for research on consumer behavior in competitive insurance markets, as unobserved concurrent activities by insurers can confound empirical tests of demand-side adverse selection (e.g., Fang et al., 2008; Finkelstein and Poterba, 2006). This paper uses an audit approach to examine whether health plans select on geography, an easily observable predictor of unpriced risk.

Regulatory approaches to contain risk selection aim at limiting plans' access to information about risk types, restricting mechanisms for selection and reducing the potential gains from selection by adjusting payments to more closely reflect individuals' expected costs. However, selection may yet emerge in markets with strict supervision and risk adjustment. In practice, even sophisticated adjustment methods are unable to eliminate all variations in risk, leaving substantial residual incentives for selection (Shen and Ellis, 2002b) which may even increase in the comprehensiveness of the adjustment formula (Brown et al., 2011). Moreover, while risk adjustment may mitigate gains from selection it can decrease incentives for efficiency by moving the payment system closer to cost-based reimbursement (van de Ven and Ellis, 2000). An optimal payment structure may maintain some uncompensated heterogeneity to balance this trade-off.

Managing the selection-efficiency trade-off from risk adjustment is particularly challenging in the case of geography. The location of enrollees is readily observable by plans and correlated with expenditure risk, two conditions that facilitate cream-skimming. However, managing geography is not straightforward. On the one hand, geography has practical appeal as a simple composite index of costs, and accounting for spatial variations can contain potentially large selection incentives. On the other hand, geography is merely correlated with a multitude of cost drivers that regulators may prefer to address separately. In particular, risk adjustment should compensate only for legitimate differences in health care needs or resource costs, e.g. costs due to morbidity or input prices. Plans should be at risk for factors that they can potentially manage, such as practice styles or moral hazard.[1] In actuality, geographic variations are due to both legitimate and objectionable factors (e.g., Fisher et al., 2003). The resulting policy trade-off can lead to residual incentives for selection on geography.

Geographic variations are pervasive in many settings, and regulating this specific selection-efficiency trade-off is a recurring concern. In the US, geographic variations in health care spending have been well established and recognized by regulators (Dartmouth Atlas Working Group, 2011). In the Medicare

[1]See also van de Ven and Ellis (2000) who distinguish factors for which solidarity is desired and which should be cross-subsidized (such as morbidity) and others for which solidarity is not desired (such as negotiating power, ability to selectively contract, and regional characteristics such as provider infrastructure).

Advantage (MA) program, private health plans receive a risk-adjusted capitation payment to assume the costs associated with providing benefits covered by the traditional fee-for-service (FFS) Medicare. The capitation payment varies according to the enrollee's risk (based on a range of disease conditions) as well as the county-specific base rate, which is a function of historical spending in the FFS program. The differences in the county rates are substantial. As illustration, in 2011 the unweighted MA rates per "aged" beneficiary ranged from $729 in Des Moines county, IA, to $1,505 in Saint-Bernard, LA (KFF, 2011). Geography also features in insurance programs for the non-elderly population. The Patient Protection and Affordable Care Act (PPACA) of 2010 requires the use of risk adjustment for the individual and small-group markets both inside and outside the state health insurance exchanges. The law does not stipulate whether a geographic adjuster should be included but explicitly suggests Medicare Advantage's methodology as a model for adjustment in the exchanges. Similar to the Massachusetts health insurance exchange, PPACA also requires adjusted community-rating, allowing limited premium variation based on a number of factors, including rating areas. The risk adjustment and rating areas therefore require definitions and policy decisions on geography. In the context of MA and the exchanges, geography is mostly discussed as policy instrument to encourage plan entry (Mcguire et al., 2011). However, plans may also exploit any within-area differences and mismatches between rating and actual market areas. The threat of geographic risk selection is particularly serious since additional rules, including the remaining allowed rating factors, could artificially generate heterogeneous premium groups within rating areas and induce selection even in presence of complementary regulation (Pauly, 1984).

Establishing geographic cream-skimming is ultimately an empirical question. The German Social Health Insurance (SHI) provides a useful context to identify selection on geography in a heavily supervised environment. Health plans in the SHI are not allowed to collect medical histories as part of the enrolment process, and they cannot refuse any applicant or vary premiums, benefits or provider networks. As in MA and the insurance exchanges, payments to plans are adjusted for the morbidity of enrolees. However, geographic variations remain a source of heterogeneity and a motive for cream-skimming. The risk adjustment system accounted for East/West differences until 2007, but a 2009 reform explicitly excluded geography from the payment formula. In an opinion on the reform, the German Constitutional Court recognized the existence of spatial variations in costs and their financial implications for plans (BVerfG, 2005). However, it argued that legitimate variations due to morbidity are sufficiently compensated by the new formula and that plans should face incentives to actively manage variations due to regional inefficiencies or patient preferences. As consequence of this policy, health plans have incentives to exploit geography to improve their risk structure and financial standing.

The aim of this paper is to assess empirically whether plans act on the prevailing financial incentives to select on geography by focusing their recruitment efforts on applicants from low-cost areas such as East Germany. To separately identify cream-skimming from potentially concurrent demand-side adverse selection, I implement a double-blind audit study in which health plans are presented with fictitious applicants who have different addresses but are otherwise identical. I measure response rates for letters, emails, and phone calls, as well as the weight and stamp value of letters as proxies for insurers' resource expenses. The findings indicate that plans are more likely to respond to applicants from East Germany, a result consistent with cream-skimming even in this tightly regulated setting. The paper also highlights the value of the audit approach for examining firm behavior and is, to my knowledge, the first audit study of selection behavior by health insurers.

2 The German context

2.1 The Social Health Insurance

The SHI is Germany's main insurance system, covering about 90 percent of the population. The remainder, mainly high-income earners and civil servants, can substitute private coverage for social health insurance. Health plans in the SHI, so-called sickness funds, provide a largely regulated benefits package, and are subject to uniform, SHI-wide funding and contracting modalities. Historically, fund membership was limited by occupation groups, guilds or companies; large general regional funds (*Allgemeine Ortskrankenkassen*) provided insurance for individuals not falling within these categories. Since the introduction of competition for enrolees in 1996, the number of funds has decreased dramatically as result of mergers (McGuire and Bauhoff, 2007). At that time, funds could choose to accept members from outside their historical base, or remain closed, an option exercised mostly by company-based funds. They could also choose to operate only in certain regions or nationwide. Once opened, funds must accept all eligible applicants within their geographic market.

In January 2008, 221 sickness funds operated in the SHI, 61 of which were open and available nationwide.[2] Funds must contract with all accredited providers and have little opportunity for selective contracting. They may not decline coverage or risk-rate premiums. In this restricted market, funds mostly compete on customer service quality and on minor variations in additional benefits that are allowed, e.g. coverage of homeopathic therapy. Since April 2007 plans may also offer new models of care, including bonus and high-deductible plans, and must offer integrated care programs and gatekeeper models.

The SHI is financed by contributions from members and, to a smaller extent, by general revenues. Since January 2009, premiums consist of an income-related contribution and a supplemental fee. All members contribute a uniform rate of 15.5 % on monthly wage earnings up to 3,675 Euro, or a maximum premium of 570 Euro.[3] Dependants are covered without additional costs to individual members, and SHI members can switch sickness funds every 18 months or when a fund raises its premium. A central collection agency, the health fund (*Gesundheitsfond*), pools the contributions and pays a risk-adjusted

[2]48 of these 61 are historically company-based funds. The social health insurance had about 70 million insured of which 51 million were members and 19 million dependants. General regional funds (AOK) covered about 24 million people, company-based funds (BKK) 14 million, guild-based funds (IKK) 6 million and substitute funds (EK) 23.6 million (BMG, 2009).

[3]About half of of the contribution is paid by employers. Retirees pay contributions based on their pension income (paid in equal share by the retirees and the pension fund). Unemployed and welfare recipients receive full or partial assistance from the relevant government agencies.

capitation to the funds. Funds are constrained to make annual losses or profits within a narrow band, and must raise community-rated, supplemental fees from their members if their expenses exceed payments from the health fund. In the event of excess profits they may provide refunds to their members.

2.2 Risk adjustment and residual geographic variation in costs

The SHI has employed a prospective risk adjustment system since 1994 and augmented the adjustment formula in January 2009. It now accounts for 80 diseases, age, gender, and indicators of whether the person receives a disability pension and is enrolled in specific disease-management programs.[4] Medicare follows a similarly comprehensive approach to risk adjustment in the Medicare Advantage program since 2004 (Brown et al., 2011). In the SHI, however, the absence of geographic adjusters leads to uncompensated variation due to spatial differences in revenues, prices and quantity of care.

The central collection mechanism effectively equalizes the contributions that sickness funds receive and in this way eliminates incentives to select high-income members in order to collect a large absolute premium. However, the structure of a funds' member base still matters for its revenues because of income-related caps on supplemental fees and copays. In particular, a fund must raise supplemental fees from its members if revenues from the central pool are insufficient. These fees are capped at one percent of income so that funds with relatively poor members must raise a higher supplemental fee for a given revenue shortfall.[5] Similarly, patient copays are subject to a cap of two percent of household net income and one percent for chronically ill. Moreover, the fees are per-member, rather than per-insured, which limits contributions from households with many dependants. As result, large and low-income households generate less total revenue. This variation is correlated with geography but not compensated by the risk adjustment system (BVerfG, 2005; Jacobs et al., 2002).

Regional variations in provider prices also affect the relative attractiveness of otherwise comparable households. The corporatist structure of the SHI requires that funds and providers negotiate as groups and largely prevents individual funds from leveraging their market power in a given region. Moreover, the difference in how geography is treated in the price schedule and risk adjustment creates a distinct mismatch between the sickness funds' revenue and expenditure structures. The majority of inpatient and outpatient payments are based on relative value scales and base rates that are initially negotiated on the federal level and may be adjusted for individual states. Hospitals are paid through a DRG system that, since 2010, uses state-specific base rates.[6] Similarly, the base rate for physician payments can vary

[4] The diseases may be further distinguished by severity, leading to 106 morbidity indicators and 152 risk groups overall (BVA, 2008). A high-risk pool providing risk-sharing for extremely expensive patients was eliminated by the reform.

[5] The fees are uniform but a high fee will be capped for poorer members, leading to a differentiated fee by income. Fees up to 8 Euros are not subject to the cap. In case of the introduction or increase in the supplemental fee, patients have the right to switch funds before the otherwise mandatory 18 months of membership are completed. The supplemental fee is fully paid by the members, unlike the general contribution which is split with the employer. News reports suggests that consumers are very responsive to supplemental fees: one sickness fund lost one-third of its members within a quarter after introducing the fee (Spiegel Online, 2010).

[6] The state rates are set to converge to -1.25 to +2.5 percent of a national base rate between 2010 and 2014. Sickness

by location, and the physician fee scale may be adjusted on the state level in case of special cost or needs structures, such as certain preventive care services.[7] The geographically differentiated inpatient and outpatient price schedules conflict with the lack of regional factors in the risk adjustment formula.

Some geographic variations in utilization also persist, since the central pool equalizes member contributions across funds but does not adjust for differences in spending by income group. Member incomes could matter for the funds' expenditure profile in two potentially offsetting ways. First, the income-related cap on copays lowers the price of care once the limit is exceeded, potentially leading to increased utilization for poorer members. Second, income and socio-economic status may be associated with differential utilization, conditional on morbidity. In addition, utilization may vary due to practice and referral patterns.

A particularly salient geographic delineation in costs is between East and West Germany. These regions had separate risk adjustment systems until 2007, when the standardized costs per capita in West Germany were 5.2 percent above those in East Germany (or 2,012 versus 1,912 Euro annually, using the pre-2009 adjustment formula (Goepffarth et al., 2010)). Table 1 shows East/West differences in several cost measures related to revenue, price and utilization.[8] These indicators suggest that, on average, West Germans are financially less attractive to funds because prices and utilization may be higher, although the higher potential revenues from supplemental fees in West Germany could counter-balance these incentives. Andersen and Grabka (2006) show that between 2000-2004, the average applicant (switcher) also tended to be a better risk than the average SHI member, and that East German switchers are consistently better risks than their West German counterparts. Switching rates are comparable over this period.

3 Sickness funds and cream-skimming

Although sickness funds are non-profit institutions, they are economic actors. In particular, while profits cannot be a direct motivation for funds, their financial position indirectly drives the objective of growth in membership.[9] Studies of switchers within the SHI suggest that consumers are price-responsive (e.g., Andersen and Grabka, 2006), so that growth is largely a function of the fund-specific supplemental fee and

funds finance recurrent costs, while state governments are responsible for funding investments in hospitals.

[7] From 2011 point values can further be modified on the district-level if a local area is classified as under- or over-served. Outpatient providers are organized in 17 regional associations (*Kassenärztliche Vereinigungen*) that coincide with states (*Bundesländer*) except for the state of North Rhine-Westphalia which has two associations. Plans provide the associations with a prospective, morbidity-adjusted budget to cover most outpatient services. The associations allocate the payment to physicians according to the relative value scale and point values.

[8] In Germany, the existence of differences in costs across geographies is well-known although not systematically documented at all levels. For instance, recent news reports explicitly suggest that high provider densities in metropolitan areas like Hamburg and Berlin generate financial distress for sickness funds with disproportionate membership in these regions (Spiegel Online, 2010). The only current and systematic documentation of lower-level variations is discussed in the results section below.

[9] A survey of sickness funds in 2000 found that most firms aim to secure their survival and increase or maintain their membership (Haenecke, 2001). Managerial compensation is subject to public scrutiny. The highest-paid board member earned below 300,000 Euro in 2008 and most fund managers earned significantly less (BRH, 2009).

hence directly related to a funds' risk structure. Size matters to funds for two reasons. First, it secures the fund's survival (or standing in a merger), and provides prestige and political clout. Prior to the market-oriented reforms of early 1990s, fund size was used to determine staffing levels and management compensation. Second, although individual funds cannot exercise their market power in negotiations with providers, they can negotiate drug rebates since 2007. Empirically, sickness funds have been responsive to incentives generated by the payment system. For instance, the introduction of a temporary adjuster for disease management programs in 2002 led to a rapid increase in the availability of such programs (e.g., Brandt, 2008).

The explicit rationale for leaving sickness funds at risk for certain cost factors is to provide economic incentives for efficient management of providers and patients. However, sickness funds have few opportunities to actively manage price or quantity. Risk rating and utilization review are prohibited while only minor modifications to premiums, copays and deductibles are feasible and tightly regulated. Given these constraints on lowering costs by improving efficiency, an appealing alternative strategy for funds is to simply skim good risks. In fact, in reviewing the market conditions and basic risk adjustment in 2006, van de Ven et al. (2007) suggest that the potential profits from risk selection in the SHI are "very high". In principle, health plans can employ several non-price instruments to achieve or maintain a profitable risk structure. For instance, they can distort their benefit package and provider networks to induce favorable adverse self-selection (Frank et al., 2000). Plans can also dump bad risks or attract good risks, e.g., by discouraging bad risks from enrolling (Shen and Ellis, 2002a). Anecdotal evidence suggests that at least some sickness funds actively engage in selective marketing by advertising additional benefits that are only attractive to healthy individuals, such as discounted sports club membership (Hoeppner et al., 2005). Yet the scope for such selection activities is limited since the largely standardized benefits package mitigates opportunities for service-level selection and because refusing or dumping patients is illegal. Marketing and advertisement is strictly regulated (BVA, 2006).[10] Funds are subject to various forms of supervision and enforcement, including by the insurance regulator, the sickness fund association, the German consumer watchdog and the media. However, these institutions mostly react to complaints rather than initiating investigations, and punishment occurs mainly through public shaming and a lingering threat of legal action and fines. To my knowledge there have been no systematic investigations of risk-selection by these institutions.

I consider an important selection tool that remains available in the German context: the differential response to applicants. While funds cannot outright reject applicants, they can manipulate the probability of signing a potential member. Funds can delay their response in order to signal applicants that they have poor service quality, an important differentiation among firms in this market. They can also withhold contracts to directly prevent membership. Although this involves a risk of being detected, fined and publicly criticized, the limited statistical power of audits permits some degree of systematic non-response

[10]In 2009, funds could spend up to 8 Euro per member on marketing activities and presents. Funds can contract with external recruitment agencies and pay signing bonuses up to 20 Euro, in addition to at most 7.56 Euro compensation for administrative expenses.

in practice. Funds can also vary the intensity of their recruitment efforts by following-up with low-risk applicants and by employing various channels of communication (e.g., letters, emails, phone calls) that affect the probability of signing.

Geographic cream-skimming is profitable for funds. As discussed above, in 2007 an East German enrollee was about 100 Euro cheaper than a West German after controlling for basic risk characteristics. This compares to an average administrative cost of roughly the same amount (McGuire and Bauhoff, 2007). However, this value does not account for the risk of detection and penalties, or the retention rate that funds expect when responding to an applicant.

4 Experiment design

The objective of this paper is to assess whether funds systematically vary recruitment efforts according to applicants' risk, as indicated by their geographic location. To identify cream-skimming separately from potentially concurrent demand-side adverse selection I implement an audit study in the spirit of Bertrand and Mullainathan's (2004, henceforth BM) work on racial discrimination in the labor market. BM respond to newspaper job ads with fictitious resumes that have randomly assigned White or African-American sounding names, and test for discrimination by comparing callback rates for emails and voicemails. In the context of my experiment, funds receive website or email requests from fictitious applicants who appear identical but for their place of residence. I measure extensive and intensive recruitment effort by comparing callback and follow-up rates for letters, emails and phone calls.

In addition to eliminating demand-side confounders, this research design has further advantages for studying risk selection. First, the communication with the funds is impersonal which avoids imperfect matching of auditor pairs and interviewer bias. It is also relatively cheap. Second, the funds' financial position is directly determined by the expected risk of applicants, and the risk predictor, geography, is clearly observed. Also, the funds' response to the applicants' requests for contracts has immediate consequences on the probability of membership. An important distinction between my setup and BM's design is that BM can modify the variable of interest (names suggestive of race) while I cannot manipulate expected profitability directly and must rely on observing responses to different geographies. In addition, geography may not only be correlated with risk but also with other factors, such as consumer behavior and preferences. This complicates the unique attribution of variations in recruitment effort to selection behavior. I address this issue below by examining the extensive and intensive effort margins, and responses to different levels of geography.

In the experiment 47 large, nationally operating funds receive requests from up to 37 fictitious applicants. Together these funds cover about 49 % of all SHI members and 79 % of members in national, open and large funds. Conceptually, the data come from a series of experiments that are replicated for each fund. Funds observe the applicants' address and phone number (revealing the location) in addition to basic information, such as the name (indicating gender and German or Turkish ethnicity) and a short message that is unique to each fund-applicant combination. Each applicant has a different name, phone number, and email and postal address. I randomly assign names, message texts and the order of contacts. The addresses are from a convenience sample of actual postal locations. Funds received about three requests per day over two weeks in January 2009, from both East and West German applicants.[11] All requests were sent in the evening. For letters I calculate the response time as difference between the next work day after the request is sent and the date of the postal stamp of the response. For emails and calls I use the actual contact date and the day of the fund's reply, to account for immediate responses. I only use voicemails for which I can clearly identify the caller/fund. Durations are in work days (Monday-Friday); immediate responses are coded as zero days. I count responses within two periods, 4 and 9 work days, or effectively 1 and 2 weeks. This choice is not restrictive in practice since, conditional on any response, most replies arrive within the first days after the request is received. Appendix A provides further details on the setup and data collection.

Tables 2 and 3 show the characteristics of the applicant pool and the participating funds. Table 2 describes the sample characteristics and the unadjusted probability of receiving any request ($y > 0$). The sample size varies by mode because funds collect different information. In particular, all funds request the address and email, and 70 % also collect the phone number. For funds that do not collect phone numbers, "any" response includes only letters and emails. For funds that collect addresses, email and phone numbers, "any" is based on responses on any of the contact modes. The sample sizes for letters and emails differ slightly because of data cleaning described in Appendix A and because several mailboxes become unavailable as the experiment ends.[12] Similarly, the sample for responses within four work days is larger since for several applicants this is the maximum duration between the last request and the last day of the experiment, when the physical mailboxes become unavailable. The unadjusted outcomes in Table 2's lower panel show that the average response rate is high, about 90 % within 9 work days, while the rate for calls is low at 9-10 %. East Germans are more likely to receive letters and emails than West Germans, and they are slightly less likely to receive phone calls. Table 3 shows the characteristics of the funds contacted in the experiment.

While I focus on selection within an operating area, the decision of funds to operate regionally or nationwide is a related mechanism of geographic selection. Selective entry has been a concern also in the Medicare Advantage program where payment rates are explicitly used to encourage participation (Mcguire et al., 2011). To the extent that the funds in my sample have deliberately chosen to operate nationally, they may consider themselves relatively more adept at managing costs through better efficiency or risk selection. My results therefore apply to this specific sample of firms. I am not aware of a systematic study of sickness funds' initial choice of operating areas or their subsequent choice of counterparts for mergers.

[11] Not all funds are paired with each applicant since I evenly expand the sample to include smaller funds as the experiment progresses.

[12] The sample sizes for "any" and letters differ by one observation due to technical problems with the associated email account.

5 Methods

I estimate the differential response by funds to applicants in East and West Germany with a series of linear probability models (LPM). To account for the intensity with which funds pursue applicants I estimate the unconditional probabilities that funds provide any response ($y > 0$) and follow-up ($y > 1$).

$$I[y_{ifd} > i] = \beta_0 + \beta_1 West_{ifd} + \beta_2 X'_{ifd} + \delta_f + \epsilon_{ifd} \qquad \text{for } i = 0, 1$$

The dependent variable differs by mode and includes letters, emails, voicemails, and a combination of these three. β_1 is the coefficient of interest. The covariates X include indicator variables for male first names and Turkish first and family names. The fund fixed-effects δ_f capture general fund characteristics and the experiments' replications across funds. In an extension I replace the indicator for West Germany with the regional cost predictors discussed above. I also replace the fund fixed-effects with interactions of the West indicator with fund characteristics. Standard errors are clustered at the fund and district level (f and d), robust to heteroskedasticity and modified for small-samples (Cameron et al., 2011; Baum et al., 2010). Critical values are also adjusted accordingly.

6 Results

6.1 Main results

The main findings in Table 4 show that funds are less responsive to requests from the more costly West German applicants. The point estimates on the West German indicator are negative throughout, with the exception of phone calls. Despite the small sample size about half of the negative coefficients are statistically significant at standard levels and the overall results are strongly suggestive of a differential response.[13] Locations in West Germany are 2.7 percentage points (3.1%) less likely to receive any response within 4 work days, and are 1 percentage point less likely to receive a follow-up letter. West Germans are also less likely to receive responses within 9 work days. Within this time frame, they are 1.7 percentage points (1.9%) less likely to receive any response, and have a 2.8 percentage point lower probability of receiving a first letter. They are also 2.7 percentage points less likely to receive a follow-up email. The coefficient for follow-up on "any mode" is marginally insignificant. There are no statistically significant differences by gender which is compensated in the risk adjustment. There are also no differences for Turkish sounding names, although these tests may be particularly limited by the available sample size. The insignificant coefficient on first-emails could be due to undifferentiated auto-replies that are sent as the fund receives a request. The last line in each Table shows the p-values of a joint test of fixed-effects

[13] For most results the clustering increases the standard errors relative to a "robust" estimator with fund-fixed effects. In two noteworthy cases in Table 4 the clustered standard errors are lower. The robust s.e. for at least one response for "any" mode are 1.51 (remaining significant at 10%) and 1.33 (becoming statistically insignificant) for 4 and 9 days respectively. However, the robust s.e. for any follow-up in 9 days decreases to 1.90 (becoming significant at 5%).

for the (randomly assigned) weekday on which the requests were received. Joint tests on the message text are also statistically insignificant (not shown). Point estimates from supplemental Cox hazard models align with the LPM results and, for callback on "any" mode and letters are statistically significant. Logit specification give results similar to those of the above LPM models, as do LPM models without fund fixed-effects.

Although the East-West difference is particularly salient, the incentives for geographic cream-skimming persist on lower levels of geography. While there are no current studies of low-level geographic variations in Germany, I can classify locations as being in high or low-cost districts, using a study by the insurance regulator that described variation in residual costs, i.e., costs net of the basic adjustment (BVA, 2004).[14] Table 5 shows the results after this district-level measure is added to the specification. The observed response rates remain relatively lower for applicants from districts with high residual costs. The estimates for any callback, any letter and any phone call within 4 days are negative and statistically significant, in addition to the "any" callback within 9 work days. Most other estimates are negative, as expected. The point estimates for the West Germany indicator are statistically insignificant but are still negative and of roughly similar size as in Table 5. This finding is consistent with selection even on more disaggregated geographies, although these tests are subject to limited statistical power.

As discussed above, a potential alternative explanation for differential response is that funds allocate their recruitment efforts according to the anticipated behavior or preferences of consumers. For instance, funds may know from experience that West Germans tend to contact more funds than East Germans, so that investments in recruitment have a lower expected return. Related, consumers may have different tastes for marketing; some may perceive multiple mailings a nuisance whereas others find them persuasive. While the experiment design cannot resolve this attribution problem, two sets of supplemental evidence suggest that funds respond to expected costs rather than consumers types. First, sophisticated marketing strategies should primarily operate on the intensive margin of recruitment. Yet applicants from West Germany and high-cost districts are less likely to receive any callback despite the negibible cost of a minimum recruitment effort: for instance, the median stamp value of the first letters is only 1.45 Euro. The findings also indicate no variations in the response modes and times, indicating that funds do not substitute costly phone calls with cheaper letters or emails and do not adjust response times by "catching up" in the longer 9-day period. Also, the marketing material is not tailored by geography. The weight and stamp value of the first letter received by East and West applicants are equally likely to be above or below the sample median, as Table 6 shows.[15] Second, the district-level results in suggest that funds vary efforts also across fine levels of geography. However, while funds can plausibly estimate risk for very

[14]The regulator's study was primarily conducted to evaluate transfer payments between states. The calculations are based on a sample of fund expenditures from 2002, excluding dental costs, using the pre-2009 adjustment formula. Districts are assigned according to the residence of the insured. District costs are available as seven-step categorical classification from a map. I classify locations as high or low cost if they fall into the three most or least expensive groups, respectively. None of the districts falls into the middle category. A specification without the West Germany indicator shows similar point estimates for the high-district indicator, with somewhat higher statistical significance.

[15]The distribution of weight is heavily skewed and most stamp values are 1.45 Euro.

disaggregated geographies using their own claims data, consumer types are likely more amorphous and dispersed. The estimated effects for smaller geographies are therefore unlikely to be driven by responses to consumer types but rather expected costs. Similarly, auxiliary results discussed in the next section, provide suggestive evidence that funds are less likely to respond and follow-up with applicants from states with high hospital (DRG) rates.

Another competing hypothesis could be that funds over-allocate marketing resources to low-cost areas by mistake rather than strategy. This could result from legacies of mergers between regional funds that varied in their backoffice setup and staffing. However, most mergers occured in the mid-1990s and sustained misallocation would imply unprofitable inefficiencies that should fundamentally not arise. Moreover, the misallocations would need to occur systematically across funds, and match the regional and district geographic variations to plausibly drive the empirical results.

6.2 Additional results

I can only observe fund responses to geography and am unable to identify the actual parameters that funds use in their risk prediction and selection rules. They may simply use a geographic aggregation or employ a more sophisticated approach, e.g., by using the proxies for prices, income and utilization from Table 1. In fact, risk adjusters from other countries suggest that basic socio-economic and health care variables are useful correlates of expected costs at various geographic levels. For example the urbanicity factors in the Dutch risk adjustment are constructed from clustering postcode areas based on "urbanization, proportion of non-Western immigrants, average income, proportion of single persons, standardized death probability, proximity of hospitals and GPs and the number of nursing home beds (per 1,000 inhabitants within a radius of 25 km)." (MinVWS, 2008, p. 10). In 2005, these factors were associated with differences in capitation payments of up to 150 Euros or about 8 percent of the mean capitation payment (Douven, 2007). In a companion set of estimations not reported here (though available upon request) I replace the indicator for West Germany with the predictors in Table 1 one at a time, in separate regressions. Most predictors are statistically insignificant, although the point estimates are generally in the expected direction. For instance, higher hospital (DRG) rates and costs per hospital case tend to be negatively related with the probability of response. Similarly the mostly positive effects of the population-to-doctor ratio suggest that funds are more responsive to requests from areas with a low provider density. *A priori* the effect of household income is ambiguous since higher incomes are less likely to reach statutory caps but may have preferences for higher levels of utilization. Here I find that the point estimates are generally negative.

My primary specification includes fund fixed-effects since, conceptually, the data are generated by a series of experiments that are replicated across funds. As funds may vary in their ability and/or need to engage in cream-skimming it is instructive to examine whether the fund characteristics in Table 3 are associated with the differential response rate. I investigate this possibility by replacing the fund fixed-effects with interactions of the West indicator and the fund characteristics, again in separate regressions.

Most estimates are mixed in sign and not statistically significant (detailed results available upon request). Funds' contribution rate is insignificant with the exception of "any" response within four days. Fund size is not associated with differential responses within four days but larger funds are generally less likely to follow-up with West Germans within the nine days period. This is contrary to the notion that larger firms could be more attuned to the potential loss of reputation if caught selecting, and may be more likely to internalize potential regulator backlash (Finkelstein and Poterba, 2006). Possibly the objectives of small and large funds diverge, with the former focusing on growth in members while the latter are concerned about their risk profile. The growth in membership and type of fund have no clear association with recruitment effort. Finally, funds with larger office networks have a lower (though statistically insignificant) probability of responding to West German applicants, which could suggest that selection operates indirectly through varying recruitment efforts across local offices. Indeed, the negative point estimates for funds with more than one office are consistent with this explanation.[16]

7 Discussion

The German Social Health Insurance is characterized by uniform benefits and community-rated premiums, and a limited set of tools for sickness funds to manage care and costs. In this setting funds face incentives to cream-skim low risks to improve their financial performance. Importantly, funds can easily observe the geographic location of applicants which is correlated with variations in costs but remains unaccounted for in the risk adjustment system. This paper considers whether sickness funds in the SHI select on geography. I implement an audit study in which funds receive requests for contract materials from applicants across Germany, and find that, for some response modes, funds are less likely to respond and follow-up on requests from applicants located in areas that are financially less attractive, such as West Germany. This differential response is unlikely to be due to marketing that is tailored to consumer differences, and could operate through varying recruitment efforts across office locations. From a policy perspective, the current SHI approach to geography is not on the frontier of the selection-efficiency trade-off: since the funds are restricted in their ability to manage regional inefficiencies, there is no scope for gains from imposing geographic risk.

In this particular context, the welfare loss from selection is likely to be relatively small. Since funds are less likely to respond at all to high-cost applicants, the results imply that consumers choose among different sets of plans. Since the benefit package and provider networks are mostly identical across plans, this limited choice generates only small losses to consumers. In the worst-case scenario of full risk segmentation, high-cost applicants are stuck in plans that provide no supplemental benefits and charge the maximum supplemental fee of 1 % of income, i.e., up to 36.75 Euro per month at the income threshold of 3,675 Euro. Importantly, in Germany even shunned applicants maintain insurance coverage from their

[16] Ideally I would consider the funds' distribution networks. From the received letters I observe the senders' zip code; however I cannot infer the senders for applicants-fund combinations that did not generate letters.

previous sickness fund. In the US, Medicare beneficiaries have the traditional FFS program as fallback if they cannot gain access to a MA plan. This is not the case in the US non-elderly market where supply-side risk selection can lead to significantly higher welfare costs as individuals may be unable to obtain adequate risk protection. US plans may also be more aggressive than their German counterparts which are subject to limits on total profits.

The persistence of cream-skimming in a market as heavily regulated as the Social Health Insurance raises the question of additional strategies to contain selection by reducing the potential gains and available mechanism to select. First, the regulator could increase monitoring and penalties for this behavior, in effect raising the costs of skimming. Second, the risk adjustment could be enhanced by including geography, at the (efficiency) cost of also accounting for sources of variation that should be not compensated. In general, potential losses may be tolerable, particularly if individual access is given a large weight relative to welfare costs from entrenching inefficient provider and patient behavior (see also Glazer and McGuire, 2009). On a practical level, geographic adjustments involve a choice of the level of geography and the time horizon of adjustment. As Newhouse (1996) notes, the skewness of medical expenditures makes small-area adjustments unstable unless spending is averaged over longer time periods, as in the US Medicare program, trading off signal and noise. Third, the regulator could allow plans to use other tools to reduce costs, including the ability to negotiate with providers and to manage care. If plans can effectively reduce regional variations with these instruments, they may find geographic cream-skimming a relatively less appealing strategy. Finally, in the German context, some sources of regional variations could be eliminated by equalizing prices across regions and compensating funds for the income-related caps on the supplemental fee and copays.

An alternative strategy has been proposed as part of "managed competition" (Enthoven, 1993) where a "sponsor" serves as intermediary between patients and plans, and in this role also manages enrolment. Enthoven explicitly envisions that this setup prevents screening and selection of applicants by plans, and acts as clearinghouse for information and transactions. This approach could also be used by the US health insurance exchanges, which can provide applicants with marketing material and withhold applicants's details until they are signed by a plan. This leaves room for alternative selection strategies such as dumping but these may be more easily detectable.

The possibility of supply-side selection is also relevant to empirical research on demand-side selection in markets for health, long-term care and car insurance, and annuities (e.g., Chiappori and Salanie, 2000; Fang et al., 2008; Finkelstein and Poterba, 2006).[17] Observational studies of selection by consumers may need to account for concurrent actions by the insurer that are outside the pricing decision. For instance, Finkelstein and Poterba (2006) test for selection by considering whether individual characteristics that are not used in the pricing of annuities are nonetheless correlated with both demand for insurance and risk occurrence. As in the German SHI, these "unused observables" can result from prohibitively high transactions costs and regulatory restrictions or, as in the case of the UK annuity market, from voluntary

[17] Cutler and Zeckhauser (2000) discuss the theory and evidence of demand-side adverse selection in health insurance.

restraint by insurers. The possibility of concurrent supply-side selection on non-price margins could have important implications for these tests and provide a rationale for why firms forego valuable information in their pricing decisions.

References

Andersen, H. H. and M. M. Grabka (2006): "Kassenwechsel in der GKV 1997 - 2004. Profile - Trends - Perspektiven," in *Jahrbuch Risikostrukturausgleich 2006*, ed. by D. Goepffarth, S. Gress, K. Jacobs, and J. Wasem, 145–189.

AOK (2010): "Uebersicht ueber die fuer 2010 gueltigen Landesbasisfallwerte in den einzelnen Bundeslaendern, Stand 13.7.2010," AOK-Bundesverband.

Baum, C. F., M. E. Schaffer, and S. Stillman (2010): "ivreg2. Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression," http://ideas.repec.org/c/boc/bocode/s425401.html.

Bertrand, M. and S. Mullainathan (2004): "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *The American Economic Review*, 94, 991–1013.

BMG (2009): "Gesetzliche Krankenversicherung: Mitglieder, mitversicherte Angehoerige, Beitragssaetze und Krankenstand Monatswerte Januar-Dezember 2008. Stand: 8. Januar 2009," Bundesministerium fuer Gesundheit.

Brandt, N. (2008): "Moving Towards more Sustainable Healthcare Financing in Germany," *OECD Economics Department Working Papers No. 612.*

BRH (2009): "Bemerkungen 2009 zur Haushalts- und Wirtschaftsfuehrung des Bundes," Bundesrechnungshof.

Brown, J., M. Duggan, I. Kuziemko, and W. Woolston (2011): "Will risk-adjustment decrease health care costs? New Evidence from the Medicare Advantage Progam," Working Paper, Princeton University.

BVA (2004): "Regionale Verteilungswirkungen des Risikostrukturausgleich," Bundesversicherungsamt.

——— (2006): "Gemeinsame Wettbewerbsgrundsaetze der Aufsichtsbehoerden der GKV vom 19. Maerz 1998 in der Fassung vom 9. November 2006," Bundesversicherungsamt.

——— (2008): "So funktioniert der neue Risikostrukturausgleich im Gesundheitsfond, Stand 16. September 2008," Bundesversicherungsamt.

BVerfG (2005): "2 BvF 2/01 vom 18.7.2005, Absatz-Nr. (1 - 287)," Bundesverfassungsgericht.

BZG (2008): "dfg Ranking. Liste der deutschen Krankenkassen, 21. Februar 2008," *Beitraege zur Gesellschaftspolitik*, 01-08.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference With Multiway Clustering," *Journal of Business and Economic Statistics*, 29, 238–249.

Chiappori, P. and B. Salanie (2000): "Testing for asymmetric information in insurance markets," *Journal of Political Economy*, 108, 56âĂŞ78.

Cutler, D. M. and R. J. Zeckhauser (2000): “The anatomy of health insurance,” in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, vol. Volume 1, Part 1, 563–643.

Dartmouth Atlas Working Group (2011): “Dartmouth Atlas of Health Care,” Dartmouth Institute for Health Policy and Clinical Practice, www.dartmouthatlas.org (accessed Nov 10, 2011).

DESTATIS (2008): “Krankenhaeuser: Einrichtungen, Betten und Patientenbewegung 2008,” Statistisches Bundesamt Deutschland.

——— (2009): “Kostennachweis der Krankenhaeuser,” Statistisches Bundesamt Deutschland.

Douven, R. (2007): “Morbidity-based risk adjustment in the Netherlands,” in *Wettbewerb und Risikostrukturausgleich im internationalen Vergleich*, ed. by E. Wille, V. Ulrich, and U. Schneider, Baden-Baden: Nomos.

Enthoven, A. C. (1993): “The History And Principles Of Managed Competition,” *Health Affairs*, 12, 24–48.

Fang, H., M. P. Keane, and D. Silverman (2008): “Sources of Advantageous Selection: Evidence from the Medigap Insurance Market,” *Journal of Political Economy*, 116, 303–350.

Finkelstein, A. and J. Poterba (2006): “Testing for Adverse Selection with "Unused Observables",” *NBER Working Paper*, 12112.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder (2003): “The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care,” *Annals of Internal Medicine*, 138, 273–287.

Frank, R. G., J. Glazer, and T. G. McGuire (2000): “Measuring adverse selection in managed health care,” *Journal of Health Economics*, 19, 829–854.

Glazer, J. and T. McGuire (2009): “Optimal Risk Adjustment,” in *The Elgar Companion to Health Economics*, ed. by A. M. Jones, Edward Elgar.

Goepffarth, D., S. Gress, and K. Jacobs (2010): *Jahrbuch Risikostrukturausgleich 2009/2010.*

Haenecke, H. (2001): “Unternehmensziele von Krankenkassen - eine empirische Analyse,” *Arbeit und Sozialpolitik*, 55, 27–33.

Hoeppner, K., S. Gress, H. Rothgang, J. Wasem, B. Bernard, and M. Buitkamp (2005): “Grenzen und Dysfunktionalitaeten des Kassenwettbewerbs in der GKV: Theorie und Empirie der Risikolektion in Deutschland,” *ZeS-Arbeitspapier*, 4.

INKAR (2007): “Indikatoren und Karten zur Raum- und Stadtentwicklung (INKAR),” Bundesamt fuer Bauwesen und Raumordnung.

Jacobs, K., P. Reschke, D. Cassel, and J. Wasem (2002): *Zur Wirkung des Risikostrukturausgleichs in der gesetzlichen Krankenversicherung. Eine Untersuchung im Auftrag des Bundesministeriums fuer Gesundheit. Endbericht*, Nomos.

KFF (2011): "Medicare health and prescription drug plan tracker. Local MA benchmark (unweighted) 2011," Kaiser Family Foundation, www.kff.org/medicare/healthplantracker (accessed Nov 10, 2011).

McGuire, T. G. and S. Bauhoff (2007): "A Decade of Choice: Tracking the German National Experience with Consumer Choice of Sickness Fund," in *Auf der Suche Nach der Besseren Loesung*, ed. by P. Oberender and C. Straub, Baden-Baden: Nomos, 145–159.

Mcguire, T. G., J. P. Newhouse, and A. D. Sinaiko (2011): "An Economic History of Medicare Part C," *Milbank Quarterly*, 89, 289–332.

MinVWS (2008): "Risk adjustment under the Health Insurance Act in the Netherlands," Ministry of Health, Welfare and Sport.

Newhouse, J. P. (1996): "Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection," *Journal of Economic Literature*, 34, 1236–1263.

Pauly, M. V. (1984): "Is cream-skimming a problem for the competitive medical market?" *Journal of Health Economics*, 3, 87–95.

Shen, Y. and R. P. Ellis (2002a): "Cost-minimizing risk adjustment," *Journal of Health Economics*, 21, 515–530.

——— (2002b): "How profitable is risk selection? A comparison of four risk adjustment models," *Health Economics*, 11, 165–174.

Spiegel Online (2010): "Hilfe, meine Krankenkasse geht pleite!" *Spiegel Online, April 15 2010 [in German].*

van de Ven, W. P., K. Beck, C. V. de Voorde, J. Wasem, and I. Zmora (2007): "Risk adjustment and risk selection in Europe: 6 years later," *Health Policy*, 83, 162–179.

van de Ven, W. P. and R. P. Ellis (2000): "Risk adjustment in competitive health plan markets," Elsevier, vol. 1 of *Handbook of Health Economics*, 755–845.

VdAK (2007): "Ausgewaehlte Basisdaten des Gesundheitswesens," Verband der Angestellten-Krankenkassen.

Warentest (2009): "Finanztest-Uebersicht aller waehlbaren Krankenkassen, Januar 2009," Stiftung Warentest.

Tables

Table 1: Summary statistics

	All of Germany		Experiment sites	
	East	West	East	West
State-level measures				
Members (1,000)	1,902	3,686	2,077	3,955
	(691)	(3,215)	(660)	(3,630)
DRG base rate (€)	2,829	2,924	2,837	2,919
	(21)	(79)	(12)	(86)
Costs per hospital case (€)	3,308	3,766	3,317	3,771
	(54)	(211)	(57)	(222)
Hospital cases per 100,000	23,117	21,996	22,988	22,222
	(1,445)	(3,177)	(1,635)	(3,430)
SHI members/pharmacy (1,000)	3.0	2.2	3.0	2.2
	(0.1)	(0.2)	(0.1)	(0.1)
Number unique states	5	11	4	8
District-level measures				
Population per doctor	775	685	680	471
	(178)	(187)	(336)	(125)
Hospital beds per 1,000	6.9	6.5	7.8	8.3
	(3.7)	(4.0)	(4.3)	(2.1)
Hhold disposable income (€)	1,190	1,505	1,189	1,587
	(49)	(160)	(55)	(212)
Unemployment rate (%)	19.3	9.9	18.0	12.4
	(3.4)	(3.4)	(3.7)	(4.0)
Number unique districts	112	327	9	13

Unweighted means and standard deviations, in parenthesis. Cost and utilization figures are not risk-adjusted. State-level base rates for 2009 DRG payments, without adjustments or cap limits. SHI members (July 2006); other state-level hospital data (2008); SHI members / pharmacy (2006). Population per doctor (2005); hospital beds (2005); household disposable income (monthly, 2005); unemployment rate (2006). The income measure includes high earners who can opt out of the SHI. For sources see Appendix Table A-2.

Table 2: Experiment sample and unadjusted outcomes

	East	West	Total
Sample sizes			
Number of fund-applicant contacts by mode			
Any mode (response $\leq$ 4 work days)	613	891	1504
Any mode (response $\leq$ 9 work days)	605	827	1432
Letters (response $\leq$ 4 work days)	613	892	1505
Letters (response $\leq$ 9 work days)	605	828	1433
Emails	613	933	1546
Calls	437	669	1106
Number of fictitious applicants †			
All applicants	15	22	37
Male	7	11	18
Turkish	2	4	6
Raw probability of any response (y>0), in %			
Response within 4 work days			
Any mode	88.42	85.41	86.64
Letters	78.14	75.67	76.68
Emails	29.53	28.94	29.17
Calls	7.09	7.32	7.23
Response within 9 work days			
Any mode	91.40	89.24	90.15
Letters	82.15	78.99	80.32
Emails	33.61	32.15	32.73
Calls	9.15	10.16	9.76

Total of 47 sickness funds; not all funds are paired with each profile. † *For largest possible dataset, actual sample may differ slightly by mode. See also Appendix A for details on data cleaning. Sample size for calls is lower because several funds do not request phone numbers in their contact forms.*

Table 3: Characteristics of funds in experiment

	Fund type			
	BKK	IKK	EK	Total
Contribution rate (%)	14.22	13.94	14.19	14.19
Growth in membership (%)	9.97	5.00	9.13	9.25
Contacted by email (%)	26	20	0	21
Contacted through website (%)	74	80	100	79
Funds with > 1 office (%)	88	60	100	87
Funds with above-median nbr of offices (%)	35	60	100	49
Fund size (1,000)	185	289	2,149	538
Number of funds	34	5	8	47

Unweighted averages. Not all funds are paired with each profile. Contribution rate (Dec 2008); fund size in members (Jan 2008); member growth (Jan 07 - Jan 08). BKK are company-based funds; IKK are guild-based funds and EK are substitute funds. For sources see Appendix Table A-2.

Table 4: Probability of response by sickness fund

	Any		Letters		Emails		Calls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	y>0	y>1	y>0	y>1	y>0	y>1	y>0	y>1
A. Within 4 work days								
West Germany	-2.73*	-2.49	-2.42	-0.96*	-1.23	-1.04	0.46	0.44
	(1.37)	(2.13)	(1.94)	(0.52)	(2.54)	(0.81)	(1.58)	(0.71)
Male	-0.36	0.32	-0.89	-0.67	0.27	0.24	2.60	-0.57
	(1.55)	(1.05)	(1.71)	(0.48)	(1.08)	(0.50)	(1.61)	(0.55)
Turkish name	-0.53	0.23	1.74	-0.40	1.21	-0.63	-1.24	1.07
	(2.35)	(2.19)	(2.46)	(0.91)	(1.78)	(1.03)	(2.08)	(1.44)
N	1504	1504	1505	1505	1546	1546	1106	1106
R2	0.29	0.49	0.40	0.10	0.51	0.63	0.34	0.29
Wald p: day of week FE	0.60	0.46	0.70	0.94	0.62	0.64	0.99	0.94
B. Within 9 work days								
West Germany	-1.68**	-3.91	-2.80*	-1.18	-1.83	-2.74*	1.31	0.01
	(0.80)	(2.39)	(1.49)	(0.83)	(2.71)	(1.35)	(2.06)	(0.91)
Male	-0.74	-0.26	-1.51	-0.94	0.37	0.43	2.43	-0.34
	(1.27)	(1.33)	(1.62)	(1.06)	(1.12)	(0.83)	(1.84)	(0.45)
Turkish name	-2.26	-1.16	0.57	-0.55	-0.33	1.46	-2.13	1.97
	(2.27)	(2.26)	(2.07)	(1.06)	(2.14)	(1.49)	(2.46)	(1.38)
N	1432	1432	1433	1433	1546	1546	1106	1106
R2	0.31	0.46	0.42	0.10	0.52	0.54	0.34	0.29
Wald p: day of week FE	0.79	0.29	0.86	0.35	0.90	0.81	0.88	0.92

** p<0.10, ** p<0.05, *** p<0.01 Units are percentage points. The dependent variable in specifications (1) and (2) is equal to 1 if the fund contacts a person once or more than once within 4 or 9 work days, respectively, and is 0 otherwise. The outcome variables in the next three pairs of specifications are for letter, email, and phone responses, respectively. OLS models with fund fixed-effects; standard errors clustered by fund and district, in parenthesis. Wald test on day of week FE from separate regression.*

Table 5: Probability of response by sickness fund to applicants in high or low-cost districts

	Any		Letters		Emails		Calls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	y>0	y>1	y>0	y>1	y>0	y>1	y>0	y>1
A. Within 4 work days								
High cost district	-2.71**	-1.97	-2.85**	-0.04	0.01	-1.19	-2.96*	-0.34
	(1.29)	(2.15)	(1.13)	(0.86)	(1.90)	(1.10)	(1.54)	(0.92)
West Germany	-2.16	-2.08	-1.83	-0.95	-1.23	-0.78	1.10	0.51
	(1.40)	(2.22)	(1.93)	(0.56)	(2.51)	(0.56)	(1.36)	(0.81)
Male	-0.18	0.45	-0.70	-0.66	0.27	0.33	2.82	-0.54
	(1.52)	(1.36)	(1.63)	(0.51)	(1.10)	(0.49)	(1.70)	(0.58)
Turkish name	-0.27	0.41	2.00	-0.40	1.21	-0.52	-0.97	1.10
	(2.39)	(2.05)	(2.92)	(0.90)	(1.84)	(0.80)	(2.23)	(1.48)
N	1504	1504	1505	1505	1546	1546	1106	1106
R2	0.29	0.49	0.40	0.10	0.51	0.63	0.34	0.29
B. Within 9 work days								
High cost district	-2.16*	-1.88	-2.16	0.03	-0.03	-1.35	-1.29	-0.63
	(1.16)	(2.14)	(1.47)	(1.26)	(1.84)	(1.09)	(2.11)	(0.86)
West Germany	-1.25	-3.54	-2.37	-1.18	-1.82	-2.45	1.59	0.14
	(0.89)	(2.45)	(1.53)	(0.90)	(2.68)	(1.49)	(2.08)	(0.86)
Male	-0.59	-0.13	-1.36	-0.94	0.37	0.53	2.53	-0.29
	(1.37)	(1.70)	(1.68)	(1.09)	(1.14)	(0.77)	(1.93)	(0.47)
Turkish name	-2.09	-1.02	0.74	-0.55	-0.32	1.57	-2.02	2.02
	(2.15)	(2.19)	(2.29)	(1.16)	(2.21)	(1.40)	(2.56)	(1.44)
N	1432	1432	1433	1433	1546	1546	1106	1106
R2	0.31	0.46	0.42	0.10	0.52	0.54	0.34	0.29

** p<0.10, ** p<0.05, *** p<0.01 Units are percentage points. Classification as high or low cost districts from BVA (2004). The dependent variable in specifications (1) and (2) is equal to 1 if the fund contacts a person once or more than once within 4 or 9 work days, respectively, and is 0 otherwise. The outcome variables in the next three pairs of specifications are for letter, email, and phone responses, respectively. OLS models with fund fixed-effects; standard errors clustered by fund and district, in parenthesis.*

Table 6: Probability that weight and stamp value of letters exceeds sample median

	Response within 4 days				Response within 9 days			
	Weight		Stamp value		Weight		Stamp value	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
West Germany	5.18	5.15	0.49	0.49	4.95	4.94	0.24	0.26
	(3.57)	(3.49)	(2.09)	(2.27)	(3.43)	(3.38)	(1.94)	(2.18)
Male		-0.44		0.33		-0.68		0.52
		(1.69)		(1.31)		(1.65)		(1.33)
Turkish name		0.98		-0.24		0.93		-0.86
		(3.17)		(1.40)		(2.77)		(1.62)
N	1122	1122	900	900	1116	1116	886	886
R2	0.58	0.58	0.36	0.36	0.57	0.57	0.32	0.32

** p<0.10, ** p<0.05, *** p<0.01 Units are percentage points. The dependent variable is equal to 1 if weight or stamp value exceeds the sample median, respectively, and is 0 otherwise. Stamp value not available for small number of letters. Median of total weight and stamp values are computed separately for response within 4 and 9 work days. Weight and stamp values are summed in the few cases where several letters arrived on same day. All models include fund fixed-effects; standard errors clustered by fund and district.*

A Experiment setup and data

A.1 Experiment setup

The fictitious applicants are provided with a real address, email and phone number. Addresses are provided by assistants selected from a snowball sample which restricts my ability to evenly balance or stratify the sample. Most importantly the addresses tend to be located in urban areas, particularly in West Germany, which also includes Berlin.

The first and family names for the fictitious applicants are randomly selected. The family names are from Wikipedia's list of the 100 most common surnames in Germany; the first names are from a list of the most popular names between 1957/8 - 2000 as provided by the *Gesellschaft für deutsche Sprache*.[18] I searched German news to identify Turkish first and family names that are common in Germany. In all cases the first and family names clearly differ from German sounding names.

The requests are made in the evening in Germany, mostly past 6pm. Each fund receives about 3 contacts per day, with a mix of applicants from high and low-cost areas. Most funds provide a contact form on their websites but some only allow contact by email; these differences across funds are captured in the fund fixed-effects. While I always provide an address, some funds do not collect phone numbers in their webforms and I do not provide a phone number in emailed requests. This generates the varying sample size by mode.

A.2 Data collection and outcome measures

Mail is collected at real postal addresses, while emails are sent to generic email accounts and phone calls are to virtual phone numbers with area codes to match the postal addresses. The voicemail greeting is a randomly assigned, impersonal prompt. The collection of emails and phone calls is automatic, while assistants at the postal addresses collect mail for the fictitious applicants. Regular test letters confirm that this collection operates properly. I delete all data from one West German location that had difficulties receiving mail for the fictitious applicant, and drop letter data for another West German applicant with general postal problems. Both deletions work against finding that West Germans have a lower response rate. I also delete very few fund-applicant combinations due to email problems, technical problems with the funds' websites etc. I flag letter data for two locations that reported minor concerns about general mail delivery, i.e., not restricted to the experiment. The results are robust to excluding these observations.

The date when funds receive the request for materials depends on the response mode. For letters I use the next working day (Mo-Fri) on which the request was received, since preparing and stamping letters is unlikely to occur at night. I use the actual contact date for emails and voicemails since funds could respond immediately using these modes. I use the postal stamp date to determine when the fund sent mailings; for seven letters this date is washed out by rain. I calculate the sent-date by subtracting two days from the received-date as noted by the assistants. While most letters have stamp values printed on the envelope, these values are missing for a small number of letters. Several of the emails may be automatic replies but these are difficult to identify as such since they may address the recipient by name, and close with the assigned agent's name. To avoid misclassifying messages, I count all emails from funds. I keep only calls if they are associated with a voicemail that clearly links them to a specific fund. Although the caller ID is often available to match calls that do not generate a voicemail, there may be a systematic pattern to suppressing the ID.

I measure the response in work days, Monday-Friday. A very small share of responses occurs on weekend days: 1.5 percent of emails (including immediate automatic acknowledgements) and 0.09 percent

[18] `de.wikipedia.org/wiki/Liste_der_häufigsten_Familiennamen_in_Deutschland` and `www.gfds.de/vornamen/beliebteste-vornamen/`, accessed November 23, 2008.

of calls occur on Saturdays or Sundays. 1 percent of letters are stamped on weekends. I move these responses forward to Monday as the next work day. The results are similar if the flagged or modified observations are excluded.

A.3 Accounting for postal delivery times

In order to determine the maximum response times (ultimately set at 4 or 9 work days, Mon-Fri) I evaluate the time between the last request to funds and the end of the experiment. Since I use the stamp date to measure fund's response and the physical mailboxes eventually become unavailable, I must allow for a reasonable postal delivery time. Table A-1 shows the distribution of delivery times for a subset of letters that were stamped within the first two weeks of the experiment. The maximum possible delivery time for this sample is 8 postal days (Mon-Sat, counting from the stamp date). As the table shows, about 98 % of letters arrive within two days, with no meaningful differences across East and West Germany. I set the maximum feasible response time for each applicant by allowing two postal days for delivery. This captures nearly all letters. Extending the delivery time to 3 postal days reduces the sample size slightly because I cannot use the very last requests: the experiment ends before the maximum response and delivery times are exhausted. This mainly affects the 9-day sample. Using 3 days for delivery does not change the quantitative and substantive results (not shown).

A.4 Covariates

Table A-2 lists the sources and dates of the covariates.

Table A-1: Postal delivery time

	Regions		Districts	
Percent delivered in:	East	West	Low	High
1 day	78.07	74.24	79.62	74.30
2 days	20.37	22.71	17.74	23.36
3 days	1.31	2.86	2.26	2.18
4 days	0.26	0.19	0.38	0.16
p-value	0.32		0.28	

In postal days, Mon-Sat. Based on 907 first letters from 35 locations. Letters stamped in first two weeks of experiment and recored as received by end of third week. p-value for Pearson's χ^2

Table A-2: Data sources for covariates

Covariate	Level	Year	Source
Fund membership	Fund	Jan 2008	BZG (2008)
Growth in membership	Fund	Jan 2007 - Jan 2008	BZG (2008)
Contribution rate†	Fund	Dec 2008	Warentest (2009)
Total number of offices	Fund	Dec 2008	Warentest (2009)
Costs per hospital case	State	2008	DESTATIS (2009)
Hospital cases	State	2008	DESTATIS (2008)
SHI members	State	July 2006	VdAK (2007)
Pharmacies	State	Dec 2006	VdAK (2007)
DRG base rates	State	2009	AOK (2010)
Hospital beds	District	2005	INKAR (2007)
Population per doctor	District	2005	INKAR (2007)
Household disposable income	District	2005	INKAR (2007)
Unemployment rate	District	2006	INKAR (2007)

† *The general contribution rate used here is shared equally between employer and employee. These rates are exclusive of an additional, uniform contribution of 0.9% that is paid solely by members.*