

Bauhoff, Sebastian; Fischer, Lisa; Göppfarth, Dirk; Wuppermann, Amelie C.

Working Paper

Plan Responses to Diagnosis-Based Payment: Evidence from Germany's Morbidity-Based Risk Adjustment

CESifo Working Paper, No. 6507

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Bauhoff, Sebastian; Fischer, Lisa; Göppfarth, Dirk; Wuppermann, Amelie C. (2017) : Plan Responses to Diagnosis-Based Payment: Evidence from Germany's Morbidity-Based Risk Adjustment, CESifo Working Paper, No. 6507, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/167493>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Plan Responses to Diagnosis- Based Payment: Evidence from Germany's Morbidity-Based Risk Adjustment

Sebastian Bauhoff, Lisa Fischer, Dirk Göppfarth, Amelie C. Wuppermann

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl

www.cesifo-group.org/wp

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

Plan Responses to Diagnosis-Based Payment: Evidence from Germany's Morbidity-Based Risk Adjustment

Abstract

Many competitive health insurance markets adjust payments to participating health plans according to their enrollees' risk – including based on diagnostic information. We investigate responses of German health plans to the introduction of morbidity-based risk adjustment in the Statutory Health Insurance in 2009, which triggers payments based on “validated” diagnoses by providers. Using the regulator's data from office-based physicians, we estimate a difference-in-difference analysis of the change in the share and number of validated diagnoses for ICD codes that are inside or outside the risk adjustment but are otherwise similar. We find a differential increase in the share of validated diagnoses of 2.6 and 3.6 percentage points (3-4%) between 2008 and 2013. This increase appears to originate from both a shift from not-validated toward validated diagnoses and an increase in the number of such diagnoses. Overall, our results indicate that plans were successful in influencing physicians' coding practices in a way that could lead to higher payments.

JEL-Codes: H510, I100, I130, I180.

Keywords: health plan payment, risk adjustment, managed competition, diagnostic coding, German Statutory Health Insurance.

*Sebastian Bauhoff**
Center for Global Development
2055 L Street NW
USA – Washington, DC
sbauhoff@cgdev.org

Dirk Göppfarth
State Chancellery of North Rhine-
Westphalia
Germany – 40190 Düsseldorf
Dirk.Goepffarth@stk.nrw.de

Lisa Fischer
German Federal Social Insurance Office
Friedrich-Ebert-Allee 38
Germany – 53113 Bonn
Lisa.Fischer@bvmf.bund.de

Amelie C. Wuppermann
Ludwig Maximilian University Munich
Faculty of Economics, Ludwigstrasse 33
Germany – 80539 Munich
amelie.wuppermann@econ.lmu.de

*corresponding author

March 21, 2017

Introduction

Competitive health insurance markets generally calibrate per-capita transfers to health plans based on the risk of their covered populations. In order to enhance the accuracy of payments, risk adjustment (RA) systems in the United States and many European countries have evolved from adjusting payments based on demographic factors to also include diagnosis-based morbidity indicators. However, these payment systems create a financial incentive for plans to report diagnoses that are included in the RA and trigger high payments (relative to resource costs). Plans can encourage coding that is appropriate (“right-coding”) or that unduly substitutes more generously-paid codes for less generously-paid ones (“upcoding”). The resulting change in coding patterns can lead to nominal changes in disease profiles (i.e., increased prevalence of certain diagnoses and/or increased severity) that do not reflect changes in actual disease patterns and severity.

A possible increase in nominal coding due to morbidity-based payments raises several concerns (Geruso and Layton, 2015; Kronick and Welch, 2014). First, in settings without overall budget cap, as in the US Medicare Advantage program and the Health Insurance Exchanges, increases in nominal coding and coding intensity that have no real basis can unduly increase government spending. In the context of Medicare, this concern has triggered repeated legislative adjustments to payments, e.g., via the Deficit Reduction Act of 2005; the Affordable Care Act of 2010; and the American Taxpayer’s Relief Act of 2012 (Kronick and Welch, 2014). Second, in contexts where RA is used to allocate a fixed budget, as in many European countries, such as in Germany, the Netherlands, Belgium and Switzerland, this behavior can generate inefficiencies by distorting the allocation of resources between competing health plans. Third, increases in nominal coding can change patient profiles, as codes are assigned to patients that lack an adequate basis for a diagnosis, or as patients with low severity are assigned to higher-severity codes. As consequence, over time the average costs for the affected diagnoses may fall, pushing down the payment associated with the specific risk adjuster. Because all plans would receive this lower payment, this effect can force all market participants to lower cost or increase revenue, potentially leading to undesirable behaviour such as risk-selection or reinforcing intensive coding. Finally, this behaviour can divert health plans’ attention from organising provision to engaging in rent-seeking. Plans that successfully manipulate coding may then use some of the additional earnings to distort consumer choices of plans, e.g., by offering premium rebates or supplemental benefits (Geruso and Layton, 2015).

In this paper we examine the impacts on coding of office-based physicians from the introduction of the morbidity-based RA in the German Statutory Health Insurance (SHI) in 2009. The “morbi-RA” replaced a

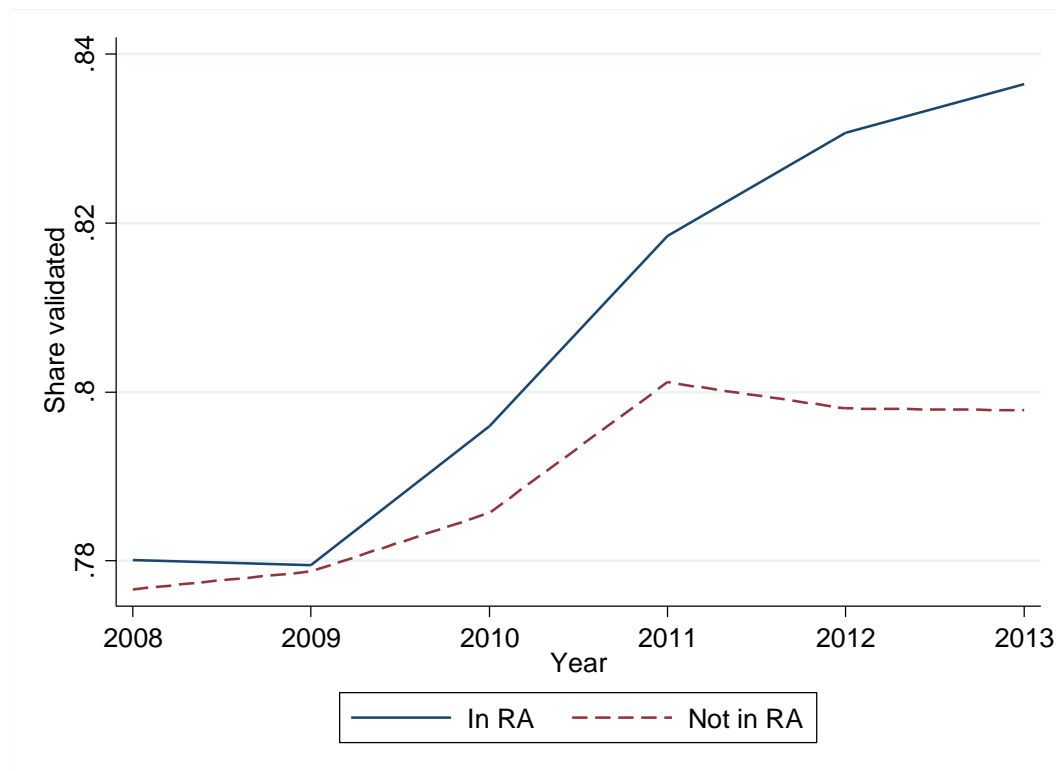
more basic system that adjusted for age, sex and disability-to-work status. The new system includes these parameters, as well as morbidity groups for 80 illnesses that are constructed based on ICD-10 diagnosis codes from hospitals and office-based physicians. Unlike in the US, German health plans are generally not allowed to own or operate health care facilities and contracting is mostly done collectively between the plan and provider associations. However, as described below, even within this heavily regulated environment, German plans have several ways to encourage physicians to adopt coding practices that are associated with (higher) payments through the RA.

We focus on a subtle payment-relevant feature of the German RA system, the designation of outpatient diagnoses. The German SHI's RA scheme only takes into account diagnoses made by office-based physicians if the latter have designated the diagnosis as "validated". Validation means that the physician is affirming the patient has the respective condition, as opposed to merely suspecting a diagnosis or recording an earlier diagnosis that is no longer relevant. In this paper, we examine changes in prevalence and count of diagnoses that are "validated" and hence taken into account by the RA scheme. We estimate the impact of the RA on the documentation of these diagnoses in difference-in-difference analyses on a random sample of administrative data used to execute the RA payments for the years 2008-2013. Specifically, we examine the change over time in the share and count of validated diagnoses at the level of an individual ICD code, for codes that were or were not part of the RA scheme. Our analyses are based on diagnoses submitted by office-based physicians who are not required to report validated diagnoses but who are required to mark each diagnoses as validated or not, and whose individual payments are based on procedures and not diagnoses codes.

Figure 1 previews our main finding that the average share of validated diagnoses increases faster for codes that are included in the RA scheme relative to those diagnoses that are excluded. The regression estimates indicate that the relative increase for these codes was 2.6 and 3.6 percentage points between 2008 and 2013. We also find that this effect is driven by both a shift from not-validated toward validated (payment-relevant) diagnoses and an increase in the number of such diagnoses. We further investigate differences in this effect across types of health plans and find that although this effect exists for most plan types, regional health plans may have experienced larger changes in coding than their competitors. This could indicate that the substantial and historical local ties of regional plans provide an effective means to shape physician coding practices, and may act as a substitute for explicit vertical integration in settings like the United States (Geruso and Layton, 2015). Our results are robust to excluding those codes and groups of codes that changed over time because of revisions of the RA system or the ICD catalog.

We argue that these effects are likely the consequences of nominal rather than real changes in morbidity, as the latter are unlikely to affect only payment-relevant codes (and should therefore be captured by our control group) and are unlikely to differ across plan types. Finally, we find no clear correlation between payments to plans for specific diagnoses and the change in the coding patterns, possibly because plans are unable to narrowly target specific diagnoses due to practical or legal constraints.

Figure 1: Share of validated diagnoses averaged across ICD codes



Notes: Based on 10% sample of claims submitted by health insurance plans to the German insurance regulator. Share validated calculated on level of ICD code and averaged across codes included in risk adjustment scheme (in RA) and excluded (Not in RA) for each year.

Research on plan responses to coding incentives in the US Medicare program has leveraged the fact that RA is only used for the Medicare Advantage (MA) component and not for the Fee-for-Service (FFS) component. FFS is used as control group to capture real changes in diagnoses that can be subtracted from the combined nominal and real changes in the MA diagnoses, after accounting for risk selection between MA and FFS. Using this approach, Kronick and Welch (2014) find that each year between 2004 and 2013, risk scores in MA rise faster than risk scores in FFS. They conclude that this rise in relative risk

score reflects changes in coding intensity rather than real increases in morbidity. Geruso and Layton (2015) investigate differences in coding intensities for FFS and MA, and among types of MA plans. They estimate that the relatively more intensive coding by MA plans generates risk scores that are 6 to 16% higher than they would have been in FFS. They also find that the risk scores are higher for MA insurers that are vertically integrated with providers, possibly because this makes it easier for insurers to influence providers' coding behavior.

A related literature on hospitals' responses to diagnosis-based payments has exploited the introduction of the diagnosis-related group (DRG) payment system or recalibrations in the payments of specific DRGs. Jürges and Köberlein (2015) examine how German hospitals responded to the introduction of DRG payments in 2003 by focusing on sharp thresholds for birth weight in DRG assignments that determine payments for preterm babies. They find that hospitals responded to the introduction of the birth weight thresholds by shifting newborns' reported birth weights from above to below the relevant thresholds, leading to DRGs with higher payment. Dafny (2005) studies US hospitals' responses to a recalibration of Medicare's DRG reimbursements in 1988. She investigates pairs of codes that are clinically similar but are associated with different payment amounts. Her findings suggest that the share of lucrative codes within a pair increased in the pairs' payment gap. She also finds that the response was primarily nominal (via coding practices) rather than real (changes in admission volumes and intensity of care). For the period after the 1988 change, Silverman and Skinner (2004) find a disproportionate increase in the prevalence of most generous codes for pneumonia and respiratory infections. A similar methodology has been used to document responses to changes in DRG payments by hospitals in Portugal and Norway (Barros and Braun, 2016; Januleviciute et al., 2016). Sacarny (2016) evaluates hospitals' responses to a 2008 reform that increased Medicare payments for claims that had detailed codes describing the patients' type of heart failure. He finds that hospitals were aware of the rewards to more detailed coding and responded accordingly. However, he also finds that the more lucrative coding diffused only slowly because physicians lacked incentives to change their coding practices.

Our analysis contributes a different perspective to the literature on plan responses, and in a context that differs from Medicare in important and informative ways. First, we highlight a novel approach to identifying coding responses induced by health plans by exploiting differential incentives of diagnoses that are included or excluded from the RA. Our approach is closely related to that used in studies of hospital coding behavior (Dafny, 2005; Sacarny, 2016) but – to our knowledge – has not yet been applied to health plans that do not directly control coding decisions. Second, we investigate changes at the level

of individual ICD codes. This is in contrast to most related work on coding by MA plans, which focuses on aggregate risk scores (an exception is Geruso and Layton (2015) who examine the transition of elderly in Massachusetts into MA or FFS). The analysis on the code level allows for controlling for time constant factors at the code level, as well as for explicitly accounting for changes in the coding system by excluding affected codes and groups of codes. By focusing on changes in designations (validated vs. others) within ICD codes, we analyze subtler changes than reclassification of patients to different diagnoses as in the research on hospital coding. The changes in designations are unlikely to affect treatment and do not directly affect physicians' pay, but nonetheless can generate substantial increases in payments to plans. Related, we observe changes over the transition from the demographic to the morbidity-based RA, which allows us to examine a major change in addition to later minor revisions within the morbidity-based system. A similar transition in Medicare in 2004 has been studied to examine risk selection by focusing on risk scores rather than diagnoses (Brown et al., 2014; Newhouse et al., 2012); and research on diagnosis coding for existing Medicare beneficiaries has not used data prior to the new RA. Our setting also facilitates a longitudinal difference-in-difference design that leverages data from before the change. - Fourth, the SHI operates as a single system so that there is no risk of selection between program components in the cross-section, e.g., as between MA and Medicare FFS. Our results also apply to the entire SHI population rather than a subgroup. Finally, we examine plan responses in a regulatory environment that is very different from the US and more similar to that of other European countries. For instance, US plans may be able to address principal-agent problems related to coding behavior through mechanisms not available to German plans, e.g., vertical integration (Geruso and Layton, 2015).

The remainder of the paper is structured as follows. In Section 2, we present additional information on the German health insurance system and describe relevant features of the morbidity-adjusted RA scheme, as well as coding incentives for hospitals and physicians, and how plans can influence them. In Section 3, we describe the administrative data used for our analysis, and in Section 4 we explain the estimation strategy. Results are presented in Section 5. Section 6 concludes with a discussion.

Background

The Statutory Health Insurance

The German SHI covers about 90% of the population.¹ The remainder is mostly covered by a separate private insurance system that restricts enrolment.² Since 1996, SHI operates as system of managed competition where members have free choice of health plans, or so-called sickness funds. In 2013, there were 137 plans (down from 212 in 2008), but not all operated nation-wide. Plans are self-administrated public institutions that offer broad coverage of benefits deemed medically necessary, including office-based physicians' treatment (including treatment by specialists), dental care, prescription drugs, hospital treatment and sick pay. Almost all benefits and the limited cost-sharing features are mandated by law. Generally, SHI members have free choice of provider but plans must offer primary care (gatekeeping) models and can combine them with financial incentives, such as reduced co-payments.

Contracting between plans and providers is mostly collective contracting, i.e., between the plans' associations and regional provider associations. There is some limited scope for selective contracting, mostly in the areas of disease management and integrated care. Payments for outpatient services are routed through the regional physician associations, which distribute payments to local office-based physicians.

The SHI is currently financed by uniform payroll tax of 14.6 % (*Beitragssatz*, literally "contribution rate"), split between employers and employees. The pension and unemployment insurance incur the employers' share for pensioners and unemployed members. The payroll tax and a subsidy from the federal budget are pooled in the Central Health Fund, which in turn pays risk-adjusted capitation payments to plans (Göpffarth and Henke, 2013). Plans are not allowed to make profits: they must cover expenditure gaps by raising supplemental contributions (*Zusatzbeitrag*) from their members, and return

¹ For additional details on the SHI see, e.g., Busse and Blümel (2014).

² The risk-rated private coverage is available primarily to higher-income earners and civil servants; the latter have strong financial incentives to select private coverage because of additional government subsidies. Switching rates between the systems have been further diminished by regulation.

excess revenues to their members, e.g., through premium rebates. In 2016, the average supplementary contribution was 1.1% of payroll, with a range of 0% to 1.9%.³

Although SHI plans are not allowed to make profits, they are nonetheless economic actors that are attentive to financial incentives and their financial position. Plans are keenly aware that their financial position is directly tied to their contribution rates and, as result, their ability to maintain and attract (healthy) enrollees. Consumers are responsive to premiums and supplemental fees (Schmitz and Ziebarth, 2016; Wuppermann et al., 2014) and shrinking plans may be at risk of bankruptcy or being the lesser partner in the wave of mergers that continues to-date. Plans are also responsive to the payment formula, as evidenced by their responses to temporary subsidies for disease management programs in the early 2000s (e.g., Brandt, 2008) and to incentives to exploit geographic variation in risk (Bauhoff, 2012).

Risk-Adjustment in the German SHI

The SHI introduced a RA system in 1994 in advance of allowing consumers free choice of health plans in 1996 (Buchner et al., 2013). The initial payment formula was based on age, sex, and disability-to-work-status as risk adjusters. The formula was updated in 2009 by adding morbidity indicators for 80 chronic and severe illnesses that are assigned to hierarchical morbidity groups (HMG; see Appendix A for details). Patients are allocated to HMGs on the basis of prior year inpatient and outpatient diagnoses. The allocation applies some diagnosis code-specific eligibility conditions such as age, sex, related pharmaceutical prescriptions, or hospital treatment. For outpatient diagnoses and secondary diagnoses from hospitals, two separate diagnoses of the same disease are required in different quarters of the year. In a few cases, pharmaceutical prescriptions are used to separate the severity of the condition. The cost data are taken from the current year while morbidity information is taken from the previous year

³ The way how health plans' contribution rates are framed has changed twice in the last decade. Until the end of 2008, each health plan had its individual contribution rate on income. In the beginning of 2009, a uniform rate was introduced, and health plans had to levy supplemental contributions or offer premium rebates in order to balance their budget. From 2009-2015, and thus also in the period that we study, these supplemental contributions and rebates were mainly cast in absolute values (i.e., 8 Euros/month) rather than as percent of payroll. In the remainder of the paper, we therefore refer to the supplementary contributions as "supplemental fees".

(prospective model). Unlike in Medicare, the SHI enforces a fixed budget that caps the level of aggregate payments to plans.

Office-based physicians provide the vast majority of diagnosis codes, about 95 % of the total (Table 1). As in Medicare FFS, German physicians have few incentives related to diagnostic codes: their reimbursement system is based on procedures rather than diagnoses. Plans and regional physician associations are jointly responsible for audits, which focus on the plausibility and legitimacy of reported procedure codes. Physicians must also report at least one diagnosis alongside any procedure but these are merely formal requirements.⁴

⁴ In contrast, German hospitals have their own coding incentives created by the comprehensive diagnosis-related group (DRG) system that may not generally coincide with the incentives facing health plans. Inpatient diagnoses do not carry designations.

Table 1: Descriptive Statistics

	2008	2009	2010	2011	2012	2013	All years
Number of individuals	7,321,764	7,275,762	7,327,604	7,279,705	7,244,172	7,274,782	7,287,298
Mean age	42.5	42.8	43.0	43.3	43.5	43.6	43.1
Share female	52.9%	52.9%	52.9%	52.8%	52.7%	52.6%	52.8%
Number of ICD codes							
# of codes in RA	3,655	3,702	3,741	3,703	3,743	3,743	3,715
# of codes not in RA	10,061	9,993	10,010	9,980	10,099	10,063	10,034
Number of diagnoses by ICD code							
Codes in RA							
Mean outpatient	11,203.04	11,639.96	12,074.16	12,411.84	12,544.12	13,098.45	12,166.23
Median outpatient	380	397	388	414	450	470	412
Mean outpatient validated	9,765.267	10,166.53	10,691.33	11,225.43	11,390.81	11,951.42	10,870.13
Median outpatient validated	273	292.5	293	330	360	387	321
Mean payment relevant	10800.47	11178.43	11758.57	12288.52	12493.09	13069.05	11,936.58
Median payment relevant	410	423	409	437	481	499	440
Codes not in RA							
Mean outpatient	13,417.87	14,061.82	14,117.34	14,080.21	14,387.92	14,982.47	14,175.07
Median outpatient	322	335	357	364	363	382	353
Mean outpatient validated	11,844.01	12,434.05	12,588.78	12,710.01	13,007.21	13,582.59	12,695.03
Median outpatient validated	241	252	269	285	280	294	270
Mean payment relevant	12362.46	12959.64	13161.74	13300.05	13639.37	14245.12	13,278.75
Median payment relevant	320	326	349	368.5	371	379	351
Share of validated diagnoses							
Codes in RA	78.01%	77.95%	79.60%	81.85%	83.07%	83.65%	80.70%
Codes not in RA	77.66%	77.88%	78.57%	80.12%	79.81%	79.79%	78.97%

Notes: 10% sample of claims submitted by health insurance plans to the German insurance regulator. Outpatient diagnoses include validated and not-validated diagnoses by office-based physicians. Payment relevant diagnoses contain validated outpatient as well as hospital diagnoses.

Germany's approach to measuring morbidity for use in the RA broadly resembles that used in other competitive health insurance markets, including the US and several European countries. For example, the RA systems used in Belgium, in Germany and in the US (Medicare Advantage and Health Exchanges) are based on in- and outpatient diagnoses. The Netherlands use a system based on inpatient diagnoses and pharmaceutical prescriptions only. In all countries, ways to capture and represent morbidity continue to evolve (Van de Ven et al., 2007; van Kleef et al., 2014).

Plans' incentives with regards to diagnostic codes from office-based physicians

The German RA varies payments based on the specific diagnosis codes, but only considers those outpatient diagnoses that are designated as "validated".⁵

Office-based physicians are required to designate their diagnoses as one of the following categories: validated (*gesichert*), suspected (*Verdacht auf*), symptomless condition post-diagnosis (*Zustand nach*) and excluded (*Ausschluss von*). In principle, a treatment episode should begin (in non-trivial cases) with a "suspected" diagnosis. A diagnoses becomes "validated" when it is established beyond doubt, e.g. by a lab test, a specialists' assessment or the physicians' own conclusions. Otherwise it becomes an "excluded" diagnosis. The designation may change over the course of an episode, e.g., as the physician confirms or rejects a suspected diagnosis. There are currently no coding guidelines that specify what is necessary to validate a diagnosis, so that the designation of a diagnosis will often be in the physician's discretion (KBV, 2011).

To increase the payout from risk adjustment, plans can aim to modify either of two features of the payment formula: the diagnosis code and the designation of the diagnosis. Inducing physicians to drastically change diagnostic information may be challenging for plans, and minor changes (e.g., within an ICD category, defined as the first three digits of the ICD code) will generally only lead to small changes in payment. By contrast, changing the designation of a diagnosis to "valid" implies a change from no

⁵ An additional payment-relevant designation applies to outpatient treatments in an inpatient facility, which are coded according to inpatient guidelines. As these guidelines prohibit code designations, health plans mark such codes as "special cases" (DIMDI, 2012). This mainly applies to physicians working at teaching and psychiatric ambulatory centers, and is rare in Germany with less than 0.1% of all outpatient diagnoses across all years of our sample. We omit these codes from our analyses.

payment to a positive payment in expectation.⁶ This will likely not lead to changes in the treatment that patients receive and thus will not trigger additional costs for the health plan. Thus, both the costs and benefits to encouraging changes to designations are favorable for plans.

SHI plans have few instruments available to align the coding practices of physicians with their financial interest in higher coding intensity. Vertical integration is not allowed, selective contracting is heavily constricted, and plans only have indirect financial relations with physicians as payments are routed through regional physician associations. There is, however, awareness among plan managers and the regulator that plans attempt to influence coding patterns with the goal of increased payments (BVA, 2015; Schmergal, 2016).

One approach is for plans to contract the regional associations to improve the coordination and management of patients with severe or chronic illnesses, e.g., through improved coordination of care, health education, home visits and consultation time (*Betreuungsstrukturvertrag*). Physicians are compensated for their participation outside their regular payments – as long as they record detailed diagnosis codes. As one example, in 2013 Berlin’s physician association informed its members of a contract with a regional plan that offered payments of 4 Euros to physicians reporting one of several hundreds of detailed codes provided in an annex, rising to 6 Euros for two codes; 8 Euros for three codes; and 12 Euros for four or more codes (KV Berlin, 2013). The contract explicitly states that physicians can only expect the additional payment if they report validated diagnoses for this plan’s enrollees. Physicians are furthermore instructed to change “suspected” or “condition post” diagnoses to “validated” ones, as long as some treatment occurred that aligns with a validated diagnosis. Another approach is to “educate” physicians about coding procedures, a practice that anecdotally has left some physicians feeling hassled by plan representatives (Schmergal, 2016). A third and more aggressive approach are requests for physicians to revise submitted diagnoses as part of a “claims review” process. The regulator has repeatedly intervened in the past years to reign in the latter two approaches (BVA,

⁶ As noted above, plans must show two separate diagnoses of the same disease in different quarters of the year in order to claim payment. As a result, it is lucrative for plans to also encourage providers to write two validated diagnoses for patients with no prior diagnosis, or an additional validated diagnosis for those patients who already have one.

2015). The regulatory agencies at the state and federal level have recently decided to tighten regulatory practice and the German federal parliament is discussing legal restrictions (Deutscher Bundestag, 2017).

Data

We conduct our primary analyses on a 10% random sample of administrative data used by the German insurance regulator (*Bundesversicherungsamt*) to construct the RA weights and to execute the payments to plans. Our dataset includes the yearly count of diagnoses that outpatient physicians and hospitals reported for the period 2008 to 2013 at the level of individual ICDs and classified using the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10). The outpatient counts are further disaggregated by the code designations. The data do not contain additional patient or provider identifiers, and we only observe the plan type (described below) rather than plan names. As the data lack patient identifiers and are only yearly counts of diagnoses, we are not able to investigate the number of validated diagnoses for specific conditions at the patient-level in different quarters of the year. Because payments from the RA to plans based on outpatient diagnoses require patients to have validated diagnoses in at least two quarters, we are further unable to evaluate whether any additional validated diagnoses actually result in higher payments. However, additional validated diagnoses will never decrease payments and generally lead to higher payment.

Although the new RA scheme was only operational in 2009, plans knew this implementation was coming and had incentives to manipulate diagnoses in 2008 already because payments to plans in year t are based on diagnoses from year $t-1$. However, the actual RA formula including the list of the 80 illnesses and associated ICD codes for 2009 was only published in the Fall of 2008 and there was considerable uncertainty regarding the specific illnesses and data sources, e.g., whether data from outpatient physicians would be used (Gaßner et al., 2010). Thus, while plans may have tried to influence coding of all conditions once they knew the RA was coming, the time span during which they could target the coding of diagnoses included in the RA for 2008 was quite limited. Furthermore, it may take some time for plans to develop and implement strategies to actually influence physicians' coding behavior. As the insurance regulator lacks reliable diagnosis data for a year prior to 2008, we have to rely on data from 2008 as the baseline year that we assume to be least affected by potential plan responses. Importantly, a coding response targeted towards payment-relevant codes already in 2008 would attenuate our results towards a null finding, while successful efforts to influence coding already in 2008 would not affect our results to the extent that all codes were affected equally.

We focus on two measures that are relevant for health plans: first, we construct the share of validated diagnoses on the ICD code level, relative to the overall number of designated diagnoses, i.e. validated, suspected, symptomless condition post-diagnosis, or excluded. Second, we focus on the number of outpatient diagnoses that count towards the RA scheme, i.e. diagnoses that are designated as “validated” by office-based physicians.

In addition to the share of validated diagnoses across all plans, we calculate this share separately for different types of plans. German plans are categorized based on their historical origins, with the groupings being regional funds (AOK, *Allgemeine Ortskrankenkassen*), company-based funds (BKK, *Betriebskrankenkassen*), guild-based funds (IKK, *Innungskrankenkassen*), substitute funds (EK, *Ersatzkrankenkassen*) and the miners’ fund (KBS, *Knappschaft*).⁷ Until 1996, most individuals in Germany were assigned to plans based on their employment. Although free choice of health plan was introduced in 1996, many plans have kept their historical organization in terms of office networks and regions within Germany that they cover until today. More specifically, many of the company-based funds only operate in a few of the currently 17 health insurance regions. In addition, they mainly have offices at the original company’s place of business. In contrast, the AOK plans that exist today grew out of a large number of small plans that each covered all individuals in a specific area who were not assigned to one of the other types of plans. Until today, the AOK type plans hold a large office network and strong presence in their regions. The EK and AOK categories had the largest enrollment in 2013, followed by the BKK and IKK funds (BMG, 2013).

Our initial primary dataset consists of a total of roughly 1.19 billion diagnoses from outpatient and hospital setting for over 43.7 million individuals over the six years. We cannot assign all observed diagnoses unambiguously to diseases that are either included in or excluded from the RA, as this assignment requires knowledge on patient sex and age in some cases. This ambiguity affects 904 code-year combinations that represent approximately 0.1% of diagnoses, which we exclude from our analysis. As Table 1 shows, our resulting analytic dataset contains information on around 13,700 ICD codes per year, of which around a quarter are included in the RA. The median number of outpatient diagnoses per

⁷ In addition, there is an agricultural plan type (*Landwirtschaftskrankenkassen*, LKK). The latter, however, do not participate in the RA scheme and are thus excluded from all analyses in this paper.

ICD code in these data fluctuates between 320-470 across the years and across those codes that are included in or excluded from the RA. The average number of diagnoses per code are a lot larger (they vary between 11,000 and almost 15,000), indicating that the distribution of diagnoses across codes is highly skewed; when analyzing the impact of the RA on the number of diagnoses we therefore use the logarithm of this variable as dependent variable. Table 1 also displays information on the mean and median number of validated outpatient diagnoses, as well as on the number of diagnoses that are payment relevant in the RA. The latter include both validated outpatient diagnoses and hospital diagnoses. Similar to the measure of outpatient data overall, the diagnoses of validated diagnoses and payment-relevant diagnoses are highly skewed. The unweighted average share of validated diagnoses for codes in the RA was 78% in 2008 and increased to 83.7% in 2013, while the share for codes outside the RA increased from 77.7% to 79.8%, again previewing our main results that the share of validated diagnoses increased faster for codes included in the RA compared to excluded codes.

For secondary analysis on whether the change in coding is correlated with plans' economic performance, we also construct three plan-specific measures based on two additional datasets: binary variables whether the plan imposed a supplemental fee or issued a rebate between 2009 and 2013, and a continuous measure of plans' enrollment in 2008 and 2012.

Information on the number of enrollees stems from an independent publisher (Müller and Lange, 2012). In this data source, enrollment is defined as paying members, not family members covered on their spouse's or parent's plan. Furthermore, the data contain information on plan specific contribution rates for 210 health plans for 2008, which compares to 212 health plans of the AOK, BKK, IKK, EK and KBS types that existed in 2008 according to official statistics (BMG, 2009); the plans with missing information cover less than 0.1% of those insured in the SHI. Due to an ongoing process of mergers in the health insurance market, the number of health plans shrank from 212 in 2008 to 137 in 2012 (BMG, 2013). For the year 2012, Müller and Lange (2012) contain information on enrollment for all 137 plans in the market. We augmented these data with information on supplemental fees and rebates from monthly records published in print and online by *Finanztest*, a publication by the leading German consumer safety group, *Stiftung Warentest*.

The two datasets are on the plan level: The first dataset contains information on the plans that existed in 2008, their contribution rate in 2008, and indicators for whether each plan levied a supplemental fee or issued a rebate between 2009 and 2013. In the construction of these indicators we take information on

mergers between plans into account, thus assigning to each 2008 plan its successors' supplemental fees or rebates. The second dataset contains information on enrollment in 2008 and 2012 for the 137 plans that existed in 2012, where again we take mergers across plans over time into account and construct total enrollment in 2008 for the plans that existed in 2012 based on all their predecessor plans. Due to missing information on enrollment in 2008, these data contain full information for 133 plans.

Methods

We implement a difference-in-difference analysis to estimate the change, over time, in the share and count of validated diagnoses using ICD codes that are included in the RA scheme as treatment group and those not included as control group. Our main estimation equation is:

$$y_{it} = \beta_0 + \beta_1 InRA_{it} + \delta' InRA_{it} \mathbf{t} + \boldsymbol{\eta}_i + \boldsymbol{\gamma}_t + u_{it} \quad (1)$$

Where y_{it} represents our outcome measures, i.e., either the share or the log number of validated diagnoses for ICD code i in year t . $InRA_{it}$ takes on the value 1 if code i is included in the RA scheme in t , and 0 otherwise; $\boldsymbol{\eta}$ reflects a vector of ICD code fixed effects that we include to capture differences across ICD codes that are constant over time $\boldsymbol{\gamma}$ represents a vector of year fixed effects that capture any changes across years which apply to all codes equally; and u_{it} is the error term. $\boldsymbol{\delta}$ is the (5x1) coefficient vector of interest. The term $InRA_{it} \mathbf{t}$ captures the interaction terms of the treatment group indicator, $InRA_{it}$, and indicator variables, \mathbf{t} , which indicate the different years after the RA scheme has been implemented, i.e., 2009-2013. The five coefficients included in $\boldsymbol{\delta}$ thus measure the increase in the share of validated diagnoses or in the number of validated diagnoses in each of the years 2009 to 2013 relative to the baseline year 2008. We estimate the coefficients in equation (1) by OLS and calculate standard errors clustered at the level of the ICD category, defined as ICD codes that share the first three digits. The clustering of the standard errors allows us to take into account a possible correlation of unobservable

shocks to the share or number of validated diagnoses across codes that belong to the same ICD category.⁸

To evaluate whether the actual payment that plans receive within the RA scheme affects their efforts to change coding, we augment this estimating equation with an interaction of the potential payment for the ICD code from the baseline year, $baselinepay_i$. This variable takes on the value 0 for all codes that are associated with diseases not included in the RA scheme, and the actual payment in the 2009 RA for the relevant HMG if the code belongs to a disease that is included in the RA scheme. We focus on the 2009 payment rates because the latter are plausibly exogenous unlike those in subsequent years: because there is an overall budget cap, any increase in volume would be offset by changes to payment rates. Increases in volumes of validated diagnoses in later years may thus have lowered payment rates. The baseline payment rate, however, was calculated before the introduction of the RA. Thus modified, the estimation equation is:

$$y_{it} = \beta_0 + \beta_1 InRA_{it} + \delta' InRA_{it} t + \zeta' InRA_{it} t baselinepay_i + \eta_i + \gamma_t + u_{it} \quad (2)$$

In the estimation of equation (2), we lose information on roughly 2,000 ICD code-year combinations compared to the estimation of equation (1) because we do not observe the baseline payment. Most of these codes were added to the ICD catalog after 2009, so that we cannot assign baseline payment without further assumptions. In some cases (85 codes), information on baseline payment is missing as we cannot assign diagnoses unambiguously to HMGs, which are the basis of payments. The assignment is not possible because it requires information on a patient's age, sex, or drug use, which is not available to us. We present results based on all diagnoses that we can assign unambiguously to HMGs. The results are quantitatively and qualitatively unchanged if we used the average payment over the different possible HMGs for each ICD code instead. Furthermore, we restrict the sample to codes that are always included in or always excluded from the RA scheme, as baseline payment should be most relevant for

⁸ Common shocks could, for example, stem from changes in the ICD catalog that affect specific ICD categories. We have also explored other levels of clustering, e.g. at the level of the disease instead of the ICD category, which did not affect our results.

these codes. This leads to an additional reduction of the sample by about 4,000 code-year combinations. Again, results are similar without this sample restriction.

For both models, our estimation strategy relies on the assumption that the share of validated diagnoses would have evolved similarly for codes included in and excluded from the RA scheme, had the RA scheme not been modified to include some of the codes. This assumption could be violated if the real prevalence increased differently across these two groups of codes, increasing the share of validated diagnoses in one but not the other. We mitigate this concern in two ways. First, we include fixed effects at the ICD code level, thus comparing changes within ICD codes over time. Importantly, this holds constant all code-specific factors that determine initial inclusion in or exclusion from the RA scheme. Second, we restrict our analysis sample to ICD codes that are associated with diseases with similar levels of chronicity before the introduction of the morbidity-based RA. Chronicity is defined as the share of patients with validated outpatient diagnoses of the disease in at least two quarters of the year relative to all patients with validated outpatient diagnoses. This measure is meant to capture the share of patients for whom a disease is chronic and was one criterion for inclusion of a disease in the 2009 RA (see Appendix A). By restricting our analysis to codes with baseline chronicity between 45-55% as measured in 2006 (the criterion for inclusion in the RA was 50%), we take into account that the development of real disease prevalence likely differs between chronic and acute diseases.

Related, our identification assumption could be violated if coding practices change in response to the RA not only for codes that are included in the RA but also for those outside the RA. One plausible reason may be learning and spillovers within physician offices: although plans may target specific sets of codes, physicians may change their designations for all codes they report. This would attenuate our findings toward finding no differential effect.

There are several other potential challenges to our identification strategy that we believe are not applicable in our case. First, to our knowledge there was no revision of the diagnostic or treatment guidelines for physicians that could have differentially affected ICD codes included or excluded by the RA. Second, at the aggregate level of the SHI, there is little risk of a sudden and substantial change in the risk profile, a concern in research on Medicare. Table 1 indicates that indeed, the share of women as well as average age stayed relatively constant over the 6 years of our analysis. In addition, any unobserved changes would have to differentially affect the included and excluded codes to bias our estimates.

In addition to the analysis for all plans, we investigate whether changes in validated diagnoses differ across different types of plans. For this, we estimate equation (1) as a fully interacted model for all types of health plans, i.e., we add to equation (1) the linear combination of all terms in the model interacted with indicator variables for four of the five types of plans. While the coefficients included in the original part of equation (1) then measure the effect for the reference type of plan (in our case the regional plans, AOK), the parts interacted with the indicator variables for the other plan types reflect the differences of the effects for each respective plan to the reference group, providing a direct test of the significance of observed differences across plan types. As switching rates are low across plans, and presumably lower across plan types (Schmitz and Ziebarth, 2016; Wuppermann et al., 2014), changes over time across plan types are unlikely to be driven by changes in risk pools across plan types.

We conduct two additional sets of analyses to assess the robustness of our findings to revisions of the ICD catalog and changes to the RA scheme. The ICD catalog is updated annually so that the number of ICD codes, as well as their meaning, can change over time. In addition, the insurance regulator revises the RA scheme every year, leading to changes in the list of illnesses that are included in the RA scheme, as well as changes in the association of ICD codes with illnesses. In our first robustness check, we restrict our analyses to individual ICD codes, as well as ICD categories, that did not change over time. In a second check, we limit our analyses to codes that are either always included in or always excluded from the RA scheme.

We also conduct secondary analyses on the potential impact of the changes in coding on the economic performance of plans. As noted above, plans cannot make a profit and may return some of the higher payments from the RA to consumers, e.g., in terms of rebates or a lower likelihood of supplemental fees that, in turn, could lead to increased enrollment. We use linear probability models to estimate the likelihood of rebates and supplemental fees over the period of 2009 to 2013, holding constant plans' contribution rate on income before the introduction of the RA (i.e., the contribution rate in 2008). We further estimate in a linear model whether enrollment on the plan level changed differently across plan types in 2012 compared to 2008. Because we can estimate the change in coding only at the plan-type level, we use indicator variables for the type in both sets of analyses.

Results

Table 2 reports the estimates from our primary analyses. Columns 1 and 2 report results for the share of validated diagnoses, columns 3 and 4 for the log number of validated diagnoses. Columns 1 and 3 are estimated on all ICD codes, while column 2 and 4 are estimated on codes with similar chronicity at baseline, i.e., a chronicity between 45-55%. As the negative coefficient estimates of the indicator for inclusion in the RA (inRA) in columns 1 and 2 indicate, the codes that were included in the RA had on average a lower share of validated diagnoses compared to the excluded codes in 2008. However, the point estimates are not statistically different from zero at conventional significance levels. The main coefficients of interest are the interaction terms between inRA and the years after the introduction of the RA (2009-2013). Each of these coefficients measures the difference in the change in the share of validated diagnoses across codes that are included in or excluded from the RA between the respective year and the baseline in 2008. The estimated difference-in-difference is about 2.6 percentage points between 2008 and 2013 based on the full sample (column 1) and, as previewed by Figure 1, the effect takes hold around 2011. One possible explanation for the lagged onset of the effect (mostly after 2010) is that plans could not anticipate which specific codes would be included in the RA until the Fall of 2008, when the details of the payment system were finalized. Instead of narrowly targeting the payment-relevant codes, they may have aimed at modifying coding behavior more generally, e.g., by providing general coding advice. Once the RA codes were revealed, targeted approaches became feasible but may have taken some time to implement, e.g. in the case of contracts between plans and physician associations that rewarded specific codes if they were validated. This staggering of plan actions could explain an initial increase in the share of validated codes for both, codes within and outside the RA that is also evident in Figure 1.

Table 2: Effect of inclusion in risk adjustment (RA) on coding of outpatient diagnoses

	Share of validated diagnoses		Ln number of validated diagnoses	
	(1) Full Sample	(2) Similar Chronicity	(3) Full Sample	(4) Similar Chronicity
In RA	-0.008 (0.006)	0.013 (0.017)	-0.012 (0.044)	0.110** (0.051)
2009 x in RA	-0.004 (0.003)	0.010 (0.010)	0.027*** (0.010)	0.018 (0.027)
2010 x in RA	0.008* (0.004)	0.013 (0.010)	0.068*** (0.016)	-0.004 (0.030)
2011 x in RA	0.014*** (0.004)	0.023** (0.011)	0.129*** (0.021)	0.019 (0.044)
2012 x in RA	0.019*** (0.005)	0.025*** (0.009)	0.169*** (0.024)	0.070 (0.051)
2013 x in RA	0.026*** (0.004)	0.036*** (0.010)	0.202*** (0.028)	0.080 (0.063)
Constant	0.779*** (0.002)	0.797*** (0.004)	5.684*** (0.014)	6.117*** (0.014)
N codes x year	82,493	9,143	81,460	9,073
N cluster	1,697	193	1,697	193
R2	0.775	0.821	0.988	0.990
Mean dep.	0.794	0.816	5.788	6.255

*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$.

Notes: Coefficients and standard errors clustered at ICD category in parentheses. All estimations include ICD and year fixed effects, (2) and (4) are restricted to codes assigned to conditions with chronicity between 45-55% before RA. Number of observations vary as codes with zero diagnoses are dropped when taking logs. Results are qualitatively similar when using a fixed effects Poisson model for the count of diagnoses including all available codes.

The observed pattern is similar for the sample of codes with similar chronicity, although the coefficients on each of the interactions are slightly larger. Columns 3 and 4 of Table 2 report the change in the (log) number of validated outpatient diagnoses for the two samples. The estimates indicate a relative increase in the number of these diagnoses of about 20% between 2008 and 2013 for the full sample (column 3) while there is no statistically detectable pattern in the restricted sample (column 4). Table 3 shows that these findings are robust to focusing only on ICD codes that do not change over time as well as subsetting to those ICD categories that did not change over time. Furthermore, the results hold when focusing only on codes that were always included in or excluded from the RA.⁹

⁹ Among the ICD categories that are always included in the RA scheme and had stable definitions of subcodes across all years, the largest increases between 2008 and 2013 in a simple-difference analysis (i.e., not relative to a control group) are for different types of cancers (ICD categories D42, C33, C12, and C68). The smallest changes were negative, for ICD categories that seem to have little in common (ICD categories U04, B45, B23 and Q00).

Table 3: Sensitivity Analyses: Effect of inclusion in RA on coding of outpatient diagnoses

	Share of validated diagnoses			Ln number of validated diagnoses		
	(1) Stable ICDs	(2) Stable category	(3) Stable RA	(4) Stable ICDs	(5) Stable category	(6) Stable RA
In RA	-0.007 (0.006)	-0.005 (0.007)	n.a.	-0.035 (0.048)	-0.018 (0.057)	n.a.
2009 x in RA	-0.004 (0.003)	-0.004 (0.003)	-0.005 (0.004)	0.027** (0.010)	0.024** (0.011)	0.028** (0.011)
2010 x in RA	0.008* (0.005)	0.009* (0.005)	0.009* (0.005)	0.071*** (0.016)	0.065*** (0.017)	0.073*** (0.017)
2011 x in RA	0.015*** (0.004)	0.016*** (0.005)	0.015*** (0.005)	0.134*** (0.021)	0.139*** (0.023)	0.145*** (0.024)
2012 x in RA	0.019*** (0.005)	0.018*** (0.005)	0.018*** (0.005)	0.165*** (0.025)	0.168*** (0.027)	0.174*** (0.028)
2013 x in RA	0.027*** (0.004)	0.027*** (0.004)	0.027*** (0.005)	0.197*** (0.029)	0.208*** (0.031)	0.212*** (0.033)
Constant	0.779*** (0.002)	0.788*** (0.002)	0.788*** (0.001)	5.672*** (0.015)	5.762*** (0.017)	5.730*** (0.006)
N codes x year	80,121	69,032	66,167	79,118	68,553	65,702
N cluster	1,660	1,458	1,396	1,660	1,458	1,396
R2	0.773	0.778	0.774	0.988	0.989	0.989
Mean dep.	0.794	0.805	0.806	5.768	5.866	5.838

*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$.

Notes: Coefficients and standard errors clustered at ICD category in parentheses. All estimations include ICD and year fixed effects. (1) and (4) restrict main sample to codes without changes in definitions over time, (2) and (5) further exclude ICD categories that see any changes over time, and (3) and (6) additionally restricts to codes that are either always included in or always excluded from RA. As assignment to RA does not vary in columns (3) and (6), coefficient of in RA dropped.

In Table 4 we investigate more closely the changes in the number of diagnoses for all outpatient diagnoses (columns 1 and 2), for the non-validated outpatient diagnoses (columns 3 and 4), and for all payment-relevant diagnoses (validated outpatient and all inpatient diagnoses, columns 5 and 6). We find that the morbidity-based RA is associated with a decrease in number not-validated diagnoses but an increase in the total number of outpatient diagnoses. The results for the combined inpatient and validated-outpatient diagnoses in columns 5 and 6 are very similar to those for the sample of overall outpatient diagnoses (columns 1 and 2) because of the quantitatively less important inpatient diagnoses. When we restrict the sample to codes that belong to conditions with similar chronicity (45-55%) before the introduction of the RA, however, none of these RA-related changes in the number of diagnoses remain significantly different from zero (columns 2, 4, and 6).

Table 4: Development of diagnoses counts over time

	<u>Ln number outpatient</u>		<u>Ln number not validated</u>		<u>Ln number payment relevant</u>	
	(1) Full Sample	(2) Similar Chronicity	(3) Full Sample	(4) Similar Chronicity	(5) Full Sample	(6) Similar Chronicity
In RA	-0.000 (0.040)	0.092* (0.050)	0.047 (0.029)	-0.110 (0.071)	0.012 (0.039)	0.136*** (0.048)
2009 x in RA	0.025** (0.011)	0.032 (0.029)	0.014 (0.013)	0.028 (0.042)	0.013 (0.010)	-0.006 (0.029)
2010 x in RA	0.058*** (0.015)	0.004 (0.029)	-0.014 (0.014)	0.039 (0.041)	0.051*** (0.016)	-0.022 (0.030)
2011 x in RA	0.113*** (0.021)	0.008 (0.043)	-0.033* (0.020)	0.000 (0.060)	0.091*** (0.020)	-0.015 (0.043)
2012 x in RA	0.148*** (0.024)	0.042 (0.051)	-0.044** (0.022)	-0.010 (0.072)	0.122*** (0.023)	0.024 (0.051)
2013 x in RA	0.175*** (0.028)	0.050 (0.062)	-0.051** (0.025)	-0.040 (0.084)	0.150*** (0.027)	0.022 (0.063)
Constant	5.901*** (0.013)	6.335*** (0.015)	4.376*** (0.010)	4.558*** (0.024)	5.939*** (0.012)	6.289*** (0.015)
N codes x year	82,493	9,143	75,763	8,515	81,910	9,094
N cluster	1,697	193	1,695	193	1,697	193
R2	0.989	0.991	0.978	0.978	0.990	0.991
Mean dep.	5.981	6.445	4.328	4.477	6.023	6.410

*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$.

Notes: Coefficients and standard errors clustered at ICD category in parentheses. All estimations include ICD and year fixed effects, (2), (4) and (6) are restricted to codes with chronicity between 45-55% before RA. Number of observations vary as codes with zero diagnoses are dropped when taking logs. Results are qualitatively similar to those after estimation of fixed effects Poisson models for counts of diagnoses for all available codes.

Together the results from Tables 2-4 indicate a relative increase in the share of validated outpatient diagnoses that are considered by the RA, i.e., diagnoses that are in expectation payment-relevant for health plans. This increase appears to originate from both, an increase in the count of validated diagnoses and a shift away from non-validated toward validated diagnoses.

Table 5 presents the estimation results based on equation (2), which includes an interaction with the payment plans receive. The results show no clear linear relationship between potential payments and the share or count of validated outpatient diagnoses. In additional analyses we investigate non-linearities by including indicators for deciles of baseline payment and find some evidence for non-linear effects for both outcomes, with the strongest response for payment-relevant codes in the mid-range of payment values, specifically the 6th decile (see Appendix Figure B.1 and Appendix Table B.1).

Table 5: Outpatient diagnoses and RA baseline payment

	Share of validated diagnoses		Ln number of validated diagnoses	
	(1) Full Sample	(2) Similar chronicity	(3) Full Sample	(4) Similar chronicity
2009 x in RA	-0.0049 (0.0037)	-0.0019 (0.0145)	0.0171 (0.0120)	0.0205 (0.0402)
2010 x in RA	0.0099* (0.0054)	0.0173 (0.0136)	0.0747*** (0.0181)	0.0086 (0.0429)
2011 x in RA	0.0161*** (0.0052)	0.0279* (0.0150)	0.1208*** (0.0253)	0.0679 (0.0597)
2012 x in RA	0.0213*** (0.0056)	0.0437*** (0.0104)	0.1484*** (0.0293)	0.1402** (0.0701)
2013 x in RA	0.0272*** (0.0048)	0.0498*** (0.0127)	0.1798*** (0.0340)	0.1370 (0.0880)
2009 x RA x baseline pay	0.0002 (0.0002)	0.0016 (0.0013)	0.0020** (0.0009)	0.0017 (0.0037)
2010 x RA x baseline pay	-0.0002 (0.0004)	-0.0004 (0.0010)	0.0008 (0.0011)	0.0021 (0.0032)
2011 x RA x baseline pay	-0.0002 (0.0002)	-0.0008 (0.0013)	0.0031* (0.0018)	-0.0057 (0.0051)
2012 x RA x baseline pay	-0.0003 (0.0004)	-0.0040*** (0.0008)	0.0029* (0.0018)	-0.0117 (0.0082)
2013 x RA x baseline pay	-0.0000 (0.0003)	-0.0026*** (0.0010)	0.0027 (0.0018)	-0.0087 (0.0081)
Constant	0.7798*** (0.0013)	0.8058*** (0.0027)	5.6756*** (0.0059)	6.0996*** (0.0109)
N codes x year	76,755	8,561	75,826	8,498
N cluster	1,620	161	1,620	161
R2	0.771	0.815	0.988	0.990

*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$.

Notes: Coefficients and standard errors clustered at ICD category in parentheses. All estimations include ICD and year fixed effects. Number of observations lower than in Table 2 because codes that switch in and out of RA as well as codes with missing or ambiguous assignment to baseline payment are excluded. Coefficient of "in RA" and of baseline payment dropped because these variables are constant over time in this sample

There are several possible explanations for the lack of a robust linear increase of the effect with the size of the payment. First, the structure of the RA payments may not be passed from the health plans to the physicians. For instance, as noted above, the payments specified in the 2013 contract between Berlin's physician association and the regional plan do not vary by diagnosis. Plans may shy away from structuring the financial rewards for physicians to closely reflect their own incentives to avoid detection and thus not offer higher payments for more lucrative diagnoses. Second, the response may be mediated by other attributes of codes that, e.g., affect how easy it is to "educate" or encourage physicians to revise the

designations of specific diagnosis. In Appendix Table B.2 we investigate heterogeneity with respect to three observable attributes of the disease (i.e., groups of diagnoses codes) for 2006: the chronicity, hospitalization rate and costs.¹⁰ The results do not indicate the presence of such heterogeneity.

In Table 6 we report estimates from the analysis of equation (1) by plan type, for the overall sample and again restricting the sample to codes with similar baseline chronicity. The AOK are the reference category against which changes for the other plans are measured. As mentioned above and also reported in Table 6, the AOK and EK are the largest plan types by enrollment, while the KBS has the smallest number of enrollees. The results indicate that for the AOK plans and starting in 2010, the share of validated diagnoses increases significantly more compared to 2008 when a code is included in the RA rather than excluded. The share increases less rapidly (relative to the AOK plans) for BKK and EK plans; the changes for the IKK and KBS are always negative compared to the AOK but not generally statistically different from zero. For the KBS the coefficients are significantly different from zero only in 2010 and 2013. For IKK, the coefficient for 2013 just falls short of statistical significance at the 10% significance level, with a p-value of .0106. The point estimates based on the restricted sample (column 2) are somewhat smaller in absolute value and the standard errors substantially larger; as result fewer of the estimated differences are significantly different from zero for this sample. Overall these results provide suggestive evidence that regional plans had a stronger coding response relative to other plan types. However, this evidence is not robust in the smaller sample and should be explored in more detail with data that contain individual plan identifiers.

¹⁰ Chronicity is defined as the share of patients that received at least one diagnosis of the condition in two different quarters of the year. Hospitalization is defined as share of patients that received at least one principal hospital diagnosis of the respective condition. Both of these measures are calculated based on diagnoses data from 2006 and published by the health insurance regulator. Costs are calculated for patients diagnosed with the condition in 2005 for the year 2006 (i.e., prospectively) and reflect the additional costs associated with the specific conditions based on claims data in 2006 holding constant other factors that determine payment from the RA, i.e., age, sex, and comorbidities.

Table 6: Heterogeneity by plan type in the share of validated diagnoses

	(1)		(2)	
	Full Sample		Similar chronicity	
AOK (2,412,219.5 enrollees) - reference group				
2008	-0.011	(0.007)	-0.003	(0.015)
2009	0.006	(0.004)	0.000	(0.013)
2010	0.018***	(0.005)	0.013	(0.012)
2011	0.024***	(0.004)	0.028***	(0.010)
2012	0.024***	(0.004)	0.031***	(0.011)
2013	0.033***	(0.005)	0.040***	(0.012)
BKK (1,273,134.6 enrollees)				
2008	0.005	(0.009)	-0.011	(0.014)
2009	-0.012*	(0.007)	0.008	(0.019)
2010	-0.016**	(0.007)	0.007	(0.017)
2011	-0.015**	(0.006)	-0.018	(0.015)
2012	-0.005	(0.006)	0.006	(0.017)
2013	-0.015**	(0.006)	-0.005	(0.015)
IKK (559,816.33 enrollees)				
2008	0.004	(0.010)	0.015	(0.025)
2009	-0.004	(0.007)	-0.006	(0.022)
2010	-0.009	(0.006)	-0.035*	(0.019)
2011	-0.008	(0.006)	-0.034	(0.021)
2012	-0.003	(0.007)	-0.034	(0.021)
2013	-0.010	(0.006)	-0.030*	(0.018)
EK (2,489,386.5 enrollees)				
2008	0.005	(0.009)	0.012	(0.020)
2009	-0.008	(0.006)	0.000	(0.017)
2010	-0.018***	(0.006)	-0.008	(0.014)
2011	-0.013**	(0.006)	-0.008	(0.014)
2012	-0.008*	(0.005)	-0.009	(0.013)
2013	-0.012*	(0.006)	-0.008	(0.013)
KBS (171,905.9 enrollees)				
2008	-0.003	(0.011)	-0.012	(0.031)
2009	-0.004	(0.007)	-0.000	(0.025)
2010	-0.017**	(0.008)	-0.031	(0.021)
2011	-0.005	(0.007)	-0.022	(0.021)
2012	-0.010	(0.007)	-0.021	(0.017)
2013	-0.018**	(0.008)	-0.044*	(0.025)
<i>N</i>	365757		41712	

*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$.

Notes: Coefficients and standard errors clustered at code and plan type in parentheses. Coefficients refer to interaction terms of respective year indicators with an indicator for being included in the RA. For types BKK-KBS the triple interaction terms of year x "in RA" and specific type are shown. Number of enrollees refers to average of 10% of officially reported enrollees for the different types across the 6 data years. All estimations include year and ICD code fixed effects.

Table 7 shows the results from the models investigating the performance of plan types in the period during which RA transitioned to diagnosis-based payments. The results in column 1 suggest that the AOK plans (the reference group) are relatively less likely to levy supplemental fees compared to the other plan types, taking into account each plan's contribution rate in 2008. The estimate for rebates are less clear and, in aggregate, regional plans do not seem to disproportionately gain enrollment by 2012 compared to 2008. These results suggest that the regional plans may have profited from the changes in coding by avoiding the introduction of supplemental fees, despite their relatively high contribution rates in 2008. However, this analysis is merely indicative, as the coding responses are not measured at the level of the individual plan and as the results presented in Table 7 do not control for other factors that may have affected whether plans introduce supplemental fees, issue rebates and loose or gain enrollment, such as changes in administrative costs or efforts to risk select enrollees.

Table 7: Differences in premiums and enrollment across plan types over time

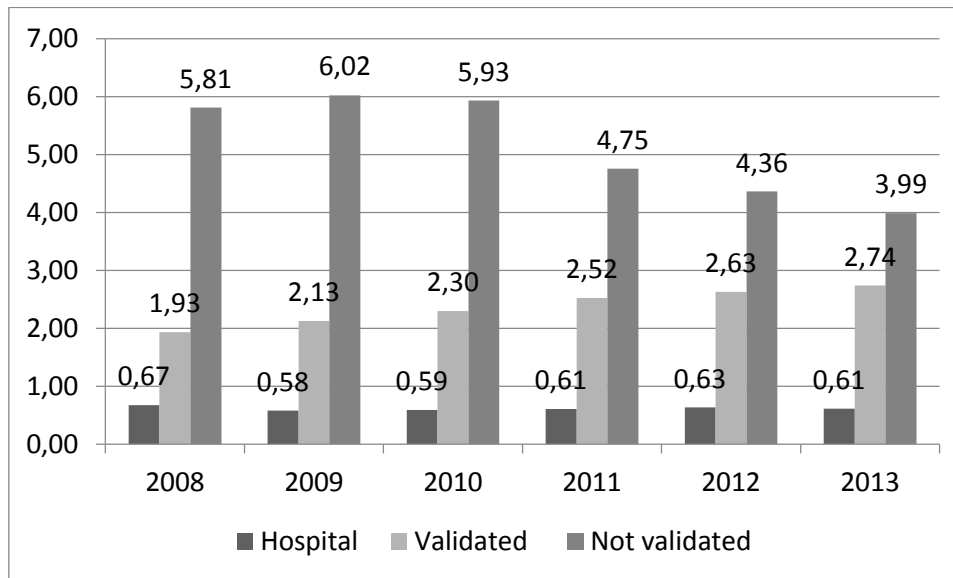
	(1) Pr Suppl. Fee	(2) Pr Rebate	(3) Plan size 2012
BKK	0.347*** (0.065)	0.011 (0.044)	-0.044 (0.804)
IKK	0.210*** (0.067)	0.191 (0.120)	0.518 (0.573)
EK	0.437*** (0.156)	0.263* (0.151)	0.081 (1.497)
KBS	0.313*** (0.072)	-0.213*** (0.069)	1.179*** (0.204)
2008 contr. rate	15.274*** (3.111)	-10.394*** (3.194)	
Plan size 2008			1.016*** (0.056)
Constant	-2.252*** (0.461)	1.533*** (0.472)	0.040 (0.845)
N plans	210	210	133
R2	0.150	0.097	0.985
Mean dep.	0.148	0.148	3.848

*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$.

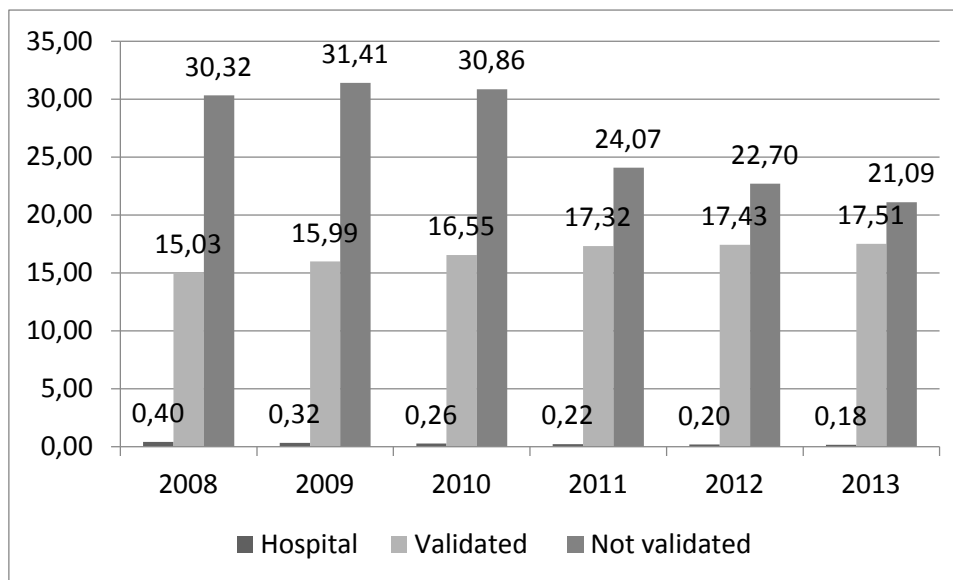
Notes: Coefficients and robust standard errors after OLS estimation in parentheses. AOK as reference plan type. Dependent variables (1): indicator for whether plan levied any supplemental fee between 2009 and 2013, (2): indicator for whether plan issued any rebate between 2009 and 2013, (3): plan size in 2012 measured as number of paying enrollees. (1) and (2) are based on all 2008 plans, (3) uses all plans that exist in 2012 and plan size in 2008 constructed using information on all mergers that happened over time. Data sources: DFG Atlas and Stiftung Warentest.

In order to further illustrate the patterns underlying the average increase in the share and number of validated diagnoses, we provide a case study of how the pattern of diagnoses evolves for two specific ICD categories, acute myocardial infarction (AMI, ICD I21) and stroke (ICD I64) in Figure 2. We selected these – merely illustrative – ICD categories as they are always included in the RA scheme and had stable definitions of subcodes (i.e., ICD codes that are associated with the ICD category) across all years. The top panel displays the prevalence of diagnoses of myocardial infarction per 1,000 insured based on hospital diagnoses only, and for two different types of physician diagnoses: claims with validated diagnoses and claims with diagnoses that are not designated as validated. The bottom panel of Figure 2 displays the same prevalence measures for stroke.

Figure 2: Case study comparing diagnoses across setting and designation
Acute Myocardial Infarction (ICD I21)



Stroke (ICD I64)



Notes: Prevalence calculated as number of diagnoses in the different settings or designation per 1,000 insured in the given year.

For both AMI and stroke, the prevalence of not-validated diagnoses is relatively large in 2008 and tends to decrease over time, while the prevalence of validated diagnoses increases. AMI and stroke thus seem to be medical conditions for which the increase in the share of validated diagnoses results from a shift from non-validated to validated ones. Although our data do not contain a further breakdown of the non-validated diagnoses that could facilitate the relevant analysis, this pattern may be explained by a shift of

diagnoses marked as symptomless condition post-diagnosis (*Zustand nach*, a non-validated code not considered by the RA scheme) toward diagnoses marked as validated.¹¹

Discussion

Our findings suggest that German health plans were successful in responding to the financial incentives embedded in the 2009 morbidity-based RA by inducing physicians to more often designate as “validated” (and hence payment-relevant) those diagnosis codes that are included in the payment formula. This increase in the share of payment-relevant validated diagnoses is relative to changes in other diagnoses that are not payment-relevant, and is more pronounced for certain plan types. It therefore most likely represents changes in nominal coding rather than real changes in morbidity. Because the set of codes included in the RA was only published in the Fall of 2008, it is unlikely that plans were able to target these codes already prior to the RA rollout in 2009. Changes in coding practices prior to 2009 would not affect our results if they affected all codes similarly, or would attenuate our estimates if they were targeted towards RA-specific codes

More specifically, our overall estimate suggests that the relative increase is 2.6 to 3.6 percentage points between 2008 and 2013. The effect takes hold two years after the specifics of the payment formula were finalized, and is present for most plan types, although regional plans (AOK) were among those that saw the largest changes in coding. One potential explanation for stronger coding responses of regional plans are the historical ties of these plans with the regional physician associations, and their relatively stronger local presence, e.g., through larger office networks.

Germany’s Statutory Health Insurance provides a useful setting to gather insights – for policy and research – about plan responses to payment incentives under managed competition. The SHI is tightly regulated, including with regards to vertical integration which has been proposed as mechanism for increases in coding intensity in Medicare Advantage (Geruso and Layton, 2015). Despite these restrictions and the

¹¹ We find similar patterns across all diagnoses that were always in the RA and were associated with stable set of ICD categories (results not shown). In aggregate for these diagnoses, the number of validated diagnoses increases at a faster rate than hospital diagnoses, while the number of non-validated diagnoses falls between 2008 and 2013.

arms-length interaction with physicians (via the physician associations), German plans have identified several ways to influence coding practices in their favor. From the research perspective, the operational features of the German plan and provider payment systems facilitate novel ways to identify coding responses, e.g., examining the subtle switch in designations of the same diagnostic codes (rather than reclassifications of codes) and using diagnoses outside the RA as control group. In this setting, there are also no concerns about selection between substantively different segments of the insurance market, as in Medicare Advantage and FFS. Finally, and unusually in the literature on coding responses, we are able to observe the full transition from the demographic-based to the morbidity-based RA. This allows us to examine a major change and minor revisions (within the morbidity based system) of the payment formula. Unlike related research (e.g., Geruso and Layton, 2015), our setup facilitates an identification strategy that leverages longitudinal data and allows for comparatively longer exposure periods.

More broadly, our results suggest that individual plans find it worthwhile to invest in changing coding practices. The instruments available to them are potentially costly, e.g., educating physicians or offering separate payments for certain coding practices. Encouraging changes to physician behaviors also raises the risk that new ailments are detected and paid for by the plans (Geruso and Layton, 2015). In addition, an individual plan's efforts may also accrue to competing plans, although our results also suggest that changes in coding behavior may be contained within certain plan types, at least initially.

Our findings have several general and specific implications. First, the alignment of coding practices with the financial incentives of the payment system (without real basis) could generate inefficiencies by inducing plans to divert resources from organizing the provision of health care services. Indeed, one plan manager estimated that, in 2014, German health plans collectively spent 1bn Euro between 2014 and mid-2016 to influence coding patterns (Schmergal, 2016). A resulting public debate has motivated legislative interest to examine and restrict plans' ability to influence coding behavior (Deutscher Bundestag, 2017, 2016; Mihm, 2016). In addition, inefficiencies could arise because plans may distort consumer choices if they spend the net gain resulting from coding changes on supplemental benefits or lowering premiums (Geruso and Layton, 2015). Second, there are several implications specific to the German approach of a fixed overall budget. In such context, an overall increase in nominal coding will distort the RA weights and may lead to inefficient allocation of resources across plans. In addition, if the marginal patient has lower severity, an increase in nominal coding will lower the average cost associated with the diagnosis, thus lowering payments from the RA for all patients and putting pressure on all plans in the market, particularly those plans that are less successful in influencing providers' coding behavior. There are also potential

implications outside the payment system, as physician and hospital diagnoses are used for other applications, e.g., to measure morbidity.

Many of these implications apply even if some or all of the change in coding is plan-specific “right-coding” rather than “upcoding,” i.e., if physicians are now merely more likely to correctly designate the diagnoses of RA conditions for a plans’ patients. That is because an individual health plan benefits from either change, as long as it occurs predominantly for its patients rather than for patients across all SHI plans. Thus, the incentive to strategically influence coding behavior and the associated distortions remain even if the induced changes leads to right-coding rather than upcoding.¹² However, right-coding could improve efficiency if “wrong-coding” initially differed across plans and subsequent right-coding mitigated this across-plan variation. Upcoding would not lead to such improvements. Definitively distinguishing between upcoding and right-coding would require an objective benchmark, e.g., independent chart reviews as in Carter et al. (1990). With the available data we cannot clearly distinguish these effects and hence cannot offer a firm conclusion on the welfare effects of the observed changes in coding.

Finally, there are several areas for future research. First, we only investigated one mechanism – the change of a diagnosis’ designation to “validated” – which does not preclude the existence of other means for plans to increase payouts from the RA system, e.g. through encouraging providers to shift from less to more lucrative codes or adding validated diagnoses for specific patients in order to fulfill the criterion of at least one diagnoses in two quarters of the year. German health plans continue to experience substantial consolidation and financial pressure; supervising and managing this competitive market will remain an important task. Second, our approach could be applied to data containing the plan names, not just their types. Knowing the plans’ identities would also allow a closer mapping of changes in coding and the economic performance of plans. Third, while we focus on diagnosis codes, future research could investigate potential effects on utilization patterns, i.e., visits and/or procedures. More broadly, our results indicate a need for more research on the strategic influences of health plans on coding behavior in European RA systems.

¹² However, as noted above, our data does not allow us to definitely distinguish whether the additional validated diagnoses translate into additional payments to plans. Nonetheless, additional validated diagnoses will never decrease payments that are already activated through existing records.

References

- Barros, P., Braun, G., 2016. Upcoding in a National Health Service: the evidence from Portugal: Upcoding in a National Health Service: the evidence from Portugal. *Health Econ.* 25, n/a-n/a. doi:10.1002/hec.3335
- Bauhoff, S., 2012. Do health plans risk-select? An audit study on Germany's Social Health Insurance. *J. Public Econ.* 96, 750–759. doi:10.1016/j.jpubeco.2012.05.011
- BMG, 2013. Gesetzliche Krankenversicherung Mitglieder, mitversicherte Angehörige und Krankenstand Monatswerte Januar 2013. Bundesministerium fuer Gesundheit, Berlin, Germany.
- BMG, 2009. Gesetzliche Krankenversicherung: Mitglieder, mitversicherte Angehörige, Beitragssätze 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 Econ. Dep. Work. Pap.* No 612.
- Brown, J., Duggan, M., Kuziemko, I., Woolston, W., 2014. How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program. *Am. Econ. Rev.* 104, 3335–3364. doi:10.1257/aer.104.10.3335
- Buchner, F., Goepffarth, D., Wasem, J., 2013. The new risk adjustment formula in Germany: Implementation and first experiences. *Health Policy* 109, 253–262. doi:10.1016/j.healthpol.2012.12.001
- Busse, R., Blümel, M., 2014. Germany health system review. *Health Syst. Transit.* 16.
- BVA, 2015. Tätigkeitsbericht 2015. Bundesversicherungsamt, Bonn, Germany.
- Carter, G.M., Newhouse, J.P., Relles, D.A., 1990. How much change in the Case Mix Index is DRG creep? *J. Health Econ.* 9, 411–428. doi:10.1016/0167-6296(90)90003-L
- Dafny, L.S., 2005. How Do Hospitals Respond to Price Changes? *Am. Econ. Rev.* 95, 1525–1547. doi:10.1257/000282805775014236
- Deutscher Bundestag, 2017. Änderungsantrag 1 der Fraktionen der CDU/CSU und SPD zum Entwurf eines Gesetzes zur Stärkung der Heil- und Hilfsmittelversorgung (Heil- und Hilfsmittelversorgungsgesetz – HHVG) (No. 18(14)0226.2). Deutscher Bundestag, Ausschuss fuer Gesundheit.
- Deutscher Bundestag, 2016. Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Harald Weinberg, Sabine Zimmermann (Zwickau), Katja Kipping, weiterer Abgeordneter und der Fraktion DIE LINKE. Mögliche „Krankfärberei der Versicherten“ durch gesetzliche Krankenkassen (November 11, 2016) (No. Drucksache 18/10153). Deutscher Bundestag.
- DIMDI, 2012. Welche Zusatzkennzeichen gibt es in ICD-10-GM und OPS und wie werden sie angewendet? (FAQ Nr. 1010) [WWW Document]. URL https://www.dimdi.de/static/de/klassi/faq/icd-10/icd-10-gm/faq_1010.htm_319159480.htm (accessed 1.10.17).
- Gaßner, M., Arndt, V., Fischer, L., Göppfarth, D., 2010. Sind die Diagnosezahlen nach Einführung des morbiditätsorientierten Risikostrukturausgleichs angestiegen? *Gesundh.- Sozialpolitik* 64, 11–19. doi:10.5771/1611-5821-2010-6-11
- Geruso, M., Layton, T., 2015. Upcoding: Evidence from Medicare on Squishy Risk Adjustment (Working Paper No. 21222). National Bureau of Economic Research.
- Göppfarth, D., Henke, K.-D., 2013. The German Central Health Fund-recent developments in health care financing in Germany. *Health Policy* 109, 246–252. doi:10.1016/j.healthpol.2012.11.001
- Januleviciute, J., Askildsen, J.E., Kaarboe, O., Siciliani, L., Sutton, M., 2016. How do Hospitals Respond to Price Changes? Evidence from Norway: How do Hospitals Respond to Price Changes? *Health Econ.* 25, 620–636. doi:10.1002/hec.3179
- Jürges, H., Köberlein, J., 2015. What explains DRG upcoding in neonatology? The roles of financial incentives and infant health. *J. Health Econ.* 43, 13–26. doi:10.1016/j.jhealeco.2015.06.001

- KBV, 2011. Ambulante Kodierrichtlinien: Diagnosesicherheit und Seitenlokalisation. Dtsch Arztebl Int. 108, A-271-A-273.
- Kronick, R., Welch, P., 2014. Measuring Coding Intensity in the Medicare Advantage Program. Medicare Medicaid Res. Rev. 4. doi:10.5600/mmrr.004.02.a06
- KV Berlin, 2013. Rundschreiben Betreuungsstrukturvertrag mit der AOK Nordost.
- Mihm, A., 2016. Betrug durch Krankenkassen: AOK wegen Manipulationen bestraft (Nov 11, 2016). Frankf. Allg. Ztg.
- Müller, A., Lange, W., 2012. Dfg-Krankenkassenatlas. MC.B, Verl. für Gesellschaftspolitik, Berlin.
- Newhouse, J.P., Price, M., Huang, J., McWilliams, J.M., Hsu, J., 2012. Steps to reduce favorable risk selection in medicare advantage largely succeeded, boding well for health insurance exchanges. Health Aff. Proj. Hope 31, 2618–2628. doi:10.1377/hlthaff.2012.0345
- Sacarny, A., 2016. Technological Diffusion Across Hospitals: The Case of a Revenue-Generating Practice.
- Schmergal, C., 2016. Diagnose auf Wunsch: Wie Krankenkassen sich beim Hausarzt einmischen (August 19, 2016). Spiegel.
- Schmitz, H., Ziebarth, N.R., 2016. Does price framing affect the consumer price sensitivity of health plan choice? J. Hum. Resour. doi:10.3368/jhr.52.1.0814-6540R1
- Silverman, E., Skinner, J., 2004. Medicare upcoding and hospital ownership. J. Health Econ. 23, 369–389. doi:10.1016/j.jhealeco.2003.09.007
- Van de Ven, W.P.M.M., Beck, K., Van de Voorde, C., Wasem, J., Zmora, I., 2007. Risk adjustment and risk selection in Europe: 6 years later. Health Policy, Health Policy 83, 162–179.
- van Kleef, R.C., van Vliet, R.C.J.A., van Rooijen, E.M., 2014. Diagnoses-based cost groups in the Dutch risk-equalization model: The effects of including outpatient diagnoses. Health Policy 115, 52–59. doi:10.1016/j.healthpol.2013.07.005
- Wuppermann, A.C., Bauhoff, S., Grabka, M., 2014. The Price Sensitivity of Health Plan Choice: Evidence from Retirees in the German Social Health Insurance (SSRN Scholarly Paper No. ID 2433588). Social Science Research Network, Rochester, NY.

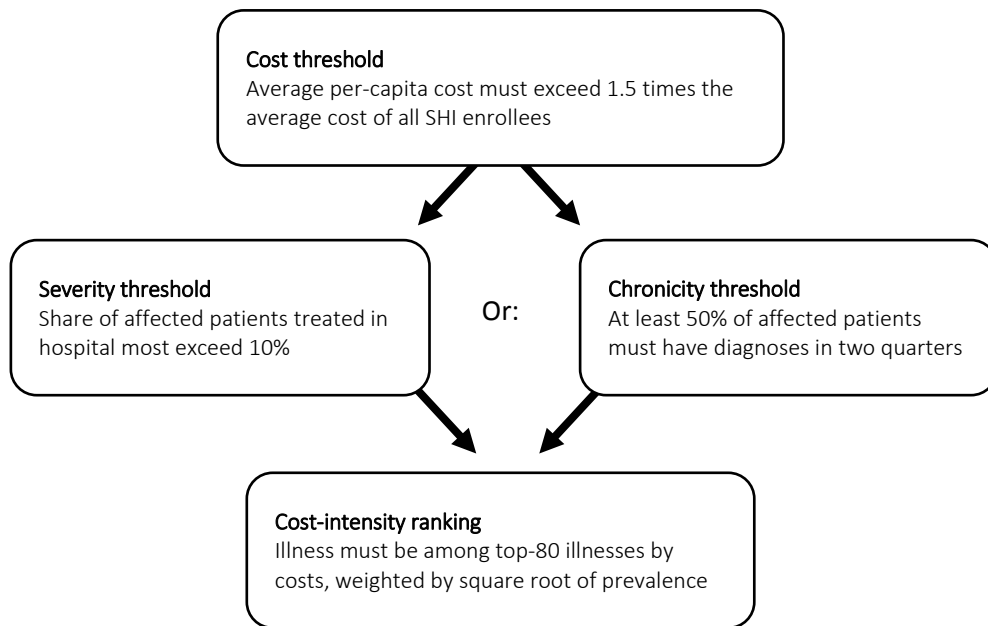
Appendix A

In this Appendix we describe the selection of the 80 diseases covered in the risk adjustment.

By law, the risk adjustment model is restricted to 80 chronic and severe diseases. The diseases are selected annually by the following process: First, the approx. 15,000 ICD-codes are merged to about 360 diseases. The next step verifies if the average cost of people diagnosed with this disease exceed 1.5 times the average costs of all SHI members (a condition set by law). Furthermore, the disease must be severe or chronic. Diseases are defined as severe if more than 10% of affected patients undergo hospital treatment. A disease is defined as chronic if the condition persists for at least two quarters for more than 50% of affected patients. After these steps, some 250 diseases remain. These are sorted by cost intensity, defined as costs weighted by the square root of prevalence. From this list, the top 80 diseases are chosen (see Appendix Figure A.1).

Payments for the HMGs are calculated annually by a WLS-regression. Individual annual health care expenditure is the dependent variable, whereas age, sex, disability-to-work-status and allocation to HMGs are the independent variable. The reciprocal of enrolment days serves as weight. Cost data is taken from the current year, while allocation to the HMGs is based on prior year diagnoses and prescriptions (prospective model). Actual payments to sickness funds are made monthly on provisional data and reconciled after the end of the year in a final settlement when all necessary data has been collected.

Appendix Figure A.1: Selection of illnesses included in the morbidity-based risk adjustment



Appendix B

Appendix Table B.1: Validated diagnoses and deciles of RA baseline payment

	<u>Share validated</u>		<u>Ln number validated</u>	
	(1) Full Sample	(2) Similar chronicity	(3) Full Sample	(4) Similar chronicity
in RA x 2009	-0.008 (0.007)	-0.007 (0.022)	0.002 (0.019)	0.004 (0.051)
in RA x 2010	-0.000 (0.008)	0.011 (0.017)	0.061** (0.026)	-0.032 (0.058)
in RA x 2011	0.005 (0.008)	0.027 (0.019)	0.073*** (0.025)	-0.037 (0.066)
in RA x 2012	0.017* (0.009)	0.041*** (0.014)	0.106*** (0.032)	0.067 (0.098)
in RA x 2013	0.024*** (0.009)	0.050*** (0.019)	0.133*** (0.038)	0.050 (0.125)
Decile 2				
x in RA x 2009	0.002 (0.008)		0.001 (0.025)	
x in RA x 2010	0.009 (0.011)		0.001 (0.034)	
x in RA x 2011	0.008 (0.010)		0.002 (0.041)	
x in RA x 2012	0.003 (0.011)		-0.011 (0.052)	
x in RA x 2013	0.005 (0.012)		-0.013 (0.065)	
Decile 3				
x in RA x 2009	0.007 (0.009)	0.021 (0.021)	0.017 (0.052)	0.205*** (0.049)
x in RA x 2010	-0.006 (0.013)	-0.004 (0.016)	-0.047 (0.042)	0.221*** (0.057)
x in RA x 2011	-0.008 (0.012)	-0.009 (0.019)	0.008 (0.058)	0.298*** (0.065)
x in RA x 2012	-0.012 (0.011)	-0.030** (0.013)	0.027 (0.068)	0.188* (0.096)
x in RA x 2013	-0.015 (0.011)	-0.034* (0.018)	-0.003 (0.065)	0.196 (0.123)
Decile 4				
x in RA x 2009	-0.013 (0.012)	0.034 (0.029)	0.022 (0.027)	0.075 (0.068)
x in RA x 2010	0.012 (0.012)	0.027 (0.029)	0.000 (0.040)	0.122 (0.082)
x in RA x 2011	0.003 (0.012)	0.022 (0.031)	0.053 (0.045)	0.290*** (0.098)
x in RA x 2012	-0.007 (0.012)	0.006 (0.022)	0.013 (0.061)	0.226** (0.108)

Table continued on next page

Appendix Table B.1 continued

x in RA x 2013	-0.004 (0.011)	0.007 (0.023)	0.043 (0.070)	0.314** (0.133)
Decile 5				
x in RA x 2009	0.004 (0.008)		0.011 (0.026)	
x in RA x 2010	0.011 (0.009)		0.008 (0.036)	
x in RA x 2011	0.009 (0.010)		0.027 (0.049)	
x in RA x 2012	-0.011 (0.011)		-0.019 (0.055)	
x in RA x 2013	-0.014 (0.012)		-0.035 (0.064)	
Decile 6				
x in RA x 2009	0.011 (0.012)		0.086* (0.048)	
x in RA x 2010	0.033 (0.022)		0.194*** (0.072)	
x in RA x 2011	0.045** (0.020)		0.333*** (0.093)	
x in RA x 2012	0.039* (0.023)		0.367*** (0.107)	
x in RA x 2013	0.035** (0.017)		0.416*** (0.140)	
Decile 7				
x in RA x 2009	0.014* (0.008)	0.001 (0.025)	0.032 (0.024)	-0.157 (0.109)
x in RA x 2010	0.015 (0.010)	-0.008 (0.022)	-0.020 (0.033)	-0.051 (0.073)
x in RA x 2011	0.019* (0.011)	-0.021 (0.019)	-0.008 (0.035)	-0.058 (0.086)
x in RA x 2012	0.006 (0.010)	-0.034* (0.019)	-0.038 (0.044)	-0.073 (0.108)
x in RA x 2013	0.010 (0.011)	-0.001 (0.018)	-0.053 (0.051)	-0.154 (0.137)
Decile 8				
x in RA x 2009	0.008 (0.009)	0.009 (0.021)	-0.004 (0.026)	-0.009 (0.052)
x in RA x 2010	-0.005 (0.014)	-0.015 (0.018)	-0.038 (0.044)	-0.016 (0.069)
x in RA x 2011	0.011 (0.013)	-0.039** (0.020)	0.004 (0.050)	-0.054 (0.077)
x in RA x 2012	0.001 (0.012)	-0.057*** (0.014)	0.028 (0.071)	-0.165 (0.108)
x in RA x 2013	0.003	-0.062***	0.020	-0.188

	(0.013)	(0.019)	(0.075)	(0.136)
Table continued on next page Appendix Table B.1 continued				
Decile 9				
x in RA x 2009	0.002 (0.014)	0.002 (0.023)	0.012 (0.029)	-0.010 (0.057)
x in RA x 2010	0.007 (0.012)	-0.007 (0.017)	-0.004 (0.043)	0.024 (0.061)
x in RA x 2011	-0.004 (0.013)	-0.021 (0.019)	-0.009 (0.043)	-0.022 (0.092)
x in RA x 2012	-0.002 (0.014)	-0.037** (0.017)	0.004 (0.046)	-0.144 (0.121)
x in RA x 2013	-0.000 (0.014)	-0.035* (0.019)	0.007 (0.050)	-0.111 (0.145)
Decile 10				
x in RA x 2009	0.013 (0.009)	0.087*** (0.021)	0.077** (0.034)	0.179*** (0.049)
x in RA x 2010	0.006 (0.013)	0.010 (0.016)	0.043 (0.041)	0.148*** (0.057)
x in RA x 2011	0.009 (0.012)	0.020 (0.019)	0.200*** (0.059)	-0.009 (0.065)
x in RA x 2012	-0.003 (0.013)	-0.119*** (0.013)	0.180*** (0.063)	-0.046 (0.096)
x in RA x 2013	0.004 (0.014)	-0.063*** (0.018)	0.165** (0.073)	-0.015 (0.123)
Constant	0.780*** (0.001)	0.806*** (0.003)	5.658*** (0.005)	6.100*** (0.010)
N codes x year	77,449	8,561	76,514	8,498
N cluster	1,620	161	1,620	161
R2	0.770	0.815	0.988	0.990

*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$.

Notes: Coefficients and standard errors clustered at ICD category in parentheses. All estimations include ICD and year fixed effects. Number of observations lower than in Table 2 because codes that switch in and out of RA excluded. Coefficients for payment deciles 2-10 measure difference to the effect in decile 1 (reference group).

Appendix Table B.2: Heterogeneity of RA effect on share validated with respect to features of diseases in 2006

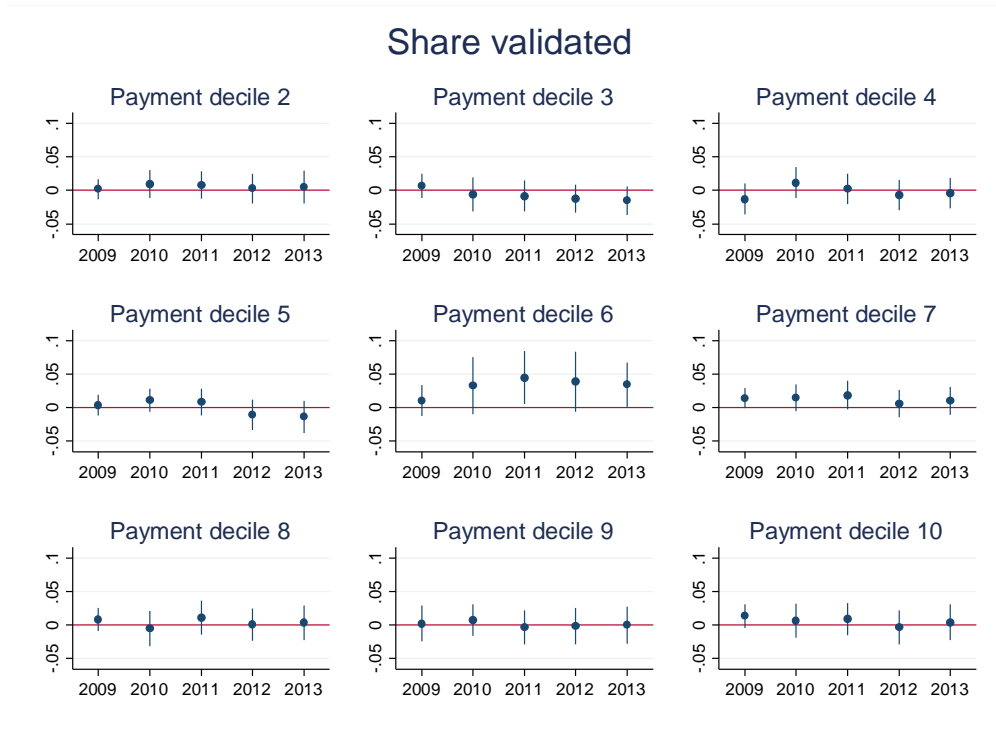
Heterogeneity w.r.t	Chronicity in 2006		Hospitalization in 2006		Costs in 2006	
	(1) Full Sample	(2) Similar chronicity	(3) Full Sample	(4) Similar chronicity	(5) Full Sample	(6) Similar chronicity
In RA	-0.003 (0.009)	0.073 (0.056)	-0.005 (0.006)	0.021 (0.021)	-0.004 (0.007)	-0.011 (0.018)
2009 x in RA	-0.010** (0.005)	0.018 (0.031)	-0.005 (0.003)	0.013 (0.014)	-0.006 (0.004)	0.022 (0.017)
2010 x in RA	0.003 (0.006)	0.013 (0.026)	0.007 (0.004)	0.013 (0.012)	0.005 (0.006)	0.024 (0.019)
2011 x in RA	0.005 (0.006)	0.022 (0.028)	0.012** (0.005)	0.021 (0.014)	0.006 (0.006)	0.033 (0.021)
2012 x in RA	0.004 (0.006)	0.045 (0.032)	0.016*** (0.005)	0.024** (0.011)	0.010 (0.006)	0.039** (0.016)
2013 x in RA	0.015** (0.006)	0.049 (0.034)	0.025*** (0.004)	0.035*** (0.012)	0.019*** (0.006)	0.047*** (0.018)
Heterogeneity x in RA	0.030 (0.043)	-1.632 (1.031)	-0.018 (0.028)	0.157 (0.187)	0.000 (0.000)	0.000 (0.000)
x 2009 x in RA	0.041* (0.023)	-0.133 (0.392)	0.005 (0.018)	0.186 (0.133)	-0.000 (0.000)	-0.000 (0.000)
x 2010 x in RA	-0.010 (0.026)	0.029 (0.346)	0.002 (0.029)	0.065 (0.121)	-0.000 (0.000)	-0.000 (0.000)
x 2011 x in RA	0.011 (0.027)	0.026 (0.360)	-0.017 (0.025)	0.027 (0.130)	-0.000 (0.000)	-0.000 (0.000)
x 2012 x in RA	0.026 (0.030)	-0.309 (0.421)	-0.019 (0.028)	-0.010 (0.109)	-0.000 (0.000)	-0.000 (0.000)
x 2013 x in RA	0.003 (0.028)	-0.200 (0.447)	-0.014 (0.026)	0.041 (0.115)	-0.000 (0.000)	0.000 (0.000)
N	73,559	9,143	73,559	9,143	73,559	9,143

*** $p < 0.01$, ** $p < 0.05$ * $p < 0.1$.

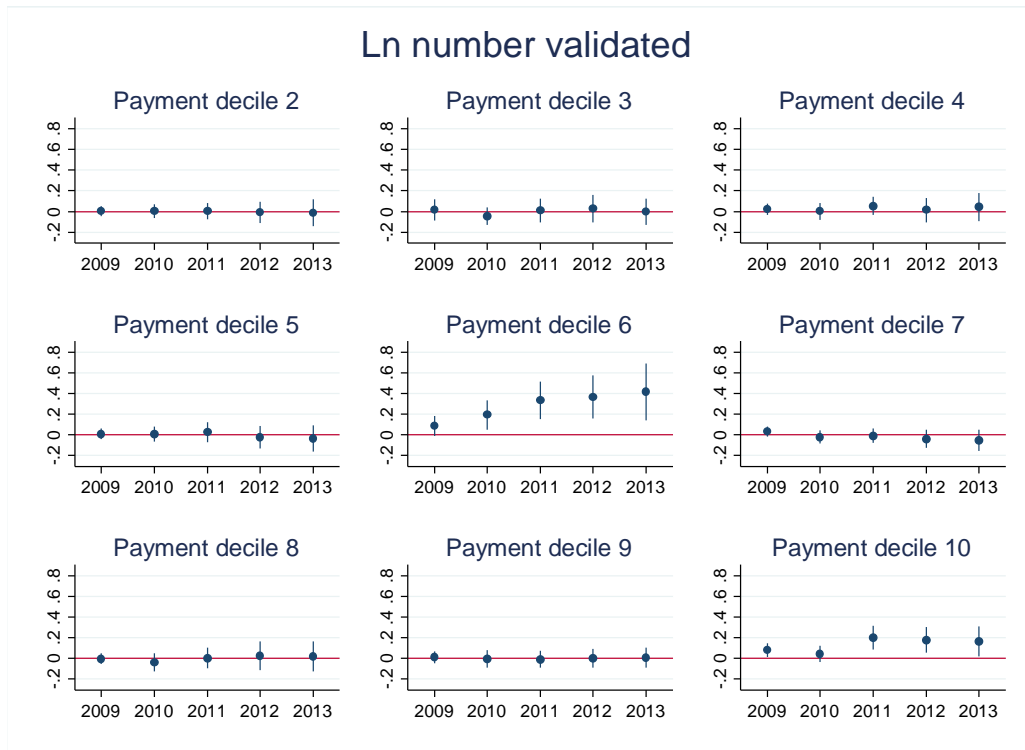
Notes: Coefficients and standard errors clustered at ICD category in parentheses. All estimations include ICD and year fixed effects. Number of observations lower than in Table 2 because of missing values of chronicity, hospitalization or costs. The main results are virtually unchanged when restricted to these codes and available upon request. Chronicity measures share of patients with at least one diagnosis in at least two quarters of the year, hospitalization measures share of patients that are hospitalized with a diagnosis of the associated condition, and costs are the following year's costs associated with the condition.

Figure B.1: Nonlinearities in the effect of payment on coding responses

Panel A:



Panel B:



Note: Point estimates and 95% confidence intervals of year x in RA x decile of pay with first decile as reference group.