–

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

hche | Hamburg Center for Health Economics

System-wide Effects of Hospital Payment Scheme Reforms:

The German Introduction of Diagnosis-Related Groups

Robert Messerle, Jonas Schreyögg

Research Paper
Year: 2022
No: 26

System-wide Effects of Hospital Payment Scheme Reforms:
The German Introduction of Diagnosis-Related Groups

Effects of an especially comprehensive hospital remuneration reform

Robert Messerle* and Jonas Schreyögg

Hamburg Center for Health Economics, University of Hamburg, Esplanade 36, 20354, Hamburg, Germany

Acknowledgements

We thank Jan Marcus and the participants of the HCHE Methodological Workshop Concerning Working with Administrative Data for their valuable comments and suggestions.

ORCID

Robert Messerle https://orcid.org/0000-0002-0354-8675

Jonas Schreyögg https://orcid.org/0000-0001-8030-2161

*Correspondence to Robert Messerle, University of Hamburg, Esplanade 36, 20354, Hamburg, Germany.
Email: robert.messerle@uni-hamburg.de, Phone: +49 42838-9524

Abstract

Hospitals account for 40 % of all healthcare expenditures and play a central role in healthcare provision. Therefore, the way hospitals are payed has major implications for the care they provide. Yet the knowledge about system wide effects of payment reforms is surprisingly thin. This study analyzes the especially comprehensive German introduction of diagnosis-related groups. In Germany, diagnosis-related groups function as sole pricing, billing and budgeting system and almost exclusively determine the turnover of hospitals. The introduction of diagnosis-related groups thus completely overhauled existing payment structures. It thereby offers a unique possibility for the analysis of payment reforms. Using aggregate OECD data and recent econometric advances, we analyze hospital activity and efficiency. We find that the reform increased hospital activity – measured as the number of discharges – significantly by around 2 percent per year. In contrast to many earlier studies, we find that diagnosis-related groups do not necessarily lead to a lowering of the average length of stay.

1. Introduction

Hospitals play a central role in healthcare provision. With on average 40 % of all healthcare expenditures they are the main recipient in almost all OECD countries (OECD, 2021a). It is therefore not surprising that hospitals are a prominent target for policy reforms. Besides restructuring of hospitals and hospital care itself, hospital financing is a recurring point of interest among policy makers. By changing the way hospitals are payed, the care they provide can be influenced. However, robust empirical evidence about effects of different funding systems is scarce. Despite insufficient knowledge about consequences, the overwhelming majority of countries arrived at Activity Based Funding (ABF) as a major source of hospital financing, mainly in the form of diagnosis-related groups (DRGs). DRGs group hospital cases based on their diagnoses into economically homogeneous groups. Thereby, DRG link remuneration to the number of cases. A hospital receives more revenue if it admits and treats more patients. Moreover, DRGs used as pricing system encourage hospitals to keep costs below the respective flat rate per case (Yardstick-competition - Shleifer, 1985).

DRGs were first introduced in the US and gradually became the basis for hospital payment schemes around the world, albeit with country-specific adaptions. Depending on the previous mode of payment, reforms' effects seem to point in opposite directions. Moving from global budgets to DRGs, as it was the case for most European countries, seemed to increase hospital activity and expenditures. In contrast in the US, where DRGs followed fee-for-service payment schemes, hospital activity initially declined (Lave, 1989). Despite it being probably one of the most important hospital healthcare interventions in recent decades, still little is known about its system wide effects.

In this paper, we use an especially extensive reform of hospital payments as a natural experiment. Germany introduced DRGs as the predominant source of financing for all acute care hospitals. It thereby differs from partial introductions in other countries. Moreover, to our knowledge Germany encompasses the only hospital system worldwide where DRGs function as sole pricing, billing and budgeting system and almost exclusively determine the turnover of hospitals. To achieve a robust estimation of the reform's effects we use three complementary quasi-experimental methods: Difference-in-Differences, Synthetic Control and Synthetic Difference-in-Differences. Aggregate country-level panel data and a comprehensive scheme of pre-dominant hospital payment systems allow us to construct a suitable control group. Thus, we do not have to resort to a pre-post-analysis but can derive causal inference.

A major motive for introducing DRG systems was to increase hospital throughput by improving efficiency. Our main outcomes of interest therefore concern the hospital activity and efficiency, represented by hospital discharges and length of stay. Following previous research, we also include secondary outcomes concerning population health status and hospital resources. Our approach allows us to complement the DRG literature with effect estimates for a uniquely comprehensive reform. It furthermore underpins the ongoing policy discussion in Germany with robust evidence. We find that the DRG reform steeply increased hospital discharges at a rate of approximately 2% annually. In contrast to previous studies, we were not able to identify any empirical impact on the length of stay. Extensive robustness tests confirm the validity of our results.

Our results add to the general research regarding DRG systems, which separates into three basic streams (see Table 1). The first stream evolves around the effects *within* DRG systems. Studies in this stream investigate hospitals' responses to changes of prices or price structures within an existing DRG system (Dafny, 2005). One of the main challenges here is to distinguish between effects at the intensive and extensive margin, e.g. between upcoding and genuine increases in discharges. The majority of studies finds that hospitals react mainly by an altered coding practice, i.e. upcoding of patients into higher priced and therefore more profitable diagnoses (Cook & Averett, 2020; Di Giacomo et al., 2017; Jürges & Köberlein, 2015) Whether hospitals also alter treatment decisions in this setting remains unclear (Bäuml, 2020). Some studies find increases in the number of discharges for surgical but not medical DRGs (Verzulli et al., 2017), which is in line with some theoretical considerations (Hafsteinsdottir & Siciliani, 2010). Changes in the quality of care, e.g. regarding in-hospital mortality were not found (Bäuml & Kümpel, 2021).

The second research stream investigates the effects of the introduction of a DRG system *itself*, primarily with a focus on the level of *individual* hospitals, diseases or population subgroups. Overall, evidence in this stream suggests that a DRG introduction causes substantial shifts to post-acute care and increases readmission rates. Studies also indicate that the transition to DRGs initially decreases the length of hospital stay (Meng et al., 2020). However, a high heterogeneity of results prohibits an unequivocal verdict. General conclusions are hampered by econometric challenges as well as inevitable design constraints (Palmer et al., 2014). Non-experimental, descriptive studies - often covering only short periods - continue to dominate the empirical literature (Valentelyte et al., 2021). Even econometrically elaborated studies exhibit a high risk of bias (Meng et al., 2020). A lack of unaffected controls, since implementations are usually nationwide (Schreyögg, 2019), also impedes conclusions by design. Furthermore, the availability of administrative data for pre-intervention periods is often insufficient.

The third research stream, to which we attribute our study, also analyzes the effects of the introduction of the DRG system *itself*. However, unlike studies in the preceding streams, this stream focuses on effects on the *system level*. This is necessary because even comprehensive studies from the previous research streams are generally limited to subgroups of the population. Kutz et al. (2019) e.g. analyze 2.5 million Swiss hospitalizations over a period of six years, but exclude i.a. all surgical patients. These varying research scopes might relate to the diverging results in research stream 2. Feess et al. (2019) find highly heterogeneous effects on the average length of stay following the German DRG reform, depending on patient and hospital characteristics. On aggregate level however, they do not find any systematic change.

System wide conclusions based on subgroup analysis are thus difficult to draw. To overcome this issue, studies in the third research stream try to establish a causal link between reforms of hospital payment schemes and subsequent developments using aggregate country-level panel data. To the best of our knowledge, only Moreno-Serra and

Wagstaff (2010) and Wubulihasimu et al. (2016) fall into this category. Both estimate the effects of changes in hospital payment systems, with DRG systems as one example. Moreno-Serra and Wagstaff find increased healthcare expenditures as well as decreased lengths of stay for Eastern Europe and Central Asia. Wubulihasimu, Brouwer, and van Baal concentrate on OECD-countries and derive first evidence for increased health expenditures and a lower mortality. The results should be regarded with caution, as they are sensitive to model specifications. They highlight the heterogeneity of reforms and their only gradual or partial introduction as major limitations for estimation. From a methodological point of view, both studies (partly) use staggered Difference-in-Differences. Recent advances have indicated that this widely used approach can be biased when treatment effect heterogeneity is present (see e.g. Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021), opening room for further research.

This paper is structured as follows. Section 2 provides an overview over the German DRG system. Sections 3 follows with a description of the data and explains our approach of constructing a data set covering the predominant hospital funding schemes for the control countries. Section 4 provides information about the methodological setup and the estimation procedure. Section 5 presents empirical results and sections 6 and 7 conclude.

Table 1: Research streams regarding DRG systems

Research Stream	**effects within DRG system**		**effects of DRG system itself**			
	price changes		**hospital/specialty/ patient level**		**system level**	
Effect	volume of care	+	volume of care	0/+	volume of care	0
	quality of care	0	length of stay	0/-	length of stay	-
	upcoding	++	mortality	0	mortality	0
			quality of care	+/-		
			readmissions or shift to post-acute care	++		
			hospital efficiency	+/-		
Studies or Reviews	Bäuml, 2020; Dafny, 2005; Jürges & Köberlein, 2015; Salm & Wübker, 2020 …		Meng et al., 2020; Palmer et al., 2014…		Moreno-Serra & Wagstaff, 2010; Wubulihasimu et al., 2016	
Caveat	methodological challenges to differentiate between effects at the in- and extensive margin		short study periods, design constraints, e.g. lack of unaffected controls, availability of administrative data for pre-intervention period, econometric challenges		payment scheme definition/ specification of control group, sample size, econometric challenges	

2. DRGs in Germany

For decades, hospital care in Germany was reimbursed mainly via per diem rates. Before 1993, rates were calculated based on full cost compensation. After 1993, a fixed global budget had to be considered in rate negotiations, but several exceptions allowed for high and steady expenditure growth rates. When Germany opted in 2000 to introduce a German DRG system (gDRG) based on the Australian Refined Diagnosis-Related Groups, it targeted several aims i.e.: Shortening the hospital length of stay, stabilizing healthcare expenditures, introducing a performance-related payment of hospitals and increasing the transparency about services and costs of hospitals. Despite experiences of other countries, the possibility of increased hospital activity was not overly focused in the legislative process.

The scope of the gDRG system is quite comprehensive: It is – with the exemption of some additional payments for e.g. especially expensive medicines – the only pricing system for hospitals. It is furthermore used for determining a hospital's budget and for direct billing purposes. With around 1.200 different DRGs, it offers a high complexity.

The gDRG system covers all operating costs for all acute inpatient stays. Hospitals receive a lump sum depending on the main diagnosis and the procedures performed. Secondary diagnoses determine the severity of the base DRG. Additions and deductions are possible if the length of stay falls above or below a DRG specific threshold. Contingency costs, e.g. to ensure the provision of emergency care, are also included in DRG case payments.[1] Long-term infrastructure investment costs are in principle financed separately by federal states through taxes. They are provided as lump sums or can be applied as case surcharges. However, real public investment in hospitals has been decreasing continuously and is less than 5% of overall hospital funding today. Federal states would need to almost triple their allocations to reach relative funding levels of 1991. In contrast, DRGs allocate about 80 % of all financial resources to hospitals. This is one of the highest share of DRG-based hospital payment internationally (Cots et al., 2011) and the reason why DRGs are the main financial parameter for German hospitals. Besides, German hospitals are tightly focused on providing inpatient care. Only few legal exceptions allow outpatient service provision, which is furthermore financially unattractive. Thus, the importance of inpatient payment schemes for hospitals is among the highest across OECD countries (see Figure 1).

The gDRG system came into force in 2003 and became the obligatory remuneration system for all acute care hospitals in 2004. Hospitals could initially apply for an exemption from the DRG system but the requirements got increasingly demanding. In 2004 reimbursement was still budget neutral and entirely based on upfront-negotiated budgets. A hospital-specific base rate meant that in the first year the payment received was the same as under the previous system. A convergence phase from 2005 to 2009 gave hospitals the opportunity to adjust step wise to a state wide uniform price system (Salm & Wübker, 2020). While losses to hospitals originating from the reform were limited, gains, which arose from the new

[1] This partly changed only in 2018.

payment system, were uncapped. Thus, strong incentives for extended activity by profiting hospitals were in place early after the introduction.[2]

Although research accompanied the gDRG reform throughout, clear causal evidence is missing. Due to the obligatory and nationwide introduction, the reform lacks a suitable control group within the German healthcare system. Therefore studies often can only describe trends without the possibility of making statements about causal directions (Schreyögg et al., 2014). External time-varying factors, regression to the mean, false assumptions regarding the functional form of underlying time trends and other threats to the internal validity render these single case time-series analysis problematic.

Koné et al. conclude in a review that there is no robust empirical evidence to indicate positive or negative effects of the DRG introduction in Germany (Koné et al., 2019). Aggregate trends indicate that the average length of stay decreased, but less sharp then before gDRG introduction. The number of cases increased as well as hospital expenditures. Inpatient discharges rose from 16.6 million in 2004 to 19.4 million in 2017. This accounted for one of the largest increases in hospital discharges in Europe – in-spite of a stable population (see Appendix Figure A. 1). A steady enlargement of the physician workforce accompanied climbing case numbers. From 2000 to 2018 the number of physicians employed in hospitals increased by almost 50%. Nursing staff – which, with respect to revenue generation, is often attributed a lower importance – showed an unsteady development and increased by only 10% overall (in FTE, OECD, 2021a).

Following the lack of conclusive studies, public discussions largely take place without evidence. Especially negative consequences on hospital care and staffing are in public focus. Eventually in 2020, the presumption that financial incentives caused by the gDRG system resulted in under-staffing of nursing personnel, led to fundamental changes. With the PpSG (Pflegepersonal-Stärkungsgesetz) the German legislative decided to remunerate nursing staff for direct patient care independently of case payments. Nursing costs, which account for around 20% of total costs, are thus excluded from DRG calculations. Hospital remuneration now consists of a combination of per case reimbursement via DRGs and a nursing staff allowance based on full cost compensation.

[2] Salm and Wübker (2020) argue that hospitals which confronted price decreases also reacted by increasing service supply, opening another channel for increased hospital activity.

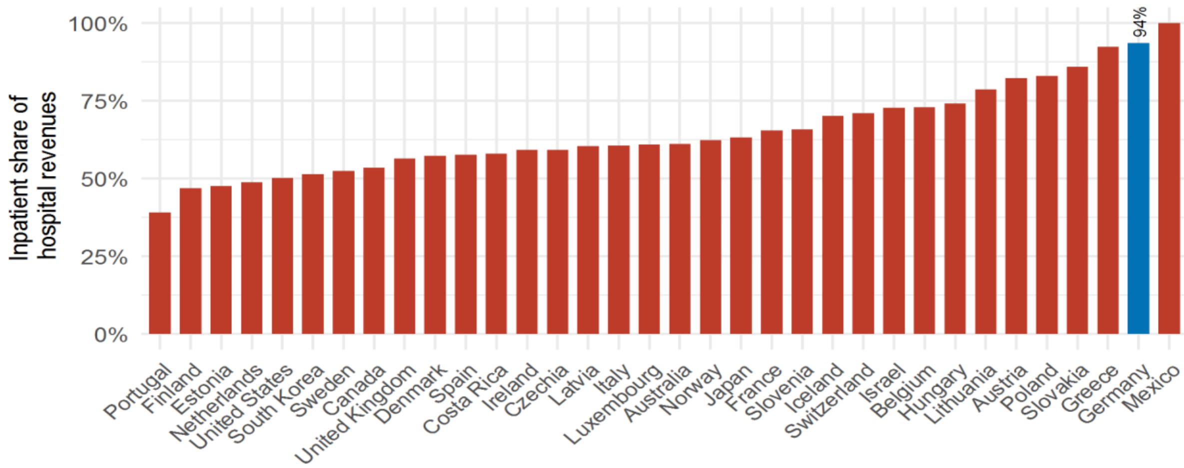


Figure 1: Importance of inpatient expenditures for hospitals in OECD countries 2017

Source: OECD (2021b)

3. Data

3.1. Classification of hospital payment systems

Table 2 summarizes our data collection process and estimation strategy. In order to investigate the effects of the gDRG reform, we constructed - in analogy to Moreno-Serra and Wagstaff (2010) and Wubulihasimu et al. (2016) - a data set describing the predominant hospital payment schemes for selected OECD and EU member states from 1994 to 2015 (see Figure 2). All European OECD countries, Australia, Canada and New Zealand as well as the EU Member States at the start of the reform (2004/2005) are included. The selection covers developed countries with in general comparable levels of healthcare and acts as starting point for the compilation of suitable controls for Germany.

In a first step, we created two basic categories with opposite incentive structures to classify hospital payment schemes: Fixed budgets (FB) or activity based funding (ABF). We classify a country as FB in a given year if global budgets or block grants are the predominant form of hospital funding. In this case, hospital revenues are mainly pre-determined depending on provider characteristics like hospital size or range of care. We chose ABF if hospitals are paid according to the characteristics of admitted patients or activities. This includes per case and DRG-based methods as well as fee-for-service. Basis for the classification is the World Health Organization’s Health System in Transition series (Busse et al., 2011; WHO, 2021) and additional literature. More details can be found in the Supporting Information.

Table 2: Outline of statistical analysis

Step	Description
1. Collecting information regarding predominant hospital payment schemes	Classify predominant hospital payment schemes for selected countries from 1994 to 2015. First step: assign Fixed Budget (FB) or Activity Based Funding (ABF) classification (a) main source is the Health Systems in Transition series (b) additional literature supplements classification Second step: decide on scope of payment scheme to differentiate between extensive and only partial ABF reforms - ABF often only affects a fraction of hospital budgets (e.g. Denmark, Italy), is limited to certain hospitals and regions (e.g. Finland, Sweden) or is used for budgeting but not for actual billing processes (e.g. Ireland) We exclude all countries that introduced an extensive form of activity based funding between 1999 and 2011 from the control group. Outside of this period, payment scheme reforms should not bias estimation.
2. Collecting and combining country level data from several sources	Aggregate (unbalanced panel) data from OECD and Eurostat. (a) main source for variables is OECD (b) additional countries and data comes from Eurostat Main outcome variables are - number of discharges per 100 000 inhabitants and - length of stay Secondary outcome variables concern - *hospital resources and expenditures as indicator of efficiency* - *population health status* Additional control variables, e.g. GDP per capita and share of population 65+, are used. Several variables exist in various definitions. See Supporting Information.
3. Applying three complementary estimation methods	Use of different estimation methods for a robust estimation. (a) A slightly extended Difference-in-Differences (DiD) model is our baseline approach. It is used for all outcomes with a credible parallel trend assumption (b) Synthetic Control Method (SCM) is used for all outcome variables; DRG reform works as prime example for “classic” SCM (c) Synthetic Difference-in-Differences (SDiD) is used for all outcome variables
4. Robustness checks	Application of several robustness checks for the different methodological approaches. (a) Different control variables, control countries, parallel trend sensitivity analysis (b,c) Placebo-in-space and placebo-in-time analysis, different control countries To validate the DRG reform in Germany as driving force for our effect estimates we further check for simultaneous healthcare reforms. Since the classification of hospital payment schemes is not straightforward for some countries, we also construct an alternative classification for the control countries.

To facilitate the identification of suited control units we add a mixed funding category to differentiate between extensive ABF reforms and the co-existence of multiple models. Many countries implemented hospital payment schemes gradually or only partially. Thereby ABF often only affects a fraction of hospital budgets (e.g. Denmark), is limited to certain hospitals and regions (e.g. Finland and Sweden) or is used for budgeting but not for actual billing processes (e.g. Ireland). We argue that in this case the change in hospital incentive structures is considerably weaker – at least at the aggregate level used in this analysis. This differentiation allows us to exclude countries with similar reforms as in Germany from the control group while retaining a reasonable size of the donor pool. For further analysis, we consider all countries that did not introduce any extensive reform of activity based funding between 1999 and 2011 as control units. Reforms years apart from the German DRG introduction should not impair our estimates. This approach excludes Estonia, France, the United Kingdom and Poland from the control group of our main analysis. We do, however, several robustness checks with different control group compositions.

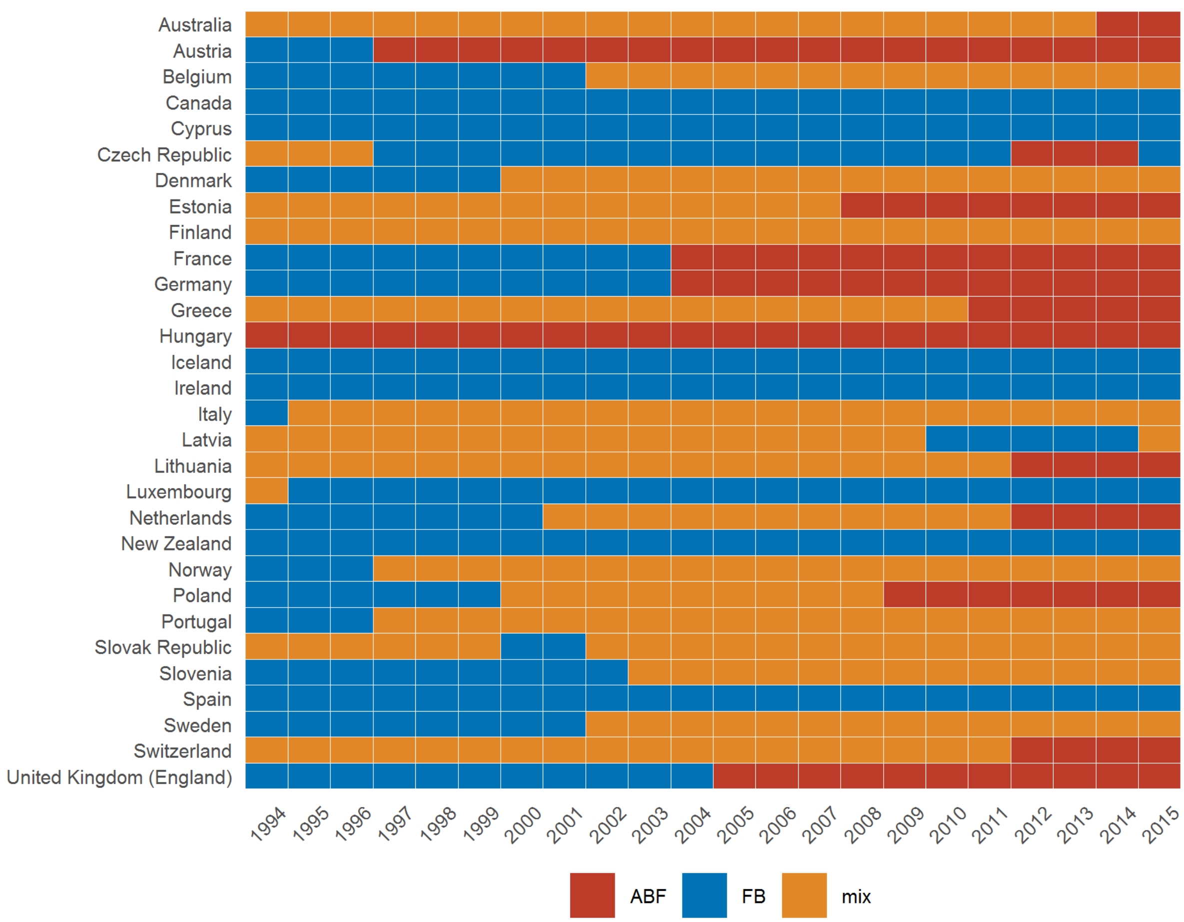


Figure 2: Hospital payment schemes in selected OECD and EU member states, 1994 to 2015.

Notes: See Supporting Information for further information.

To consider other kind of major reforms, which might affect hospital activity at the aggregate level, we screened the health policy literature for the relevant control countries and excluded them if necessary (Clemens et al., 2014; Dubas-Jakóbczyk et al., 2020). However, we cannot control for the smaller gradual adaption of healthcare systems. Yet, it shows regularly difficult to provide consistent evidence for the effects of minor system changes (Milstein & Schreyögg, 2016; Purdy et al., 2012). It is therefore implausible to expect pronounced effects at the aggregate national level absent major reforms.

3.2. Variables

We use unbalanced country-level panel data from OECD sources, complemented by data from Eurostat (Eurostat, 2021; OECD, 2021a). Our main outcomes of interest concern the hospital activity and efficiency. We use hospital discharges and the average length of stay as outcome variables. Secondary outcome variables concern hospital resources and expenditures as well as population health status. As baseline, we control for changes in GDP per capita to account for possible budgetary constraints, e.g. caused by the impact of the financial crisis beginning 2007/2008. To capture time varying effects on the demand side, we follow previous empirical work and use the share of the population aged 65 or above (Wubulihasimu et al., 2016). More details on the data can be found in the Supporting Information.

Table 3 gives an overview. Even on aggregate level, certain health related data is not available. Table 3 furthermore underlines that Germany showed exceptional high values for several outcomes before and after the gDRG introduction. For our analysis, we transformed outcome and control variables to natural logarithms to allow straightforward interpretation of the results.

Table 3: Data Description

	Germany		Others[1]	
Variables	**2003**	**2012**	**2003**	**2012**
Hospital Discharges	201	230	160 (29)	155 (29)
Average Length of Stay	9.3	7.8	7.1 (28)	6.5 (29)
Share of population 65+	18 %	21 %	15 % (29)	16 % (29)
GDP	36977	42822	34326 (29)	37640 (29)
Inpatient Expenditures	1156	1357	852 (20)	891 (27)
Outpatient Expenditures	965	1103	770 (22)	848 (27)
Hospital Expenditures	1218	1432	1076 (20)	1275 (27)
Hospital Physicians	1.5	1.9	1.6 (17)	1.8 (25)
Hospital Nurses	4.2	4.4	4.9 (17)	5.0 (25)
Hospital Beds	6.6	6.2	4.5 (25)	3.7 (28)
Private Hospital Beds	2.1	2.5	0.6 (13)	0.6 (18)
Average Idle Bed Capacity	1.4	1.3	1.1 (19)	1.0 (21)
Potential Years of Life lost	5258	4162	6322 (28)	4907 (28)
Life Expectancy	78.6	80.6	77.5 (29)	80.0 (29)

[1] In brackets: Number of countries with data available.

Notes: Number of Discharges, (Idle) Beds, Physicians and Nurses per 1000 inhabitants. Potential years of life lost per 100000 inhabitants, aged 0-75. GDP and expenditures per inhabitant.

4. Methods

4.1. Empirical Approach

The general objective of our approach is to obtain unbiased estimates for the effect of the gDRG introduction on several outcomes. For a robust estimation, we make use of three complementary methods.

a. Difference-in-Differences (DID)
b. Synthetic Control (SC)
c. Synthetic Difference-in-Differences (SDID)

Although DID and SC are normally used in different empirical settings, they are closely related (Arkhangelsky et al., 2021): A standard *Difference-in-Differences* approach can be considered as an *unweighted* linear regression *with* unit and time fixed effects. Without covariates it can be expressed as follows (Arkhangelsky et al., 2021)

$$\left(\hat{\tau}^{DID}, \hat{\alpha}, \hat{\beta}\right) = \underset{\alpha,\beta,\tau}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \alpha_i - \beta_t - W_{it}\tau)^2 \right\}$$

with Y_{it} being the outcome of interest, α_i unit and β_t time fixed effects. W_{it} denotes a binary treatment and τ the treatment effect. The *Synthetic Control Method (Abadie et al., 2015; Abadie, 2021)* - "arguably the most important innovation in the policy evaluation literature in the last 15 years" (Athey & Imbens, 2017) - in contrast drops the unit fixed effects α_i and instead adds unit weights $\hat{\omega}_i^{SC}$ into the regression function (Arkhangelsky et al., 2021). It therefore can be considered as a *weighted* linear regression *without* unit fixed effects.

$$\left(\hat{\tau}^{SC}, \hat{\alpha}, \hat{\beta}\right) = \underset{\alpha,\beta,\tau}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \beta_t - W_{it}\tau)^2 \, \hat{\omega}_i^{SC} \right\}$$

The third, very recently proposed method *Synthetic Difference-in-Differences* combines aspects of a standard DID model and the SC estimator. Similar to DID it includes unit α_i and β_t time fixed effects. Like SC, it uses unit weights $\hat{\omega}_i^{SDID}$ to align pre-treatment outcome trends among treatment and control units. In contrast to SC, SDID allows for an intercept term in weight optimization. Thus, pre-treatment outcomes of control and treated units do not need to match at level, instead matching on trends is sufficient. SDID additionally incorporates time weights $\hat{\lambda}_t^{SDID}$ to balance pre- and post-treatment periods. The time weights are chosen so that the weighted average of pre-treatment outcomes predicts the average post-treatment outcome for each control unit up to a constant. Thereby time weights can improve estimation by diminishing the influence of pre-treatment periods that are very different from post-treatment periods (Arkhangelsky et al., 2021).

$$\left(\hat{\tau}^{SDID}, \hat{\alpha}, \hat{\beta}\right) = \underset{\alpha,\beta,\tau}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \alpha_i - \beta_t - W_{it}\tau)^2 \, \hat{\omega}_i^{SDID} \, \hat{\lambda}_t^{SDID} \right\}$$

Both sets of weights are then used in a two-way fixed effects regression to get an estimate for the average treatment effect. Arkhangelsky et al. (2021) demonstrate that SDID has attractive properties with regard to bias and variance compared to the SC and DID estimators.

In all three methods other developed countries, which were not exposed to payment scheme reforms of similar extent, function as control group (see Figure 2 and Supporting Information for a list of control units and additional information). The key assumption is therefore similar across all models: the outcome variables in Germany would have shown similar trajectories as the (weighted set of) control countries, absent the introduction of the German DRG-system.

a. Difference-in-Differences

A Difference-in-Differences model represents our baseline. Standard DID models estimate one time additive effects of a binary treatment on the outcome level. We slightly deviate and use the count of years since the DRG introduction (Z_{it}) as treatment indicator to allow for a structural trend break. We thereby follow previous literature on the effects of payment scheme reforms (Wubulihasimu et al., 2016). This approach is more similar to an event study design and allows for lasting dynamic treatment effects. With only one treated unit and hence absence of heterogeneous effects and varying treatment timing we do not have to consider the recent insights on continuous treatments (Callaway et al., 2021) or staggered DID (Goodman-Bacon, 2021). Our approach leads to the estimation of the following equation

$$Y_{it} = \alpha_i + \beta_t + \rho Z_{it} + \gamma X_{it} + \epsilon_{it}$$

with Y_{it} being the outcome of interest, α_i country fixed effects, β_t time fixed effects, X_{it} country-specific time-varying covariates.[3] Countries and years are indexed by i and t. The identifying assumption is that untreated potential outcomes evolve in parallel in the treatment and control group after conditioning on observables ("parallel trends"). The common way to back this assumption is to test for pre-treatment differences in trends. However, several recent papers have highlighted issues with this approach (Rambachan & Roth, 2021; Roth, 2022).

b. Synthetic Control Method

While nowadays applied to other settings and extended by different approaches (see Abadie (2021) for an overview), SC initially aimed at aggregate interventions affecting only one individual unit (Abadie et al., 2010; Abadie & Gardeazabal, 2003). Thereby the analysis of the gDRG introduction is a prime example for its "classic" application. The underlying idea is that – at the aggregate level – a weighted combination of unaffected controls might provide better comparisons than individual unaffected units. The selection of the control groups is formalized by a data driven procedure. In contrast to a standard DID model, SC does not require parallel outcome trends upfront (Abadie 2021). However, the identifying assumption is closely related. SC depends on finding a weighted set of control units, which matches the treated unit pre-treatment as closely as possible ("convex hull condition"). The identifying assumption then is that in the absence of treatment, the treated unit would follow the weighted set's outcome trajectory in the post-treatment periods.

Including a continuous treatment Z_{it} as in the DID framework is not possible for SC. Thus, we estimate an average treatment effect on a binary treatment indicator. Pre-intervention covariates are used in the weighting algorithm.

Critics of SC argue that it provides specification-searching opportunities since its results can be highly volatile (Ferman et al., 2020). This is caused by the inherent sparsity of weights, which can lead to widely different results using only slightly different specifications.

c. Synthetic Difference-in-Differences

The common way of backing parallel trends assumptions was recently disputed (Rambachan & Roth, 2021; Roth, 2022). In the case of our analysis, the assumption might not hold for some outcome variables (see Appendix Figure A. 10). The convex hull condition for SC is also difficult to guarantee for particular variables since Germany is "extreme" with

[3] By our definition of Z_{it} we assume a dynamic linear effect which can be interpreted as an annual year on year change.

regard to its hospital sector.[4] Such is that it may not be closely approximated by a synthetic control.

We therefore include the recently proposed *Synthetic Difference-in-Differences* (Arkhangelsky et al., 2021) into our approaches to bridge both methods and address these concerns. SDID compiles a weighted control group assuring approximately parallel trends by construction. Time weights diminish the role of periods that are very different from the considered post-treatment periods. Together, these adjustments make the estimation strategy more plausible and are close to the current empirical practice of upfront selecting suitable controls and periods, but in a more transparent way (Arkhangelsky et al., 2021). The identifying assumption is similar to SC: In the absence of treatment, the treated unit would follow the outcome trajectory of the time and unit weighted control set in the post-treatment periods.

Similar to SC, a continuous treatment Z_{it} is not possible in the SDID framework. Thus, we also estimate a binary treatment effect. The principal use case of the SDID framework is focused on outcomes only. Nonetheless, covariates X_{it} can be incorporated by applying SDID to the residuals of the regression of Y_{it} *on* X_{it} (Arkhangelsky et al., 2021).

Arkhangelsky et al. (2021) demonstrate that their estimation approach is at least as good as DID or SC with respect to variance and bias. They present different methods for inference. They recommend a bootstrap or jackknife variance estimation to conduct asymptotically valid inference. However, both methods are designed for settings with large panels and many treated units. For inference with $N_{tr} = 1$, i.e. only one treated unit, both approaches are not well defined. As third approach, Arkhangelsky et al. (2021) construct confidence intervals based on placebo evaluations, which are widely used in the SC framework. While it allows for some basic inference, it is likely that this method leads to rather too large intervals.

All analyses were carried out using R, version 4.0.3. Fixest, Augsynth and SynthDiD were primarily used as packages for estimation (Arkhangelsky et al., 2021; Ben-Michael et al., 2021; Bergé, 2018).

4.2. Robustness checks

The application of different methods already provides a certain level of robustness. We performed comprehensive additional checks. We include additional control variables, e.g. the total number of hospital beds, the respective share of private hospital beds or expenditure variables. We also tested whether our results critically depend on the choice of control countries. To this end, we used four different sets of control units. As described above, our main control group includes all countries that did not introduce any extensive reform of activity based funding between 1999 and 2011. Additionally, we built a set of control units where we excluded any country classified as activity based funding. A third, very narrow control group is based on fixed budget classification only. For the fourth set of

---

[4] E.g. the number of discharges as well as the average length of stay are both exceptionally high compared to other OECD countries, see Table 3.

control countries, we used an alternative classification of payment schemes (see Supporting Information).

The gDRG system became the obligatory remuneration system for all acute care hospitals in 2004. The budgetary consequences only started 2005 and were followed by a transition phase. While we therefore use 2005 as the year of intervention in our main analysis we repeat it with different treatment timings (up to t=2007).

Within the Synthetic Control framework, placebo tests are often recommended. The underlying idea is to replace the exposed unit with different units that were not exposed. By estimating a “placebo-in-space-test”, we are able to assess the magnitude of randomness in the data. Backdating the intervention is a similar approach, which can be applied to the SDID framework as well. It works as “placebo-in-time test” (Abadie, 2021) and assigns a different treatment timing instead of a different treatment unit. If the estimation is able to reproduce the trajectory of the outcome variable prior to treatment and does not estimate any backdated effect, it provides credibility to the estimation itself (Abadie, 2021). In addition, we also use recent advances and provide results from a conformal inference method for SC (Chernozhukov et al., 2021).

Finally, to establish a plausible link between the changes in observed outcomes and the gDRG introduction we also discuss alternative explanations for our results.

5. Results

5.1. a. Difference-in-Differences

The DID regression results are shown in Table 4. Model 1 represents the baseline specification, controlling for changes in GDP per capita and the share of the population aged 65 or above. Models 2 to 6 add additional control variables. On the supply side, we consider the number of hospital beds. Additionally, it could be hypothesized that rising private hospital bed numbers play a role and therefore also enter the regression. Overall healthcare expenditures control for varying scopes of healthcare cuts, e.g. due to the financial crisis. Furthermore, outpatient expenditures and doctor consultations work as additional proxies for demand side factors and possible shifts to other sectors.

For the main outcome hospital discharges, the coefficients remain stable in size and sign throughout all models. Thus, our estimation results seem robust to model alterations. The estimates for the structural growth effect $\rho_{[Hospital\ Discharges]}$ range from 0.019 to 0.037 and are significant at the 1-%-level across all models. This implies an additional growth in hospital discharges of at least 2% annually. We do not provide DID results for our other main outcome, average length of stay, because we do not consider the parallel trend assumption to hold (see discussion below).

Regarding our secondary outcomes, the results are rather mixed. We find a small, positive effect on the number of hospital physicians in our baseline model. Following the addition of other controls, effect size and significance vary. We do not find any effects on nursing staff levels. Interestingly, the results consistently indicate, that the German DRG introduction had a negative effect on the population health status. However, additional analysis with socioeconomic control variables, as e.g. unemployment rates, rendered the results insignificant.

The validity of our DID estimates is conditional on the assumption that absent the German DRG-reform the outcomes would follow a common trend. Although this assumption itself is untestable, parallel pre-treatment trends can provide some reassurance. To this end, Figure A. 10 depicts an event-study-plot with estimated coefficients and 95% confidence intervals for our main and secondary outcomes. The underlying regression equation is given by

$$Y_{it} = \alpha_i + \beta_t + \rho_t(Treated_i * Time_t) + \gamma X_{it} + \epsilon_{it}$$

where α_i and β_t refer to country and time fixed effects. $(Treated_i * Time_t)$ refers to the interaction terms of treatment and time dummies. They replace the single treatment variable Z_{it} and therefore lead to annual effect estimates ρ_t. The pre-intervention coefficients $\rho_{t[t<2005]}$ for hospital discharges are close to zero and not statistically significant. This holds true for all secondary outcomes as well. While some pre-treatment point-estimates are different from zero and might hint at diverging pre-trends, the confidence intervals always cover the zero. For our second main outcome, the average length of stay, there are clear signs of differing pre-trends. We infer that we can support the parallel trends assumption for most outcomes; but it does not hold for the latter.

Table 4: DID – Estimated impact of the gDRG introduction

	(1)		(2)		(3)		(4)		(5)		(6)	
Outcome	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**
Hospital Discharges	0.0286*** (0.0047)	530	0.0194** (0.0069)	478	0.0230** (0.0067)	235	0.0328*** (0.0058)	500	0.0367*** (0.0063)	415	0.0289** (0.0089)	390
Average Length of Stay												
Inpatient Expenditures	0.0081. (0.0047)	394	0.0019 (0.0048)	368	0.0035 (0.0044)	198	0.0162** (0.0056)	394	0.0089 (0.0063)	384	0.0030 (0.0039)	360
Hospital Physicians	0.0087** (0.0030)	316	0.0075. (0.0040)	303	0.0009 (0.0057)	171	0.0103** (0.0029)	306	0.0087* (0.0039)	254	0.0055 (0.0041)	250
Hospital Nurses	-0.0004 (0.0030)	314	-0.0010 (0.0027)	301	-0.0046. (0.0022)	160	0.0020 (0.0020)	304	-0.0002 (0.0026)	252	-0.0001 (0.0026)	248
Life Expectancy	-0.0010** (0.0003)	542	-0.0010* (0.0004)	480	-0.0010** (0.0003)	233	-0.0009* (0.0003)	513	-0.0012*** (0.0002)	423	-0.0014*** (0.0002)	394
Potential Years of Life lost	0.0069. (0.0034)	528	0.0041 (0.0038)	475	0.0050* (0.0022)	222	0.0055. (0.0031)	509	0.0062* (0.0027)	417	0.0076* (0.0029)	388
Stand. Death Rates	0.0065** (0.0018)	545	0.0050* (0.0022)	487	0.0054*** (0.0013)	234	0.0052** (0.0015)	515	0.0075*** (0.0011)	423	0.0090*** (0.0011)	394
Controls for:												
Baseline	✓		✓		✓		✓		✓		✓	
Hospital Beds			✓		✓						✓	
Private Hospital Beds					✓						✓	
Healthcare Expenditure							✓				✓	
Outpatient Expenditures									✓		✓	

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Notes: All estimates include country and year fixed effects, with outcomes and control variables in log form. Robust standard errors clustered at country level in brackets. Results for Average Length of Stay are not provided because we do not assume the parallel trend assumption to hold.

The event-study-analysis in Figure A. 10 is quite revealing in two additional ways. First, it underlines that the effect of the gDRG-introduction is dynamic. Instead of a sudden additive effect, it probably induced a structural trend break and should be better captured by our extended DID design. Second, it corroborates the findings of a delayed impact of the reform. Budgetary consequences only affected hospitals to varying degrees from 2005 on. A transitional phase is expected therefore.

5.2. b. Synthetic Control

Table 5 summarizes the results of the SC approach. In contrast to DID, a continuous treatment variable Z_{it} is not feasible. The estimates are therefore average effect estimates. For a better comparison with the DID results in Table 4, we added an annual linear interpretation of the effect estimates.[6] Inference is based on the conformal inference procedure from Chernozhukov et al. (2021). We estimated synthetic controls individually for each outcome and we provide the corresponding weights in the Appendix B (Table A. 1). Additionally, we made use of a feature of the Augmented Synthetic Control from Ben-Michael et al. (2021) and estimated a model with multiple outcomes used in the weighting algorithm at once.[7] We restricted it to our two main outcomes in order to achieve a sufficient fit.

In general, the results are – where significant – comparable in size and sign to DID. The estimated effect on hospital discharges is only insignificant in the multiple SC model (2) with additional covariates used in weighting. However, this model leads to a bad pre-intervention fit. The estimates regarding our second main outcome, the average length of stay, are close to zero and insignificant. In contrast to DID, we do not find effects on the number of hospital physicians.

The multiple outcomes approach results in a synthetic Germany composed of Czech Republic (50 %), Switzerland (21 %), Austria (15 %) and Norway (11 %). Due to Germany's special characteristics with one of the oldest populations worldwide and a lavish supply of beds, a perfect fit is not possible (see Table 6). Depending on the outcome, unit weights and therefore the synthetic Germany vary (see Appendix B Table A. 1).

[6] Derived from the ATT under the assumption of linear effects.
[7] In which case inference is based on bootstraps.

Table 5: SC – Estimated impact of the gDRG introduction

	(1)		(2)	
Outcome	**ATT Estimate**	**approx. Annual Growth**	**ATT Estimate**	**approx. Annual Growth**
Single SC				
Hospital Discharges	0.1300** (0.0080)	0.0247	0.1126* (0.0100)	0.0216
Average Length of Stay	-0.0109 (0.9380)	-0.0022	-0.0052 (1.0000)	-0.0010
Inpatient Expenditures	-0.0111 (0.9400)	-0.0022	-0.0391 (0.9250)	-0.0080
Hospital Physicians	0.0953 (0.2060)	0.0184	0.1186 (0.8730)	0.0227
Hospital Nurses	-0.0150 (0.4060)	-0.0030	-0.1053 (0.4330)	-0.0220
Life Expectancy	-0.0012** (0.0040)	-0.0002	-0.0019. (0.0660)	-0.0004
Potential Years of Life lost	0.0263* (0.0120)	0.0052	-0.0048* (0.0110)	-0.0010
Stand. Death Rates	0.0344 (0.1280)	0.0068	0.0198 (0.1520)	0.0039
Multi SC				
Hospital Discharges	0.1023* (0.0419)	0.0197	0.0422 (0.3211)	0.0083
Average Length of Stay	-0.0378 (0.7425)	-0.0077	0.0019 (0.9880)	0.0004
Variables used in weight determination:				
Baseline	✓		✓	
Hospital Beds			✓	
Healthcare Expenditure			✓	
Outpatient Expenditures			✓	

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Single SC: Outcome specific control group weights
Multi SC: One set of control group weights across all outcomes.
Notes: All estimates with outcomes and control variables in log form. P-values from the conformal inference procedure from Chernozhukov et al. (2021) are shown in brackets. An approximate interpretation of the ATT as (linear) annual growth was added for comparison with DID results.

Table 6: Comparison of Real and Synthetic Germany for t=2003

Country	Hospital Discharges	Average Length of Stay	GDP	Share of population 65+	Hospital Beds	Average Idle Bed Capacity
Germany	20,143.80	9.30	36,977.35	17.50	6.57	1.43
Synthetic Germany	19,630.37	9.47	33,094.54	14.58	5.42	0.97
Unweighted Average of Control Units	15,225.09	7.11	32,186.56	14.44	4.19	0.94

Notes: Weights based on Multi SC Model with Hospital Discharges and Average Length of Stay as outcomes and GDP per capita and share of population 65+ as controls. Hospital beds only shown for comparison purposes. All variables re-transformed from natural logarithm.

5.3. c. Synthetic Difference-in-Differences

Table 7 shows results for the Synthetic Difference-in-Differences approach. We restricted our estimations to the baseline control variables (GDP and share of population 65 plus). Since SDID needs balanced panel data, including additional control variables with fragmented availability would shrink the donor pool for control units prohibitively. SDID distributes weights across many countries. Across all outcomes, no country gets a weight higher than 20%. Time weights in contrast are very sparse. We provide the weights in the Appendix B Table A. 2.

The estimates for our outcomes are similar in sign and size to our previous results. As described in the methods section, the confidence intervals are based on placebo evaluations (Arkhangelsky et al., 2021). It is likely that this method leads to rather large intervals. In our case, it leads to 95-%-confidence intervals always including the null. The effect for hospital discharges is significant only at the 10-%-level. Notwithstanding these limitations, the SDID approach broadly confirms our findings.

Table 7: SDID – Estimated impact of the gDRG introduction

Outcome	**ATT Estimate**	**approx. Annual Growth**
Hospital Discharges	0.1246. (0.0886)	0.0238
Average Length of Stay	-0.0270 (0.0952)	-0.0055
Inpatient Expenditures	0.0501 (0.1099)	0.0098
Hospital Physicians	0.0492 (0.0844)	0.0097
Hospital Nurses	-0.0268 (0.0921)	-0.0054
Life Expectancy	-0.0035 (0.0050)	-0.0007
Potential Years of Life lost	0.0349 (0.0606)	0.0069
Stand. Death Rates	0.0222 (0.0413)	0.0044

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Notes: All estimates with outcomes and control variables in log form. GDP per capita and share of population 65+ used as control variables. Standard errors from placebo-evaluations in brackets. An approximate interpretation of the ATT as annual growth was added for comparison with DID results.

5.4. Further robustness checks

The sections above already include first robustness checks. As a comprehensive extension, we ran several additional checks. We focused on our main outcome, hospital discharges, since other estimates were inconclusive.

We checked the robustness of our estimates by using different sets of control units as described in section 4.2 across all three methods. Generally, signs and sizes of the estimates remained similar. Excluding all countries with comprehensive ABF reforms from our control group, independent of timing, led to mostly higher effect sizes. Using only fixed budget countries led to volatile results due to the remaining small sample size. Especially SC was prone to a high volatility. SDID in contrast seemed more robust to variations because it spreads weights more equally by design (see Appendix Table A.4 ff. for results).

Following the critique of Rambachan and Roth (2021) we also conducted a sensitivity analysis of our DID estimates to assess violations of the parallel trends assumption. The approach allows us to judge the validity of our results based on the observed worst-case violation in the pre-treatment period. The results are robust to roughly half the level of the maximal pre-treatment parallel trend violation (see Appendix Figure A. 11). We consider this sufficient since this maximal pre-treatment violation occurred several years before the intervention. Allowing for a non-linearity in the differential trend between Germany and the control countries that is about half the maximum observed in the pre-treatment period would also render our DID results insignificant (see Appendix Figure A. 11 and Table A.13).

For SC we estimated placebo effects for each control country as a placebo-in-space test. The change in Germany exceeds corresponding changes in all other countries (see Figure A. 12). This test statistic provides a placebo-p-value of 0.047 for SC. Germany also outranks the placebo treatments regarding the quality of fit, as the ratio of post- vs. pre-treatment root mean square prediction error underlines (also in Figure A. 12).

Placebo-in-time tests, i.e. bringing the reform forward to an earlier point in time, are generally difficult to conduct when limited data is available. Still, Figure A. 13 indicates, that our approach provides credible trajectories. Germany and its (synthetic) control only begin to diverge after the gDRG reform, despite backdating it to the year 2000. Repeating our analysis with postponed treatment timing ($t = 2006\ and\ 2007$) led to overall similar effect estimates with slightly higher annual values due to a shorter time span.

After the application of several robustness checks and different methods, the main threat to the credibility of our results remains the presence of other systematic reforms within Germany, which could explain the sudden change in hospital activity. If there were further reforms with an impact on hospital activity, an attribution of the estimated effects to the DRG reform would not be possible. To rule out other external factors, mainly two other areas concerning the hospital sector have to be considered (Dubas-Jakóbczyk et al., 2020): hospital governance or respective reforms in other sectors. The first dimension covers reforms with respect to hospital capacity, management and quality governance. The latter describes reforms in primary healthcare, long-term care or integrated care models.

In Germany, federal states are responsible for hospital capacity planning. Most planning processes evolve around a mere update of historically derived bed numbers. In general, these algorithms only use a limited set of parameters within a Hill-Burton formula, such as the number of residents in a region or the frequency of hospitalizations. Activity based planning or quality requirements play, even today, a negligible if any role. Hospital governance reforms therefore cannot explain any variation, especially not of this size, in hospital activity.

Regarding the relations with other sectors, a reform of the German outpatient payment system in 2007/2008 deserves some scrutiny. Until then, total outpatient remuneration (within the public health insurance system) was negotiated as capitation per insured person, covering all outpatient services within all specialties. Thus, additional services were financed by lowering the average value of every other service. With the overhaul of the payment mechanisms in 2007/2008 the remuneration changed to a morbidity-based system with (largely) fixed values for each activity. The declared goal was to shift the morbidity risk from physicians to sickness funds. In contrast to other countries (He & Mellor, 2012), this outpatient payment reform had no notable direct effect on hospitals in Germany since they provide almost exclusively inpatient services. However, changes in the provision of outpatient services can also have indirect effects on hospitals, either via ambulatory care sensitive conditions (Purdy et al., 2009) or by changed referral patterns. The German outpatient reform was accompanied by significant additional funding and the – scarcely available – data shows no hint of decreasing outpatient activity (see figure A.2). It is therefore unlikely that it caused additional inpatient hospital activity in line with our estimations.

Another possible event causing different trends among the considered countries could be the financial crisis 2007/2008 and the following European debt crisis. While these crises affected all countries including Germany – albeit to a different degree – and we control for GDP, we cannot completely rule out any influence. Some countries cut healthcare budgets significantly and the phased introduction of the gDRG system overlaps with these cuts. However, since our results are more driven by increases in hospital activity in Germany than by decreases in the control countries, a plausible channel through which the financial crises could have caused these changes is not apparent. Furthermore, first effects already become visible before the timing of the financial crisis.

5.5. Summary

For illustrative purposes, Figure 3 plots the results for our two main outcomes for all three methods.[8] Overall, the results are comparable. All methods find unequivocal effects for hospital discharges. The results indicate that the gDRG introduction increased the number of hospital discharges by over 2% per year. For the average length of stay, we were not able to identify a consistent effect using SC and SDID while DID results are not credible due to differential pre-trends.

Regarding the secondary outcomes, we find no or no unequivocal effects. Comparative plots are provided in the Appendix B. Conclusions regarding the in Germany intensely discussed physician and nursing staff levels are difficult due to limited pre-treatment data and – in the case of nursing staff – quite idiosyncratic trajectories. The downward trend in nursing staffing levels seems to have reversed after the gDRG introduction. However, any causal attributions are prohibited by data limitations.

A closer look at the weights emphasizes the differences of our estimation approaches (see Appendix B Table A. 1 and Table A. 2). The optimization procedure of SC assigns zero weights to several control units. Therefore, the control group for one outcome can be very different from the control group for another outcome. Slight changes of the included covariates, the considered period or the weight optimization algorithm can lead to very different control unit weights. Thus, the results are highly responsive to small variations. SDID and DID in contrast use comparatively distributed respectively uniform weights and are therefore less affected by individual country comparisons.

[8] Without covariates because within the SDID framework from Arkhangelsky et al. (2021) they are incorporated by applying SDID to the residuals of the regression of Y_{it} on X_i. Outcome trajectories are therefore not straightforward comparable in models with time-varying covariates. The corresponding weights are provided in the Appendix.

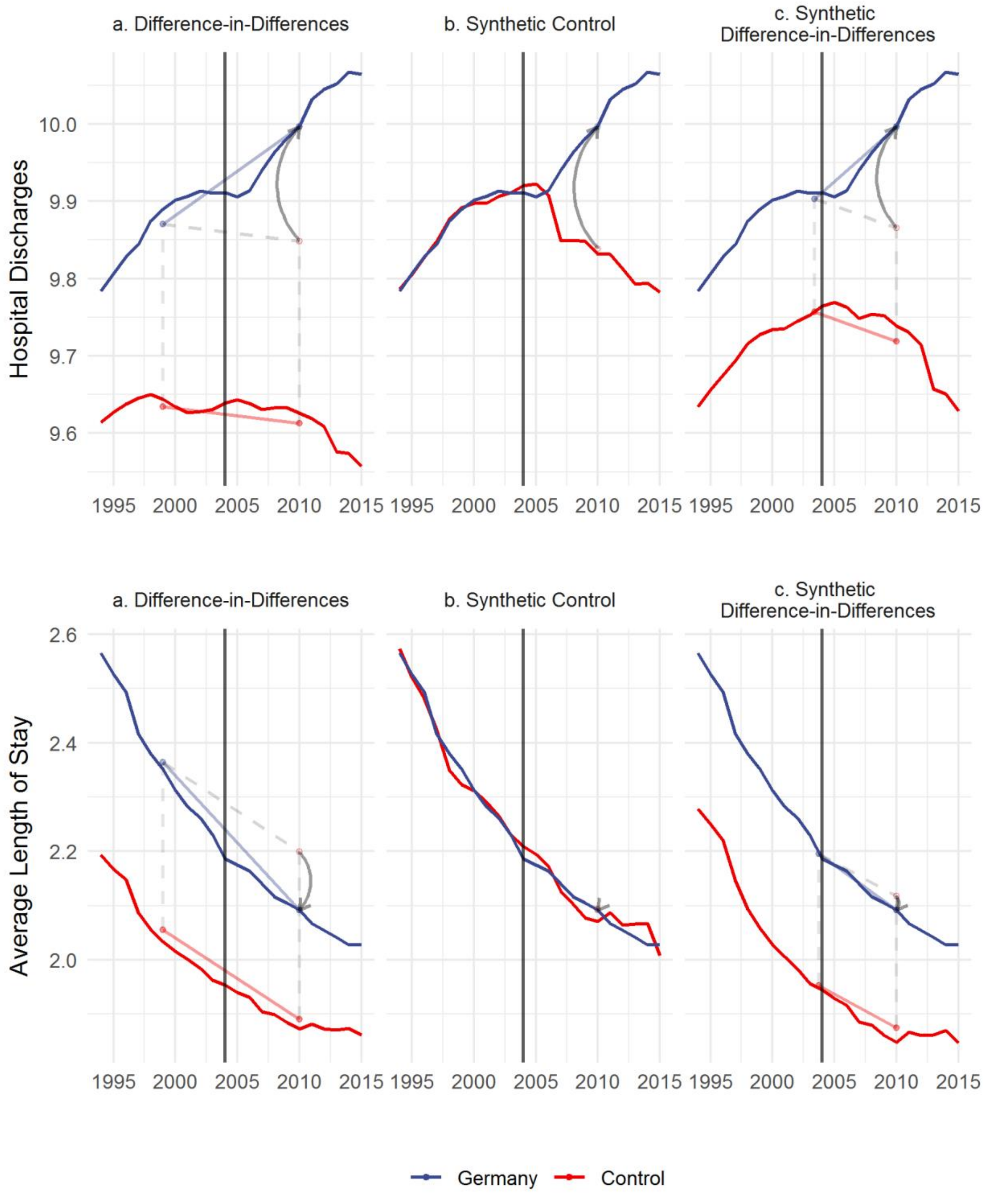


Figure 3: Illustrated impact of the gDRG introduction on main outcomes

6. Discussion

With this study, we contribute to the surprisingly thin literature on system-wide effects of hospital payment schemes by investigating the effects of the introduction of DRGs in Germany. So far non-experimental, descriptive studies remain the most commonly used method for assessing effects of hospital financing. Therefore, recent reviews have highlighted the importance of quasi-experimental approaches (Valentelyte et al., 2021). Our study addresses this gap by applying recent econometric advances on a comprehensive reform.

We use a country-level aggregate data set to compare the development of German hospital activity despite not having a control group in Germany itself. Due to this panel structure, we are able to control for underlying aggregate trends regarding economic and demographic variables as well as unobserved across-country influences. The case of Germany is particularly interesting for two reasons. First, it experienced one of the largest rises in hospital discharges among OECD and EU countries. Second, many countries implemented DRGs only partially, for certain aspects of hospital funding or restricted to individual regions. In contrast, the German DRG system is one of the most comprehensive implementations of activity based funding worldwide. It functions as sole pricing, billing and budgeting system and almost exclusively determines the turnover of hospitals.

Our triple quasi-experimental estimation approach allows us to combine individual advantages of each method and to gauge an ensemble estimate. While all methods perform differently depending on the real data generating process, our results unanimously indicate that the gDRG reform steeply increased hospital discharges. We estimate a prolonged effect of approximately 2% annually, meaning that over our estimation period of 10 years discharges increased by more than 20% on a per capita basis. To put it into perspective: Based on the average case costs in 2005 and everything else being equal, it led to additional expenditures of over one billion Euro added on top of Germany's hospital expenditures each year. Starting from an already very high level of hospital activity, the gDRG reform put an even stronger focus on inpatient care. This finding is somewhat contrary to previous studies of DRG reforms, which could not identify unequivocal evidence for an increased volume of care (Meng et al., 2020; Palmer et al., 2014). However, the majority of studies analyzed specific procedures and diagnosis and not a system-wide impact. Our results are robust across all models and methods.

In contrast to some of the existing literature, we were not able to demonstrate any effects on the length of stay. While absence of evidence is no evidence for absence, following the theoretical incentive structure one would expect to find pronounced declines. Internationally, a majority of studies found such connections between DRGs and initial declines in length of stays (Koné et al., 2019; O'Reilly et al., 2012). However, the few studies for Germany often did not consider existing pre-trends (Lotter et al., 2014; Reinhold et al., 2009). Some previous research on the gDRG reform showed a strong correlation between the change in the number of cases and the change in the length of stay. Hospitals with stronger decreases in the average length of stay experienced a particularly large number of additional cases (Schreyögg et al., 2014). Yet, the analysis was limited to data after the reform, unable to cover preexisting trends. For the closely related Swiss DRG reform Busato and Below (2010) did not find any impacts on the length of stay. Moreover, more recent studies on the gDRG introduction using quasi-experimental approaches found no overall effect on the length of stay either (Feess et al., 2019). Instead, the length of stay only changed for subsets of the population depending on patient and hospital characteristics.

A probable reason for our somewhat unexpected finding is the high idle hospital capacity of Germany (see Table 3). Due to lavish technical resources, German hospitals did not have to reduce the length of stay in order to provide the economically induced increase in the number of discharges. Our results seem to suggest that the overall prevailing trend of shorter stays was sufficient to accommodate more patients.

We also find less pronounced effects on population health and the number of hospital physicians. However, attributing these effects to the gDRG reform proves harder for two reasons. First, the results are not as robust. The effects on population health become insignificant when controlling for socioeconomic changes while the effects on hospital physicians are only significant within the DID framework under the parallel trends assumption. Second, the uptake of these effects coincides with probable alternative explanations. The effects on population health start after the financial crisis. Regarding hospital physicians and their working hours the European Court of Justice issued a judgement (C-151/02) of great importance in 2003; probably raising the number of physicians needed.

We were not able to identify any effects on the nursing staffing level. An idiosyncratic trajectory in Germany and data limitations prohibited any attribution to the gDRG reform. Thus, this highly relevant question for Germany and probably other countries remains open for further research.

Our study has several limitations. First, even on aggregate level only limited data is available for the considered period. Depending on the specific variable, data starts being available beginning mid-1990s. However, for many variables only few pre-intervention data exist, sometimes limited to selected countries. Therefore, data scarcity influences the choice of controls. Estimating counterfactuals based on a more comprehensive data set might have resulted in different estimates. On the other hand, the use of country-level data limits the extent to which low-level data errors data can affect estimation. Data quality issues might affect analysis on the level of individual hospitals or cases but are averaged out on country-level.

Second, other unobserved factors might have influenced hospital activity in Germany and the control countries. We include different variables to control for time varying factors. However, there are limits to how far controlling is possible with aggregated data. Finally, the construction of suitable control groups poses a limitation in itself. Our payment scheme classification is at least subject to certain interpretations.

7. Conclusion

Research on the effects of payment reforms is surprisingly thin. Our paper builds on an especially comprehensive payment scheme reform in Germany to reduce this gap. To the best of our knowledge, we provide the first cross-country empirical analysis of the German DRG-Introduction. Using aggregate panel data, we benefit from recent methodological advances and find a pronounced influence on hospital activity by strongly increased discharge numbers. Somewhat unexpectedly, we do not find any evidence for a shorter length of stay. Many countries introduced DRGs to enable additional hospital activity by decreasing the length of stay. In Germany however, the DRG reform targeted efficiency via shorter hospital treatment episodes as major goal itself, without the intention of increased hospital activity.

Our results complement two different strands of literature. First, it adds to the ongoing policy discussion in Germany over the long-term effects of the gDRG reform. For German hospitals, DRGs are the almost exclusive source of revenue. That is why the changed financial incentives induced steeply increased hospital discharges. However, German hospitals are also equipped with high (idle) capacities, which presumably explains why the reform did not reduce the overall length of stay. Based on our results and judged only on an aggregate level, the gDRG reform failed to achieve one of its major goals.

Second, our results also add to the overarching literature regarding the impact of activity based funding. Our evidence suggests that DRGs can lead to an increase in discharge rates but do not necessarily decrease the length of stay. Using quasi-experimental approaches on a comprehensive reform, our research adds relevant insights to the so far thin literature on system-wide effects of hospital financing. However, given the limitations of our study, these findings must still be interpreted with due caution.

Overall, our results suggests that hospitals do indeed respond to incentives induced by payment scheme reforms and effects are visible even on a system-wide level. As a result, policy makers should carefully align desired changes with the incentives caused by payment mechanisms. There is a need of pre-specified and subsequently thoroughly evaluated policy goals. Moreover, policy makers should be cautious when transferring previously found effects to other jurisdictions. The introduction of DRGs in Germany underlines the possibility of unexpected consequences within a different context.

References

Abadie, A. (2021). Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature*, *59*(2). https://doi.org/10.1257/jel.20191450

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, *59*(2), 495–510. https://doi.org/10.2139/ssrn.1950298

Abadie, A., & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, *93*(1), 113–132. https://doi.org/10.1257/000282803321455188

Arkhangelsky, D., Athey, S., Hirshberg, D., Imbens, G., & Wager, S. (2021). Synthetic Difference In Differences. *American Economic Review*, *111*(12), 4088–4118. https://doi.org/10.3386/w25532

Athey, S., & Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

Bäuml, M. (2020). How do hospitals respond to cross price incentives inherent in diagnosis-related groups systems? The importance of substitution in the market for sepsis conditions. *Health Economics*, 1–18. https://doi.org/10.1002/hec.4215

Bäuml, M., & Kümpel, C. (2021). Hospital responses to the refinement of reimbursements by treatment intensity in DRG systems. *Health Economics*, *30*(3), 585–602. https://doi.org/10.1002/hec.4204

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The Augmented Synthetic Control Method. *Journal of the American Statistical Association*, 1–34. https://doi.org/10.1080/01621459.2021.1929245

Bergé, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. *CREA Discussion Papers*(13).

Busato, A., & Below, G. von (2010). The implementation of DRG-based hospital reimbursement in Switzerland: A population-based perspective. *Health Research Policy and Systems*, *8*, 31. https://doi.org/10.1186/1478-4505-8-31

Busse, R., Geissler, A., Quentin, W., & Wiley, M. M. (Eds.). (2011). *European Observatory on Health Systems and Policies series*. *Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. Open University Press.

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2021). *Difference-in-Differences with a Continuous Treatment*. http://arxiv.org/pdf/2107.02637v2

Chaisemartin, C. de, & D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, *110*(9), 2964–2996. https://doi.org/10.1257/aer.20181169

Chernozhukov, V., Wüthrich, K., & Zhu, Y. (2021). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls. *Journal of the American Statistical Association*, 1–16. https://doi.org/10.1080/01621459.2021.1920957

Clemens, T., Michelsen, K., Commers, M., Garel, P., Dowdeswell, B., & Brand, H. (2014). European hospital reforms in times of crisis: Aligning cost containment needs with plans for structural redesign? *Health Policy (Amsterdam, Netherlands)*, *117*(1), 6–14. https://doi.org/10.1016/j.healthpol.2014.03.008

Cook, A., & Averett, S. (2020). Do hospitals respond to changing incentive structures? Evidence from Medicare's 2007 DRG restructuring. *Journal of Health Economics*, *73*, 102319. https://doi.org/10.1016/j.jhealeco.2020.102319

Cots, F., Chiarello, P., Salvador, X., Castells, X., & Quentin, W. (2011). Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals. In R. Busse, A. Geissler, W. Quentin, & M. M. Wiley (Eds.), *European Observatory on Health Systems and Policies series. Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals* (pp. 75–92). Open University Press.

Dafny, L. S. (2005). How Do Hospitals Respond to Price Changes? *The American Economic Review*, *95*(5), 1525–1547. https://doi.org/10.1257/000282805775014236

Di Giacomo, M., Piacenza, M., Siciliani, L., & Turati, G. (2017). Do public hospitals respond to changes in DRG price regulation? The case of birth deliveries in the Italian NHS. *Health Economics*, *26 Suppl 2*, 23–37. https://doi.org/10.1002/hec.3541

Dubas-Jakóbczyk, K., Albreht, T., Behmane, D., Bryndova, L., Dimova, A., Džakula, A., Habicht, T., Murauskiene, L., Scîntee, S. G., Smatana, M., Velkey, Z., & Quentin, W. (2020). Hospital reforms in 11 Central and Eastern European countries between 2008 and 2019: A comparative analysis. *Health Policy (Amsterdam, Netherlands)*, *124*(4), 368–379. https://doi.org/10.1016/j.healthpol.2020.02.003

Eurostat. (2021). *Health Database*. https://ec.europa.eu/eurostat/web/health/data/database

Feess, E., Müller, H., & Wohlschlegel, A. (2019). Reimbursement schemes for hospitals: the impact of case and firm characteristics. *Applied Economics*, *51*(15), 1647–1665. https://doi.org/10.1080/00036846.2018.1528334

Ferman, B., Pinto, C., & Possebom, V. (2020). Cherry Picking with Synthetic Controls. *Journal of Policy Analysis and Management*, *39*(2), 510–532. https://doi.org/10.1002/pam.22206

Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, *225*(2), 254-277. https://doi.org/10.3386/w25018
Hafsteinsdottir, E. J. G., & Siciliani, L. (2010). Drg prospective payment systems: Refine or not refine? *Health Economics*, *19*(10), 1226–1239. https://doi.org/10.1002/hec.1547
He, D., & Mellor, J. M. (2012). Hospital volume responses to Medicare's Outpatient Prospective Payment System: Evidence from Florida. *Journal of Health Economics*, *31*(5), 730–743. https://doi.org/10.1016/j.jhealeco.2012.06.001
Jürges, H., & Köberlein, J. (2015). What explains DRG upcoding in neonatology? The roles of financial incentives and infant health. *Journal of Health Economics*, *43*, 13–26. https://doi.org/10.1016/j.jhealeco.2015.06.001
Koné, I., Zimmermann, B. M., Nordström, K., Elger, B. S., & Wangmo, T. (2019). A scoping review of empirical evidence on the impacts of the DRG introduction in Germany and Switzerland. *The International Journal of Health Planning and Management*, *34*(1), 56–70. https://doi.org/10.1002/hpm.2669
Kutz, A., Gut, L., Ebrahimi, F., Wagner, U., Schuetz, P., & Mueller, B. (2019). Association of the Swiss Diagnosis-Related Group Reimbursement System With Length of Stay, Mortality, and Readmission Rates in Hospitalized Adult Patients. *JAMA Network Open*, *2*(2), e188332. https://doi.org/10.1001/jamanetworkopen.2018.8332
Lave, J. R. (1989). The effect of the Medicare prospective payment system. *Annual Review of Public Health*, *10*, 141–161. https://doi.org/10.1146/annurev.pu.10.050189.001041
Lotter, O., Stahl, S., Beck, M., Loewe, W., & Schaller, H.-E. (2014). Entwicklung der Fallpauschalenvergütung in verschiedenen chirurgischen Fachgebieten [Development of diagnosis-related groups in different surgical disciplines]. *Zentralblatt für Chirurgie - Zeitschrift für Allgemeine, Viszeral-, Thorax- und Gefäßchirurgie*, *139 Suppl 2*(S 02), e109-15. https://doi.org/10.1055/s-0031-1271532
Meng, Z., Hui, W., Cai, Y., Liu, J., & Wu, H. (2020). The effects of DRGs-based payment compared with cost-based payment on inpatient healthcare utilization: A systematic review and meta-analysis. *Health Policy*, *124*(4), 359–367. https://doi.org/10.1016/j.healthpol.2020.01.007
Milstein, R., & Schreyögg, J. (2016). Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. *Health Policy*, *120*(10), 1125–1140. https://doi.org/10.1016/j.healthpol.2016.08.009
Moreno-Serra, R., & Wagstaff, A. (2010). System-wide impacts of hospital payment reforms: Evidence from Central and Eastern Europe and Central Asia. *Journal of Health Economics*, *29*(4), 585–602. https://doi.org/10.1016/j.jhealeco.2010.05.007
OECD. (2021a). *Health statistics*. https://doi.org/10.1787/data-00900-en

OECD. (2021b). *Health statistics: Health expenditure and financing* [Inpatient curative and rehabilitative care as share of current expenditure on health (all functions) for hospitals as provider]. https://doi.org/10.1787/data-00900-en
O'Reilly, J., Busse, R., Häkkinen, U., Or, Z., Street, A., & Wiley, M. (2012). Paying for hospital care: The experience with implementing activity-based funding in five European countries. *Health Economics, Policy, and Law*, *7*(1), 73–101. https://doi.org/10.1017/S1744133111000314
Palmer, K. S., Agoritsas, T., Martin, D., Scott, T., Mulla, S. M., Miller, A. P., Agarwal, A., Bresnahan, A., Hazzan, A. A., Jeffery, R. A., Merglen, A., Negm, A., Siemieniuk, R. A., Bhatnagar, N., Dhalla, I. A., Lavis, J. N., You, J. J., Duckett, S. J., & Guyatt, G. H. (2014). Activity-based funding of hospitals and its impact on mortality, readmission, discharge destination, severity of illness, and volume of care: A systematic review and meta-analysis. *PloS One*, *9*(10), e109975. https://doi.org/10.1371/journal.pone.0109975
Purdy, S., Griffin, T., Salisbury, C., & Sharp, D. (2009). Ambulatory care sensitive conditions: Terminology and disease coding need to be more specific to aid policy makers and clinicians. *Public Health*, *123*(2), 169–173. https://doi.org/10.1016/j.puhe.2008.11.001
Purdy, S., Paranjothy, S., Huntley, A., Thomas, R. L., Mann, M. K., Huws, D. W., Brindle, P., & Elwyn, G. (2012). *Interventions to reduce unplanned hospital admission: a series of systematic reviews: Funded by National Institute for Health Research*. Final report 2012.
Rambachan, A., & Roth, J. (2021). *An Honest Approach to Parallel Trends: Working Paper.*
Reinhold, T., Thierfelder, K., Müller-Riemenschneider, F., & Willich, S. N. (2009). Gesundheitsökonomische Auswirkungen der DRG-Einführung in Deutschland--eine systematische Ubersicht [Health economic effects after DRG-implementation--a systematic overview]. *Das Gesundheitswesen*, *71*(5), 306–312. https://doi.org/10.1055/s-0028-1119399
Roth, J. (2022). Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends. *American Economic Review: Insights*(Forthcoming).
Salm, M., & Wübker, A. (2020). Do hospitals respond to decreasing prices by supplying more services? *Health Economics*, *29*(2), 209–222. https://doi.org/10.1002/hec.3973
Schreyögg, J. (2019). Changes in Hospital Financing and Organization and Their Impact on Hospital Performance. *Oxford Research Encyclopedia of Economics and Finance.* Advance online publication. https://doi.org/10.1093/acrefore/9780190625979.013.380
Schreyögg, J., Bäuml, M., Krämer, J., Dette, T., Busse, R., & Geissler, A. (2014). *Forschungsauftrag zur Mengenentwicklung nach § 17b Abs. 9 KHG: Endbericht*. Hamburg.

Shleifer, A. (1985). A Theory of Yardstick Competition. *The RAND Journal of Economics*, *16*(3), 319. https://doi.org/10.2307/2555560
Valentelyte, G., Keegan, C., & Sorensen, J. (2021). Analytical methods to assess the impacts of activity-based funding (ABF): A scoping review. *Health Economics Review*, *11*(1), 17. https://doi.org/10.1186/s13561-021-00315-1
Verzulli, R., Fiorentini, G., Lippi Bruni, M., & Ugolini, C. (2017). Price Changes in Regulated Healthcare Markets: Do Public Hospitals Respond and How? *Health Economics*, *26*(11), 1429–1446. https://doi.org/10.1002/hec.3435
WHO. (2021). *European Observatory on Health System and Policies*. https://eurohealthobservatory.who.int/
Wubulihasimu, P., Brouwer, W., & van Baal, P. (2016). The Impact of Hospital Payment Schemes on Healthcare and Mortality: Evidence from Hospital Payment Reforms in OECD Countries. *Health Economics*, *25*(8), 1005–1019. https://doi.org/10.1002/hec.3205

Appendix A – General

Trajectory in Germany compared to Europe

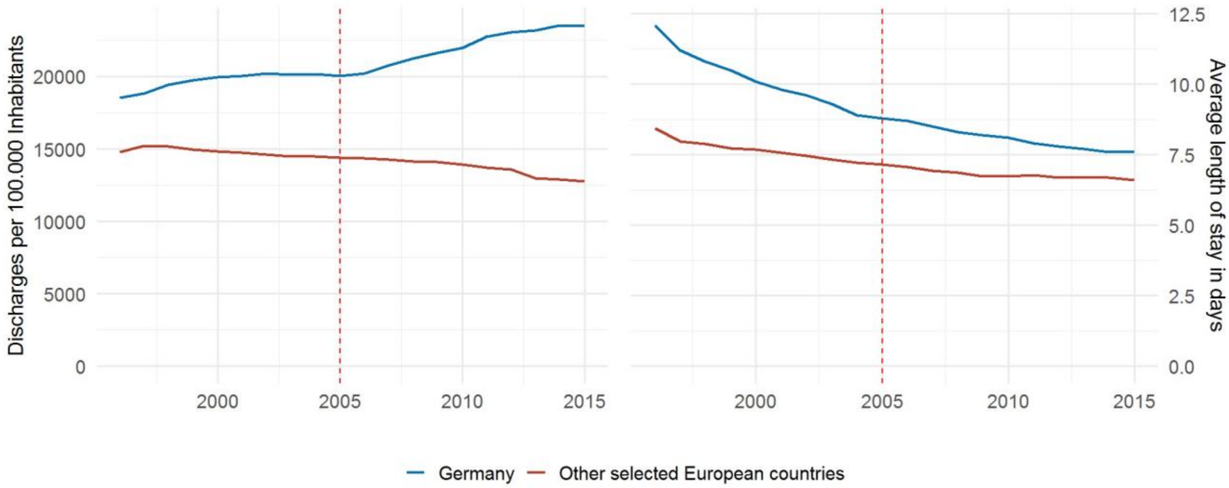


Figure A. 1: Curative care hospitalizations and length of stay.

Source: OECD 2021

Outpatient exposure

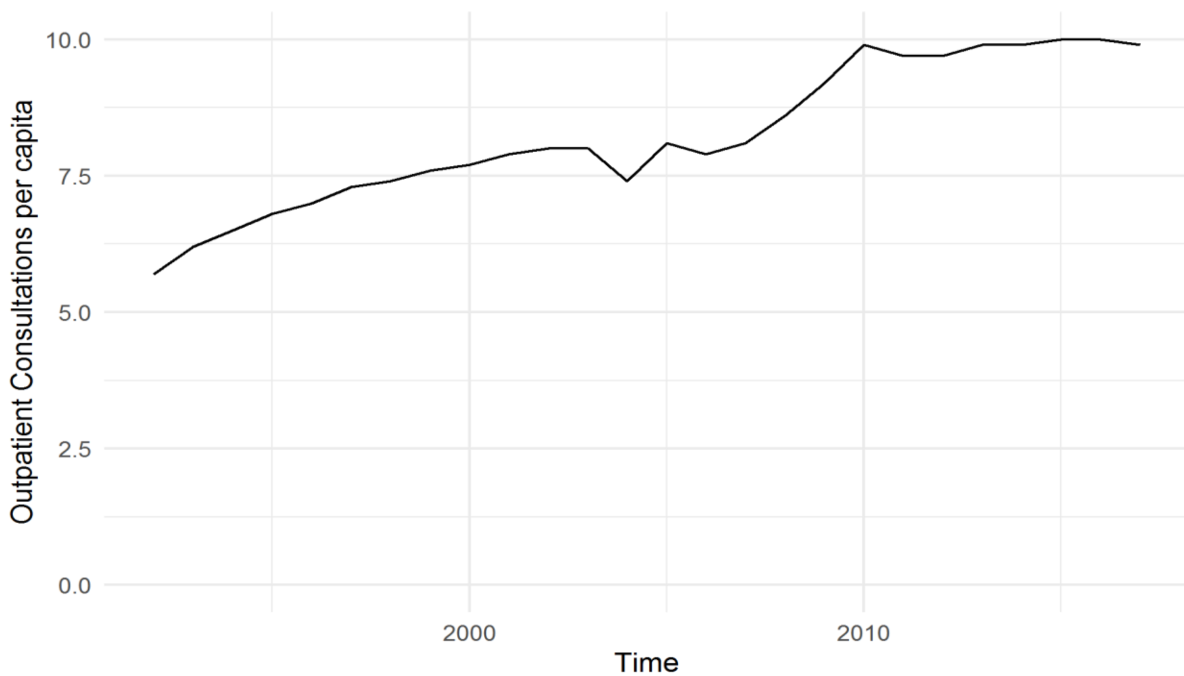


Figure A. 2: Doctor’s consultation outside of hospital per capita in Germany

Source: OECD 2021

Appendix B – Additional results

SC country weights

Table A. 1: SC - Weights for baseline model

Country	Hospital Discharges	Average Length of Stay	Inpatient Expenditures	Hospital Physicians	Hospital Nurses	Life Expectancy	Potential Years of Life lost	Stand. Death Rates	Multiple (main) Outcomes
AUS									
AUT	0.48		0.46			0.14	0.25	0.04	0.16
BEL			0.34	0.22			0.18	0.22	
CAN									
CHE		0.68		0.16	0.06				0.21
CYP									
CZE		0.27							0.52
ESP				0.33	0.18				
FIN							0.03		
GRC	0.16		0.20						
HUN						0.06	0.07		
IRL									
ISL									
ITA	0.26			0.28	0.58	0.52	0.39	0.28	
LTU		0.05			0.01	0.08	0.01		
LUX						0.09	0.07	0.15	
LVA								0.03	
NLD					0.18				
NOR	0.09					0.11			0.11
NZL									
PRT								0.23	
SVK									
SVN									
SWE	0.01							0.06	

SDID country weights

Table A. 2: SDID – Country weights for baseline model

Country	Hospital Discharges	Average Length of Stay	Inpatient Expenditures	Hospital Physicians	Hospital Nurses	Life Expectancy	Potential Years of Life lost	Stand. Death Rates
AUS	0.00	0.03	0.06	0.05	0.05	0.04	0.05	0.05
AUT	0.13	0.05	0.07	0.07	0.08	0.07	0.05	0.05
BEL			0.07	0.06	0.05	0.03	0.04	0.04
CAN	0.00	0.01	0.06		0.00	0.03	0.04	0.04
CHE	0.00	0.05		0.07	0.13	0.05	0.05	0.04
CYP	0.06	0.06		0.08	0.04	0.01		
CZE	0.02	0.04	0.02			0.07	0.05	0.05
ESP	0.13	0.05	0.08	0.07	0.06	0.04	0.05	0.04
FIN	0.03	0.00	0.06	0.07		0.04	0.04	0.04
GRC	0.17	0.05	0.08	0.05	0.01	0.02	0.04	0.04
HUN	0.09	0.06	0.04			0.06	0.04	0.04
IRL	0.00	0.02		0.07	0.04	0.02	0.04	0.03
ISL	0.01	0.09	0.08			0.04	0.03	0.04
ITA	0.05	0.05		0.08	0.11	0.06	0.05	0.05
LTU	0.13	0.08		0.08	0.13	0.11	0.04	0.05
LUX			0.05			0.06	0.06	0.06
LVA							0.03	0.03
NLD	0.00	0.08	0.06	0.03	0.13	0.02	0.04	0.04
NOR	0.08	0.04	0.06	0.06		0.02	0.03	0.03
NZL				0.07	0.10	0.06	0.05	0.05
PRT	0.05	0.04	0.07			0.06	0.05	0.05
SVK	0.00	0.08				0.01	0.04	0.04
SVN	0.06	0.08	0.08	0.07	0.06	0.05	0.05	0.05
SWE	0.00	0.04	0.06			0.02	0.03	0.04

SDID time weights

Table A. 3: SDID – Time weights for baseline model

Year	Hospital Discharges	Average Length of Stay	Inpatient Expenditures	Hospital Physicians	Hospital Nurses	Life Expectancy	Potential Years of Life lost	Stand. Death Rates
1994	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1995	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
1996	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1997	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1998	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00
2002	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37
2003	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2004	0.95	0.97	1.00	0.92	1.00	1.00	1.00	0.63

Corresponding weights of Figure 3

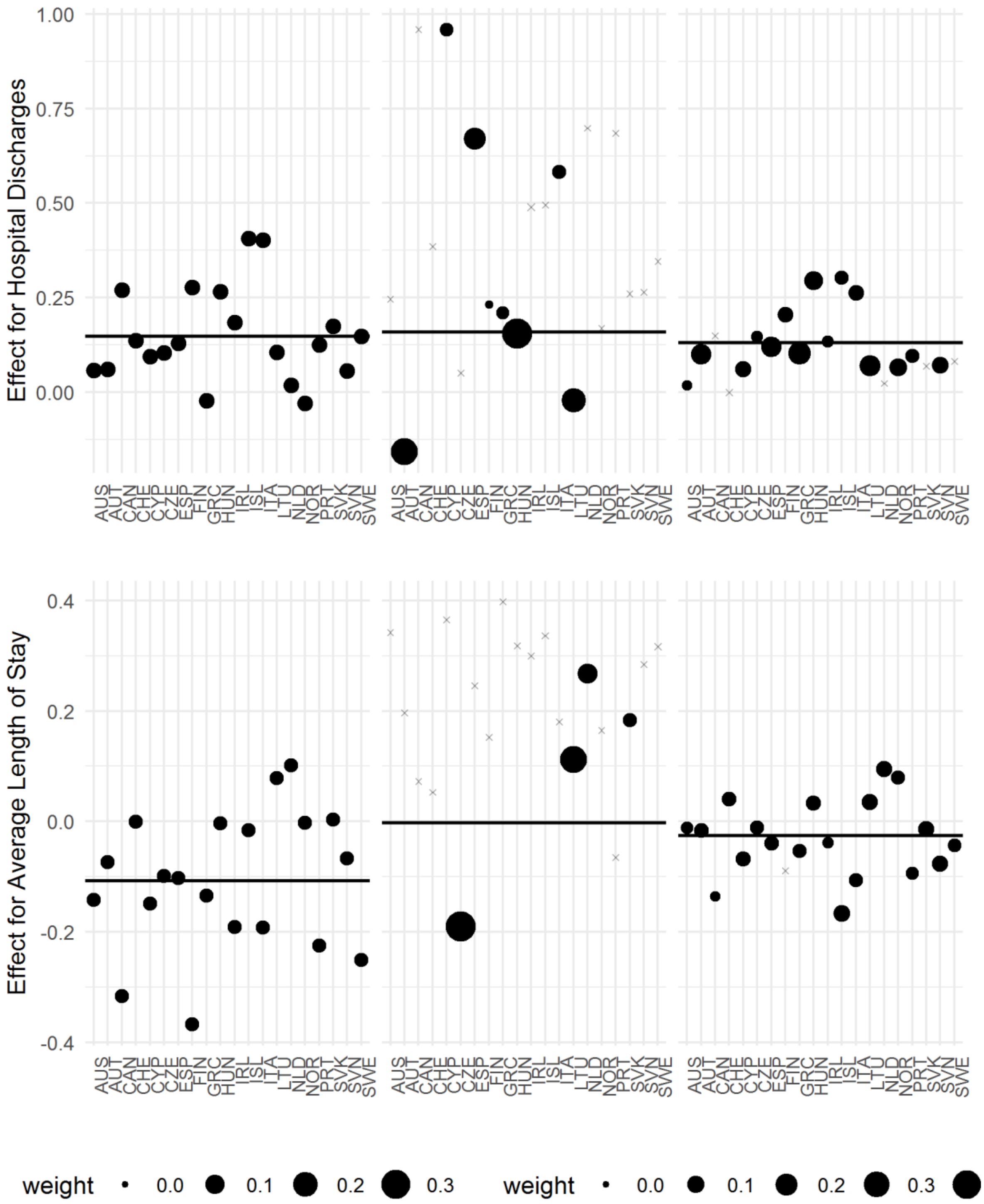


Figure A. 3: Corresponding weights of Figure 3

Graphical Illustration for Number of Hospital Nurses

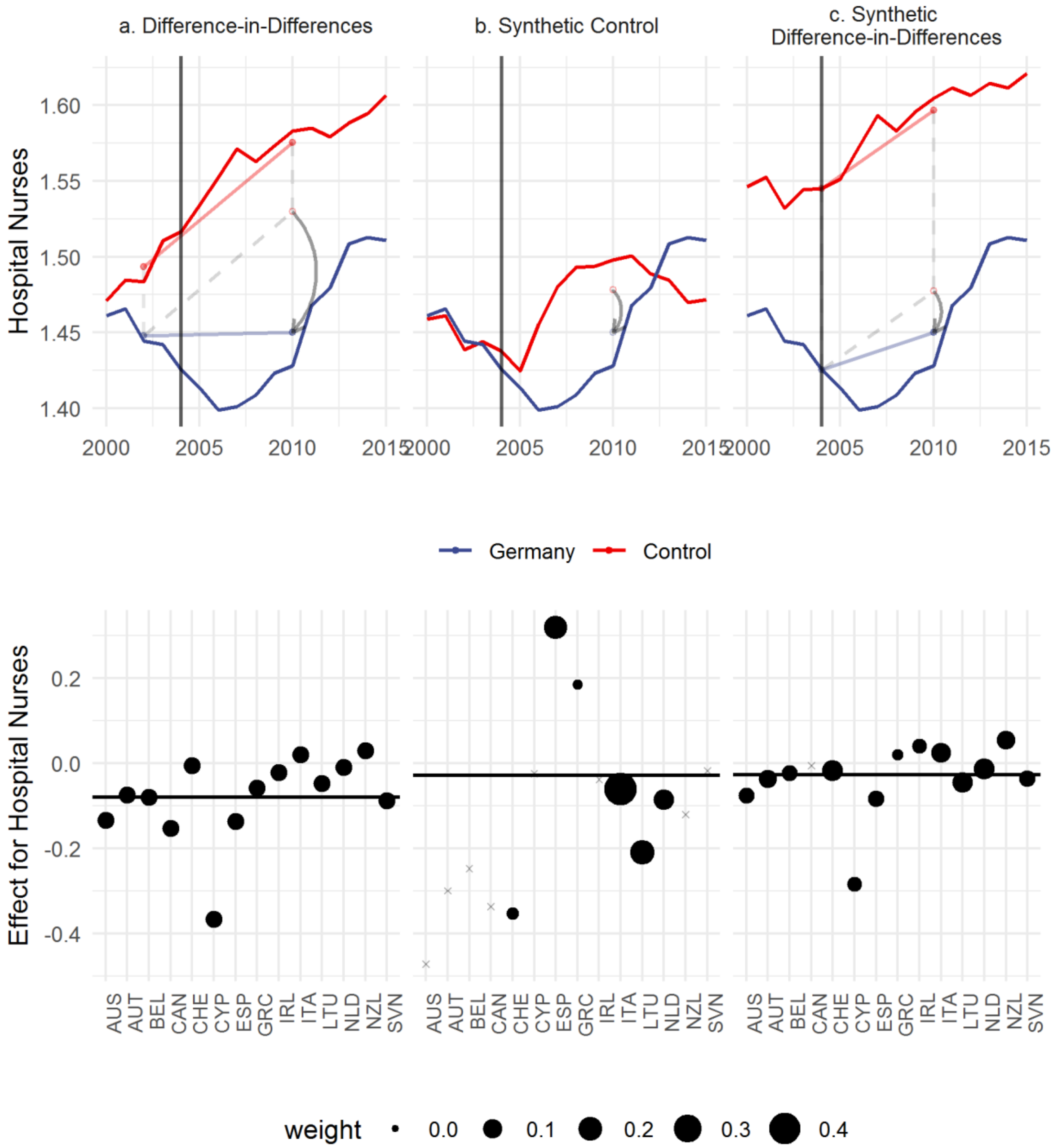


Figure A. 4: Estimated impact of the gDRG introduction on hospital nurses

Notes: All calculations without covariates. In the upper half, overall effects are shown. Below, country-by-country outcome differences with weights indicated by dot size are plotted. The black horizontal line indicates the overall estimate.

Graphical Illustration for Number of Hospital Physicians

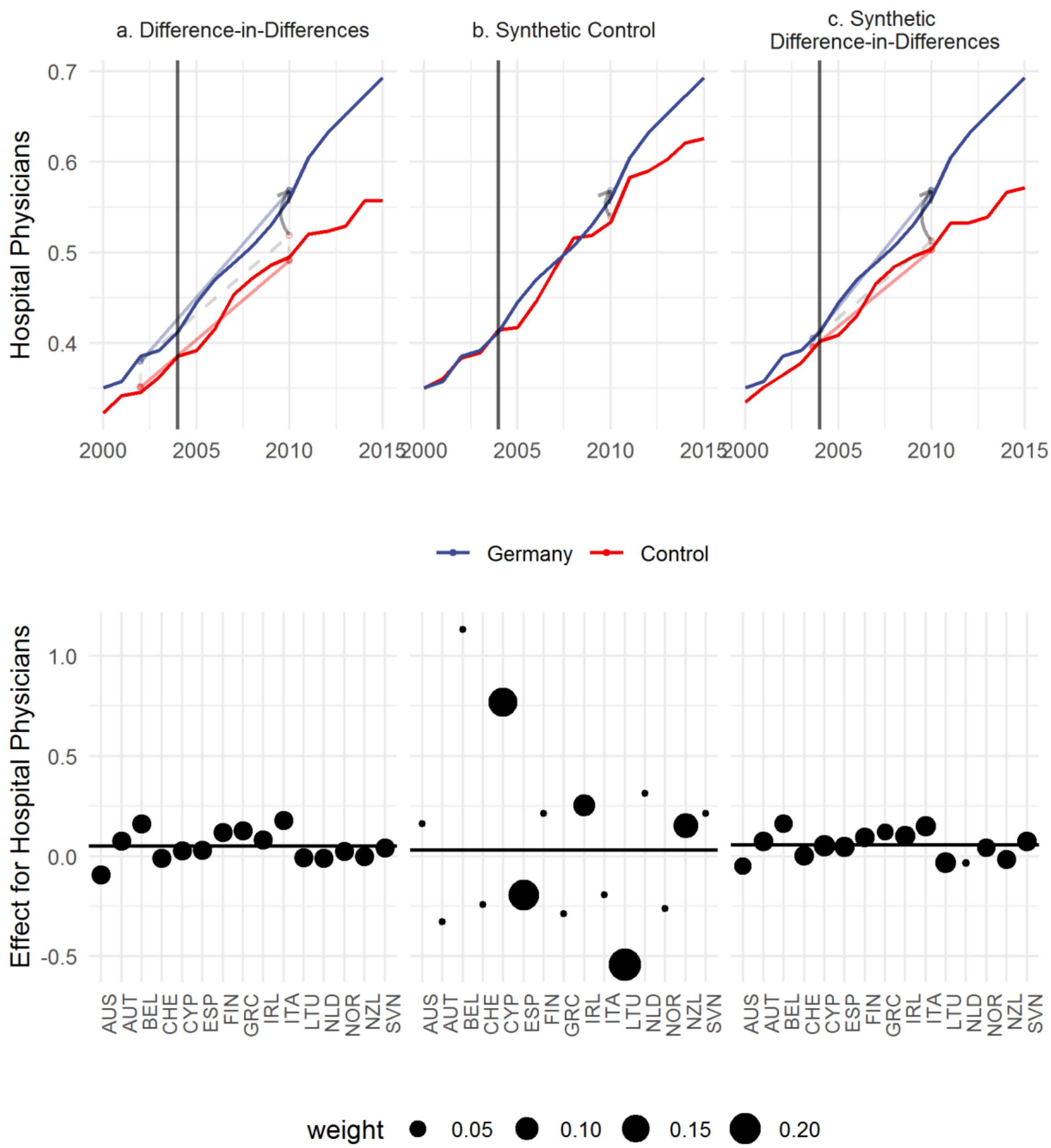


Figure A. 5: Estimated impact of the gDRG introduction on hospital physicians

Notes: All calculations without covariates. In the upper half, overall effects are shown. Below, country-by-country outcome differences with weights indicated by dot size are plotted. The black horizontal line indicates the overall estimate.

Graphical Illustration for Inpatient Expenditures

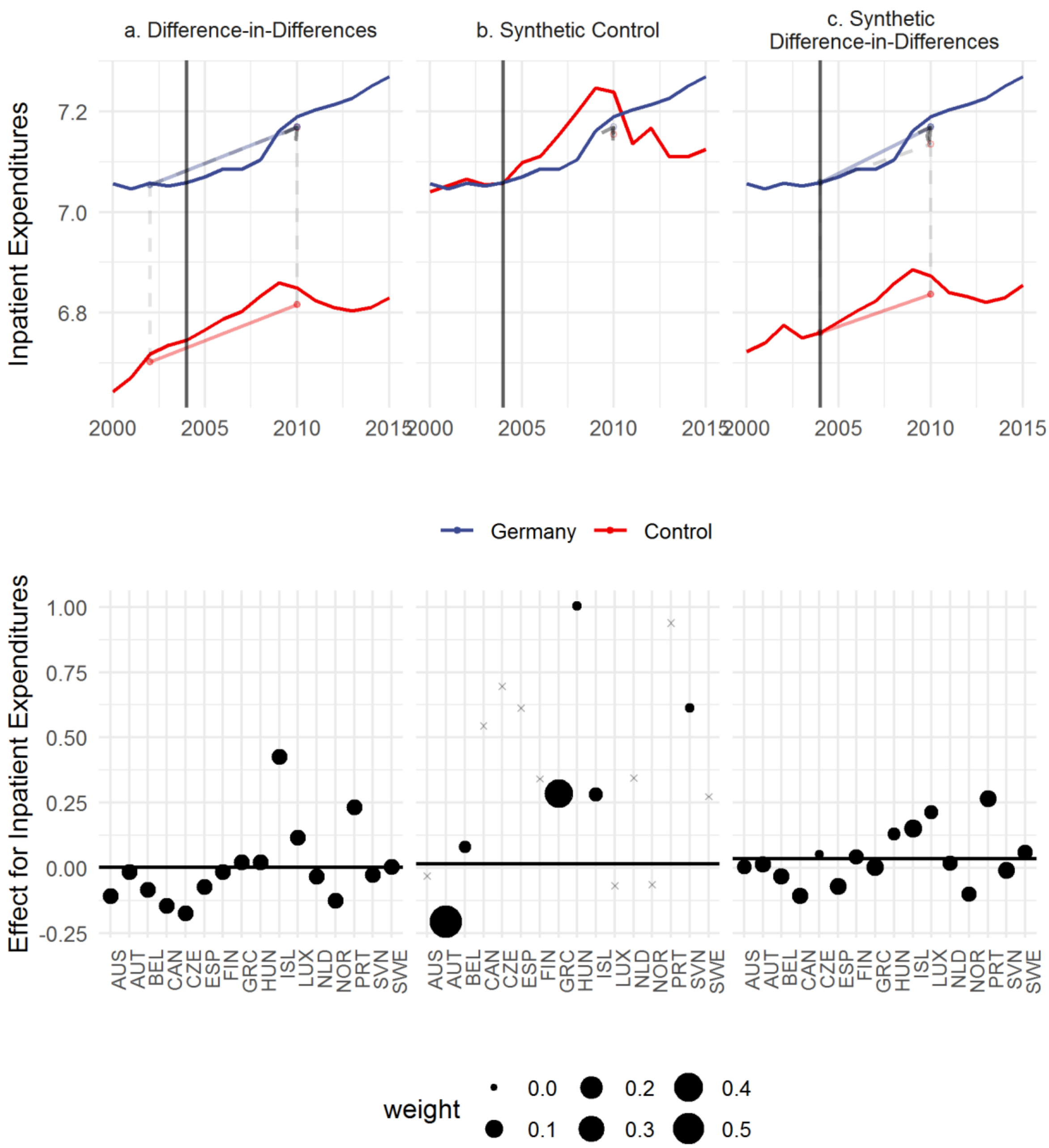


Figure A. 6: Estimated impact of the gDRG introduction on inpatient expenditures

Notes: All calculations without covariates. In the upper half, overall effects are shown. Below, country-by-country outcome differences with weights indicated by dot size are plotted. The black horizontal line indicates the overall estimate.

Graphical Illustration for Life Expectancy

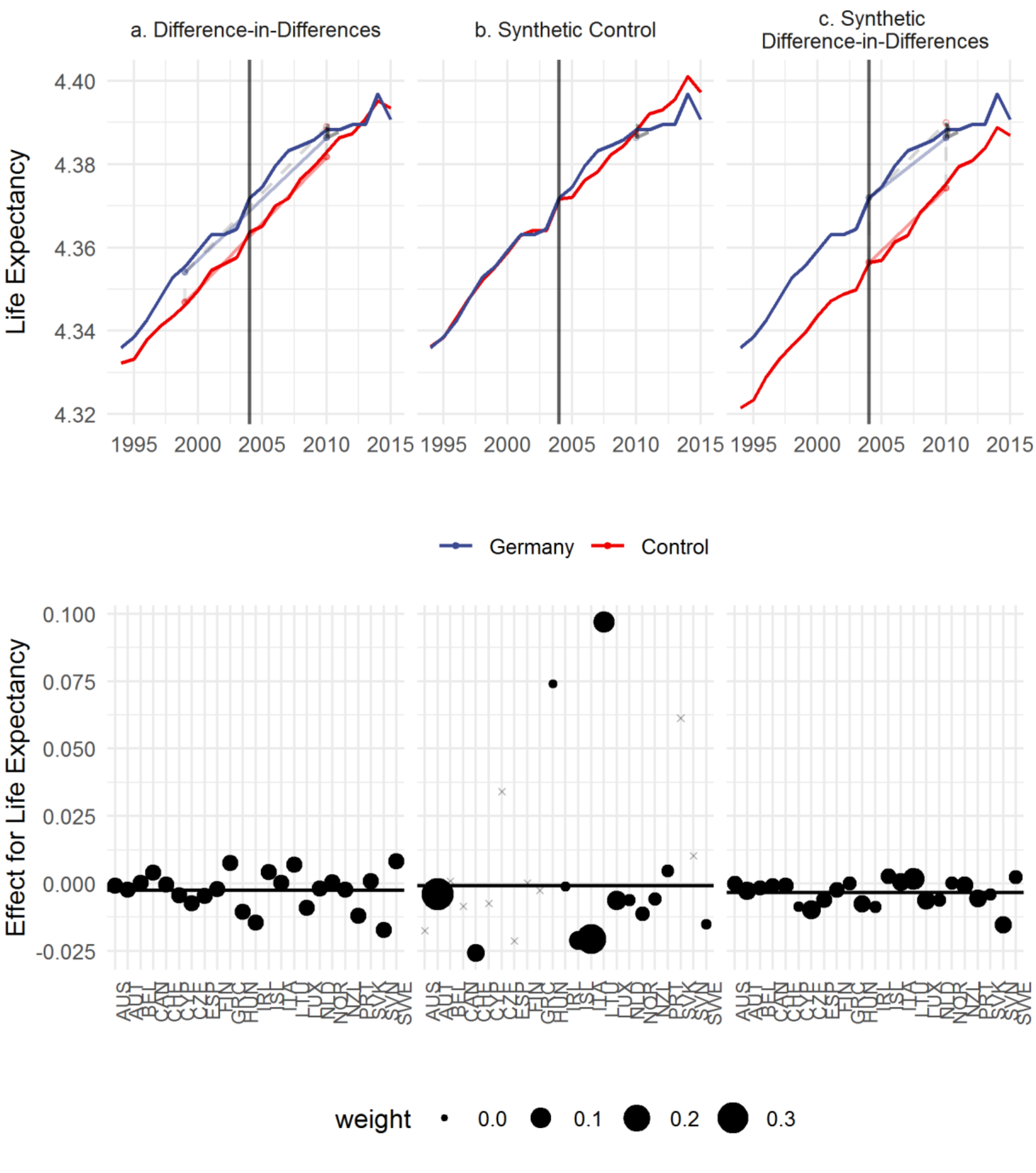


Figure A. 7: Estimated impact of the gDRG introduction on Life Expectancy

Notes: All calculations without covariates. In the upper half, overall effects are shown. Below, country-by-country outcome differences with weights indicated by dot size are plotted. The black horizontal line indicates the overall estimate.

Graphical Illustration for Stand. Death Rates

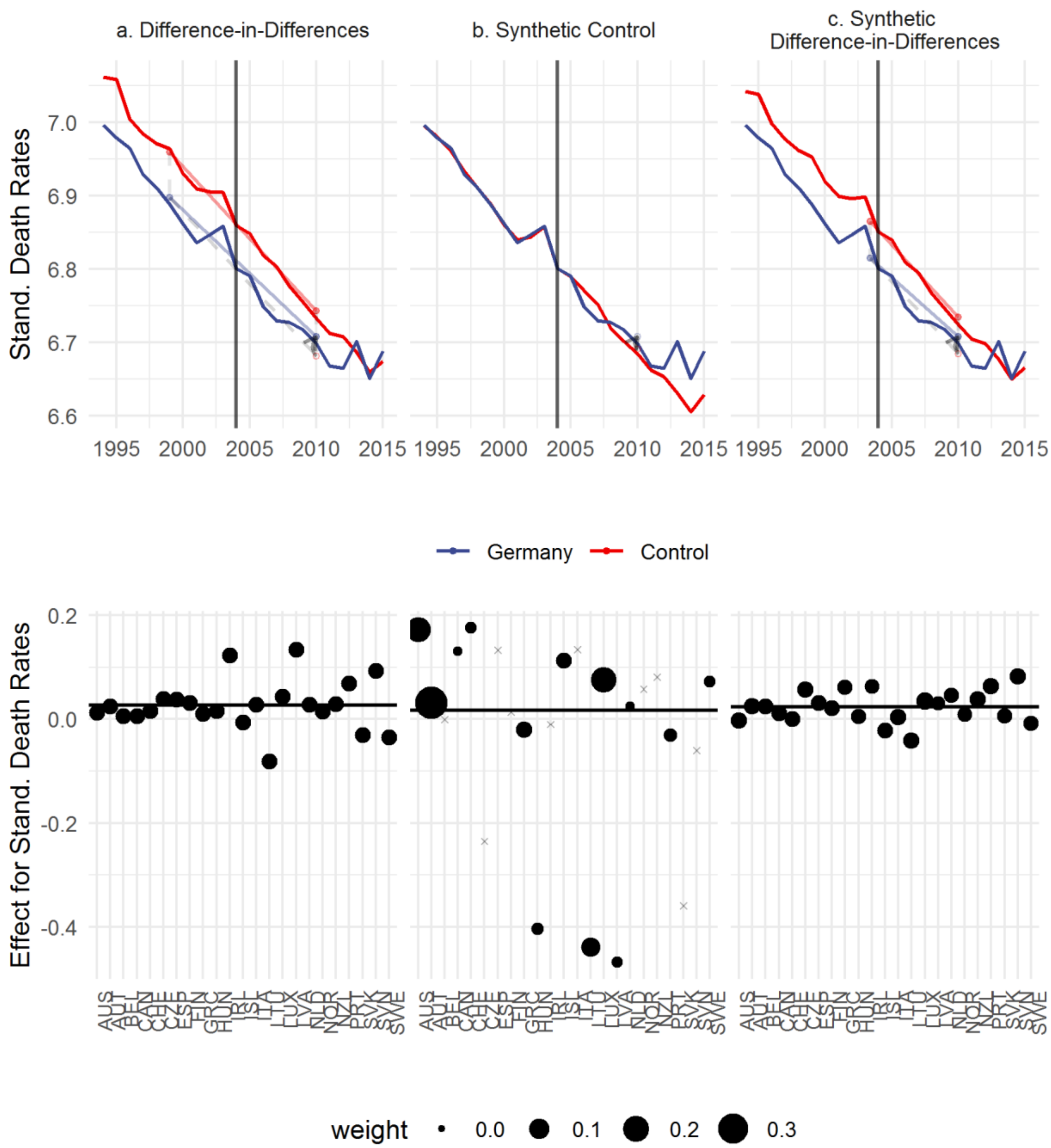


Figure A. 8: Estimated impact of the gDRG introduction on Standardized Death Rates

Notes: All calculations without covariates. In the upper half, overall effects are shown. Below, country-by-country outcome differences with weights indicated by dot size are plotted. The black horizontal line indicates the overall estimate.

Graphical Illustration for Potential Years of Life lost

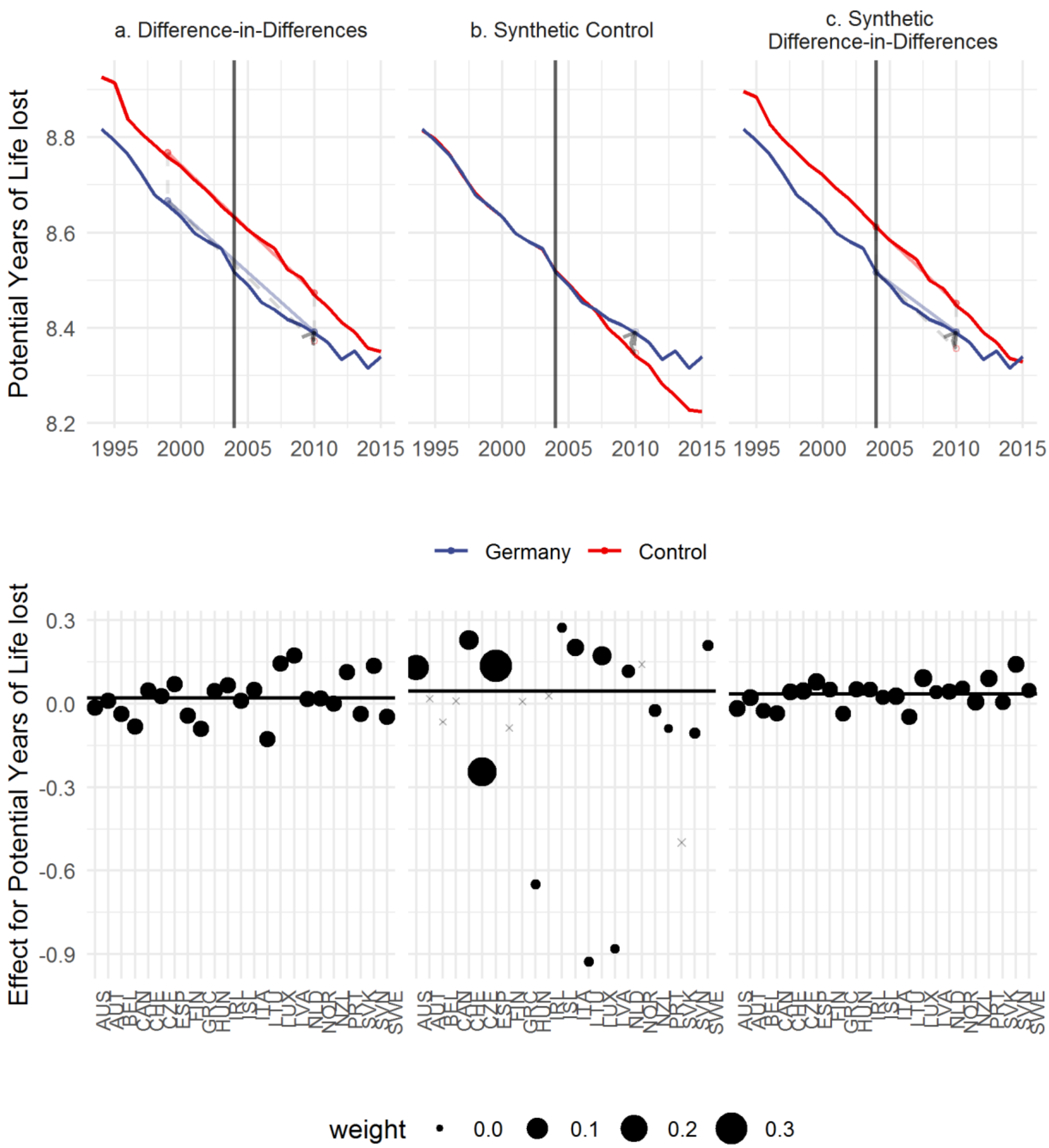


Figure A. 9: Estimated impact of the gDRG introduction on Potential Years of Life lost

Notes: All calculations without covariates. In the upper half, overall effects are shown. Below, country-by-country outcome differences with weights indicated by dot size are plotted. The black horizontal line indicates the overall estimate.

Appendix C – Robustness Checks

Event study plot – Parallel trends

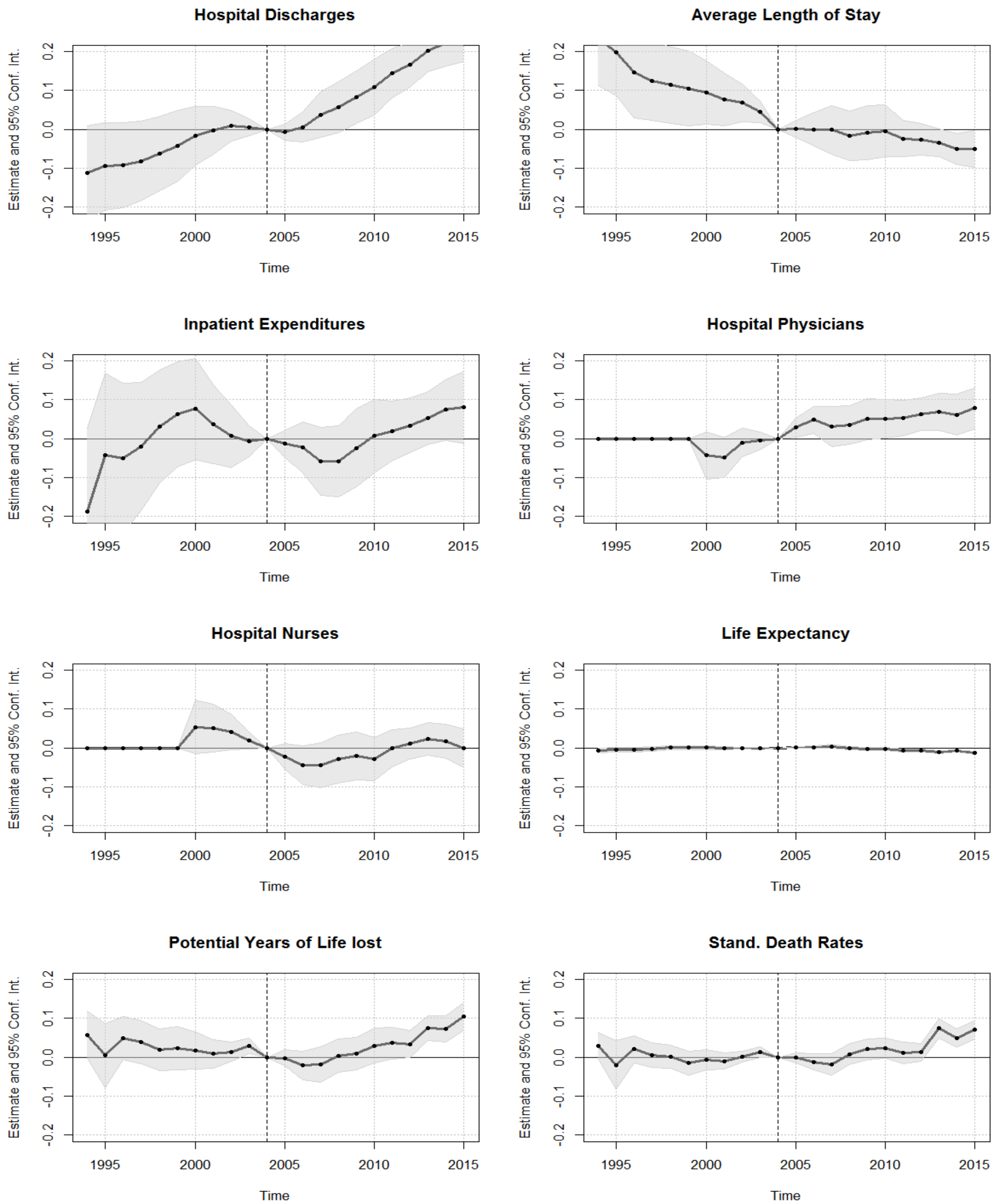


Figure A. 10: Event study plot for the DRG-introduction in Germany

DID results without countries with ABF classification

Table A.4: DID – Estimated impact of the gDRG introduction without countries with ABF classification

	(1)		(2)		(3)		(4)		(5)		(6)	
Outcome	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**
Hospital Discharges	0.0352*** (0.0065)	332	0.0276** (0.0092)	290	0.0302*** (0.0050)	140	0.0435*** (0.0070)	308	0.0459*** (0.0076)	261	0.0367** (0.0104)	238
Average Length of Stay												
Inpatient Expenditures	0.0155 (0.0096)	243	0.0043 (0.0080)	219	0.0078* (0.0030)	112	0.0241** (0.0079)	243	0.0217. (0.0106)	236	0.0101 (0.0061)	214
Hospital Physicians	0.0131** (0.0032)	185	0.0151*** (0.0032)	172	0.0147. (0.0065)	93	0.0154*** (0.0028)	175	0.0194*** (0.0035)	141	0.0144** (0.0037)	137
Hospital Nurses	-0.0018 (0.0058)	183	-0.0019 (0.0057)	170	0.0045 (0.0049)	82	0.0016 (0.0043)	173	-0.0038 (0.0034)	139	-0.0043 (0.0030)	135
Life Expectancy	-0.0014** (0.0004)	344	-0.0016** (0.0004)	292	-0.0015*** (0.0002)	138	-0.0012* (0.0004)	321	-0.0011* (0.0004)	269	-0.0014*** (0.0003)	242
Potential Years of Life lost	0.0121** (0.0040)	330	0.0109* (0.0042)	287	0.0078** (0.0023)	127	0.0086* (0.0038)	317	0.0063 (0.0041)	263	0.0086. (0.0041)	236
Stand. Death Rates	0.0096*** (0.0017)	347	0.0087*** (0.0019)	299	0.0068*** (0.0010)	139	0.0067** (0.0019)	323	0.0072** (0.0019)	269	0.0092*** (0.0016)	242
Controls for:												
Baseline	✓		✓		✓		✓		✓		✓	
Hospital Beds			✓		✓						✓	
Private Hospital Beds					✓						✓	
Healthcare Expenditure							✓				✓	
Outpatient Expenditures									✓		✓	

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Notes: All estimates include country and year fixed effects, with outcomes and control variables in log form. Robust standard errors clustered at country level in brackets. Results for Average Length of Stay are not provided because we do not assume the parallel trend assumption to hold. Parallel pre-trends might not hold for alle estimated models.

DID results with fixed budget only classification

Table A.5: DID – Estimated impact of the gDRG introduction with FB countries only

	(1)		(2)		(3)		(4)		(5)		(6)	
Outcome	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**
Hospital Discharges	0.0270 (0.0157)	138	0.0180** (0.0047)	101	-0.0027 (0.0109)	56	0.0317* (0.0105)	122	0.0539** (0.0139)	107	0.0202* (0.0055)	84
Average Length of Stay												
Inpatient Expenditures	0.0502** (0.0129)	109	-0.0006 (0.0172)	87	-0.0051. (0.0015)	42	0.0256* (0.0075)	109	0.0473* (0.0130)	106	0.0062 (0.0083)	84
Hospital Physicians	0.0104 (0.0072)	89	0.0204** (0.0039)	76	0.0094 (0.0070)	49	0.0187*** (0.0026)	79	0.0236** (0.0038)	64	0.0165* (0.0056)	60
Hospital Nurses	0.0032 (0.0108)	89	-0.0040 (0.0068)	76	0.0147 (0.0068)	49	0.0121*** (0.0017)	79	0.0060. (0.0030)	64	0.0022 (0.0054)	60
Life Expectancy	-0.0002 (0.0004)	154	-0.0007. (0.0003)	107	-0.0008 (0.0004)	56	-6.61e-5 (0.0004)	137	0.0004 (0.0006)	115	-0.0006. (0.0002)	88
Potential Years of Life lost	0.0006 (0.0040)	132	0.0053 (0.0066)	94	0.0168* (0.0020)	43	0.0011 (0.0042)	131	0.0008 (0.0072)	109	0.0086 (0.0073)	82
Stand. Death Rates	0.0023 (0.0015)	149	0.0047. (0.0019)	106	0.0159** (0.0025)	55	0.0023 (0.0021)	137	0.0008 (0.0028)	115	0.0061* (0.0017)	88
Controls for:												
Baseline	✓		✓		✓		✓		✓		✓	
Hospital Beds			✓		✓						✓	
Private Hospital Beds					✓						✓	
Healthcare Expenditure							✓				✓	
Outpatient Expenditures									✓		✓	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$.
Notes: All estimates include country and year fixed effects, with outcomes and control variables in log form. Robust standard errors clustered at country level in brackets. Results for Average Length of Stay are not provided because we do not assume the parallel trend assumption to hold. Parallel pre-trends might not hold for alle estimated models.

DID results with alternative classification

Table A.6: DID – Estimated impact of the gDRG introduction with alternative classification

	(1)		(2)		(3)		(4)		(5)		(6)	
Outcome	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**	**Est. ρ**	**Obs.**
Hospital Discharges	0.0320*** (0.0053)	486	0.0231** (0.0079)	434	0.0206** (0.0064)	256	0.0363*** (0.0062)	462	0.0408*** (0.0064)	388	0.0331** (0.0094)	363
Average Length of Stay												
Inpatient Expenditures	0.0067 (0.0051)	373	-9.73e-5 (0.0057)	347	0.0046 (0.0038)	219	0.0152* (0.0059)	373	0.0095 (0.0059)	363	0.0030 (0.0045)	339
Hospital Physicians	0.0096** (0.0031)	291	0.0072 (0.0042)	278	0.0031 (0.0047)	191	0.0119*** (0.0029)	281	0.0107* (0.0043)	248	0.0083. (0.0047)	244
Hospital Nurses	-0.0012 (0.0036)	289	-0.0017 (0.0027)	276	-0.0034 (0.0023)	180	0.0032 (0.0020)	279	0.0002 (0.0030)	246	9.64e-5 (0.0026)	242
Life Expectancy	-0.0009** (0.0003)	498	-0.0008* (0.0003)	436	-0.0010*** (0.0002)	254	-0.0009** (0.0003)	475	-0.0010*** (0.0003)	396	-0.0011*** (0.0002)	367
Potential Years of Life lost	0.0045 (0.0037)	484	-0.0001 (0.0041)	431	0.0048* (0.0020)	243	0.0042 (0.0034)	471	0.0037 (0.0029)	390	0.0044 (0.0027)	361
Stand. Death Rates	0.0061** (0.0018)	501	0.0035 (0.0026)	443	0.0056*** (0.0012)	255	0.0058*** (0.0012)	477	0.0063*** (0.0011)	396	0.0073*** (0.0009)	367
Controls for:												
Baseline	✓		✓		✓		✓		✓		✓	
Hospital Beds			✓		✓						✓	
Private Hospital Beds					✓						✓	
Healthcare Expenditure							✓				✓	
Outpatient Expenditures									✓		✓	

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Notes: All estimates include country and year fixed effects, with outcomes and control variables in log form. Robust standard errors clustered at country level in brackets. Results for Average Length of Stay are not provided because we do not assume the parallel trend assumption to hold. Parallel pre-trends might not hold for alle estimated models.

SC results without countries with ABF

Table A.7: SC - Estimated impact of the gDRG introduction without countries with ABF

	(1)		(2)	
Outcome	**ATT Estimate**	**approx. Annual Growth**	**ATT Estimate**	**approx. Annual Growth**
Single SC				
Hospital Discharges	0.2295*** (0.0000)	0.0422	0.4191* (0.0170)	0.0725
Average Length of Stay	0.0496 (1.0000)	0.0097	0.1507 (1.0000)	0.0285
Inpatient Expenditures	0.0796 (0.9950)	0.0154	0.0545 (0.9740)	0.0107
Hospital Physicians	0.1033 (0.2510)	0.0198	0.2208 (0.2140)	0.0407
Hospital Nurses	-0.0187 (0.1220)	-0.0038	-0.1556 (0.1030)	-0.0333
Life Expectancy	-0.0064 (0.5580)	-0.0013	0.0003 (0.9180)	0.0001
Potential Years of Life lost	0.0377* (0.0390)	0.0074	-0.0567* (0.0470)	-0.0116
Stand. Death Rates	0.0348 (0.1900)	0.0069	0.0007 (0.7390)	0.0001
Multi SC				
Hospital Discharges	0.1810 (0.5342)	0.0338	0.1691*** (0.0000)	0.0317
Average Length of Stay	0.1709*** (0.0000)	0.0321	0.1646*** (0.0000)	0.0309
Variables used in weight determination:				
Baseline	✓		✓	
Hospital Beds			✓	
Healthcare Expenditure			✓	
Outpatient Expenditures			✓	

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Single SC: Outcome specific control group weights
Multi SC: One set of control group weights across all outcomes.
Notes: All estimates with outcomes and control variables in log form. P-values from the conformal inference procedure from Chernozhukov et al. (2021) are shown in brackets. An approximate interpretation of the ATT as (linear) annual growth was added for comparison with DID results.

SC results with fixed budget only classification

Table A.8: SC - Estimated impact of the gDRG introduction with Fixed Budget

	(1)	
Outcome	**ATT Estimate**	**approx. Annual Growth**
Single SC		
Hospital Discharges	0.4956*** (0.0000)	0.0838
Average Length of Stay	0.2628 (1.0000)	0.0478
Inpatient Expenditures	0.0825 (0.8700)	0.0160
Hospital Physicians	0.1389* (0.0390)	0.0264
Hospital Nurses	-0.1054 (0.8890)	-0.0220
Life Expectancy	-0.0014 (0.5810)	-0.0003
Potential Years of Life lost	0.0196 (0.1340)	0.0039
Stand. Death Rates	0.0179 (0.6660)	0.0036
Multi SC		
Hospital Discharges	0.5422*** (0.0000)	0.0905
Average Length of Stay	0.3123*** (0.0000)	0.0559
Variables used in weight determination:		
Baseline	✓	

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Single SC: Outcome specific control group weights
Multi SC: One set of control group weights across all outcomes.
Notes: All estimates with outcomes and control variables in log form. P-values from the conformal inference procedure from Chernozhukov et al. (2021) are shown in brackets. An approximate interpretation of the ATT as (linear) annual growth was added for comparison with DID results.

SC results with alternative classification

Table A.9: SC - Estimated impact of the gDRG introduction with alternative classification

	(1)		(2)	
Outcome	**ATT Estimate**	**approx. Annual Growth**	**ATT Estimate**	**approx. Annual Growth**
Single SC				
Hospital Discharges	0.1321** (0.0080)	0.0251	0.1263** (0.0070)	0.0241
Average Length of Stay	-0.0102 (0.9370)	-0.0020	-0.0083 (1.0000)	-0.0017
Inpatient Expenditures	0.0690 (0.8580)	0.0134	-0.0356 (0.9380)	-0.0072
Hospital Physicians	0.1024 (0.5540)	0.0197	0.1190 (0.8840)	0.0227
Hospital Nurses	-0.0333 (0.3570)	-0.0068	-0.1061 (0.4140)	-0.0222
Life Expectancy	-0.0012* (0.0100)	-0.0002	-0.0017. (0.0580)	-0.0003
Potential Years of Life lost	0.0263* (0.0120)	0.0052	-0.0054** (0.0080)	-0.0011
Stand. Death Rates	0.0346 (0.1230)	0.0068	0.0192 (0.1610)	0.0038
Multi SC				
Hospital Discharges	0.1020* (0.0417)	0.0196	0.0421 (0.3742)	0.0083
Average Length of Stay	-0.0368 (0.7460)	-0.0075	-0.0005 (0.9973)	-0.0001
Variables used in weight determination:				
Baseline	✓		✓	
Hospital Beds			✓	
Healthcare Expenditure			✓	
Outpatient Expenditures			✓	

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Single SC: Outcome specific control group weights
Multi SC: One set of control group weights across all outcomes.
Notes: All estimates with outcomes and control variables in log form. P-values from the conformal inference procedure from Chernozhukov et al. (2021) are shown in brackets. An approximate interpretation of the ATT as (linear) annual growth was added for comparison with DID results.

SDID results without countries with ABF classification

Table A.10: SDID - Estimated impact of the gDRG introduction without countries with ABF classification

Outcome	ATT Estimate	approx. Annual Growth
Hospital Discharges	0.1030. (0.0718)	0.0198
Average Length of Stay	-0.1032 (0.0891)	-0.0215
Inpatient Expenditures	0.0424 (0.2173)	0.0083
Hospital Physicians	0.0814 (0.0660)	0.0158
Hospital Nurses	-0.0359 (0.1447)	-0.0073
Life Expectancy	-0.0036 (0.0050)	-0.0007
Potential Years of Life lost	0.0456 (0.0705)	0.0089
Stand. Death Rates	0.0252 (0.0390)	0.0050

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.

Notes: All estimates with outcomes and control variables in log form. GDP per capita and share of population 65+ used as control variables. Standard errors from placebo-evaluations in brackets. An approximate interpretation of the ATT as annual growth was added for comparison with DID results.

SDID results with fixed budget only classification

Table A.11: SDID - Estimated impact of the gDRG introduction with Fixed Budget

Outcome	**ATT Estimate**	**approx. Annual Growth**
Hospital Discharges	0.1098 (0.1848)	0.0211
Average Length of Stay	-0.1075* (0.0567)	-0.0225
Inpatient Expenditures	0.0550 (0.3307)	0.0108
Hospital Physicians	0.0454 (0.1124)	0.0089
Hospital Nurses	0.0048 (0.1701)	0.0009
Life Expectancy	-0.0059. (0.0043)	-0.0012
Potential Years of Life lost	0.0574 (0.0585)	0.0112
Stand. Death Rates	0.0247 (0.0199)	0.0049

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$.

Notes: All estimates with outcomes and control variables in log form. GDP per capita and share of population 65+ used as control variables. Standard errors from placebo-evaluations in brackets. An approximate interpretation of the ATT as annual growth was added for comparison with DID results.

SDID results with alternative classification

Table A.12: SDID - Estimated impact of the gDRG introduction with alternative classification

Outcome	ATT Estimate	approx. Annual Growth
Hospital Discharges	0.1068 (0.1013)	0.0205
Average Length of Stay	-0.0368 (0.0924)	-0.0075
Inpatient Expenditures	0.0137 (0.2035)	0.0027
Hospital Physicians	0.0470 (0.0731)	0.0092
Hospital Nurses	-0.0286 (0.1089)	-0.0058
Life Expectancy	-0.0031 (0.0042)	-0.0006
Potential Years of Life lost	0.0338 (0.0548)	0.0067
Stand. Death Rates	0.0251 (0.0468)	0.0050

*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.
Notes: All estimates with outcomes and control variables in log form. GDP per capita and share of population 65+ used as control variables. Standard errors from placebo-evaluations in brackets. An approximate interpretation of the ATT as annual growth was added for comparison with DID results.

Deviation from parallel trends

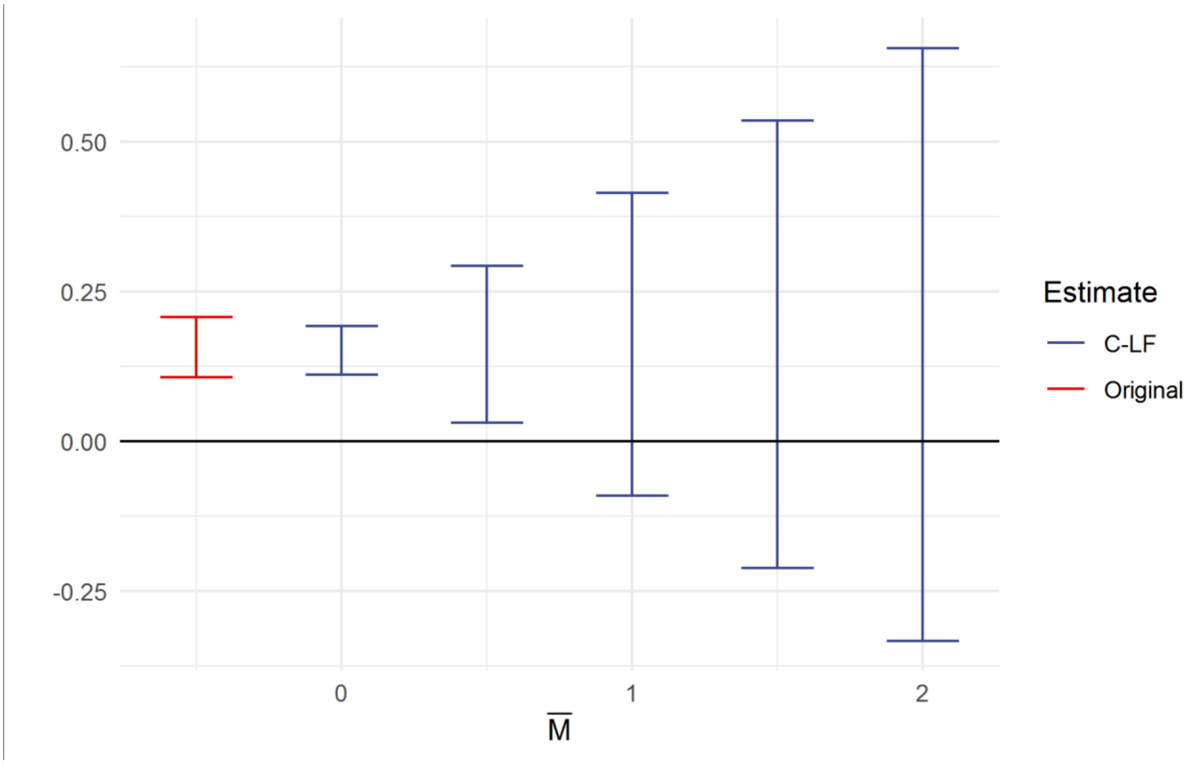


Figure A. 11: Sensitivity analysis for parallel trends - hospital discharges.

Notes: Estimate for t = 2010. See Rambachan and Roth (2021) for more information regarding setup and interpretation.

Deviation from linear trends

Table A.13: Sensitivity for deviation from linear trend

lb	ub	method	Delta	Mbar
0.1515	0.2525	C-LF	DeltaSDRM	0.0000
-0.0909	0.5152	C-LF	DeltaSDRM	0.5000
-0.3737	0.7980	C-LF	DeltaSDRM	1.0000
-0.6364	1.0000	C-LF	DeltaSDRM	1.5000
-0.9394	1.0000	C-LF	DeltaSDRM	2.0000

Notes: Estimate for t = 2010. See Rambachan and Roth (2021) for more information regarding setup and interpretation.

.

Placebo-in-space test for Hospital Discharges

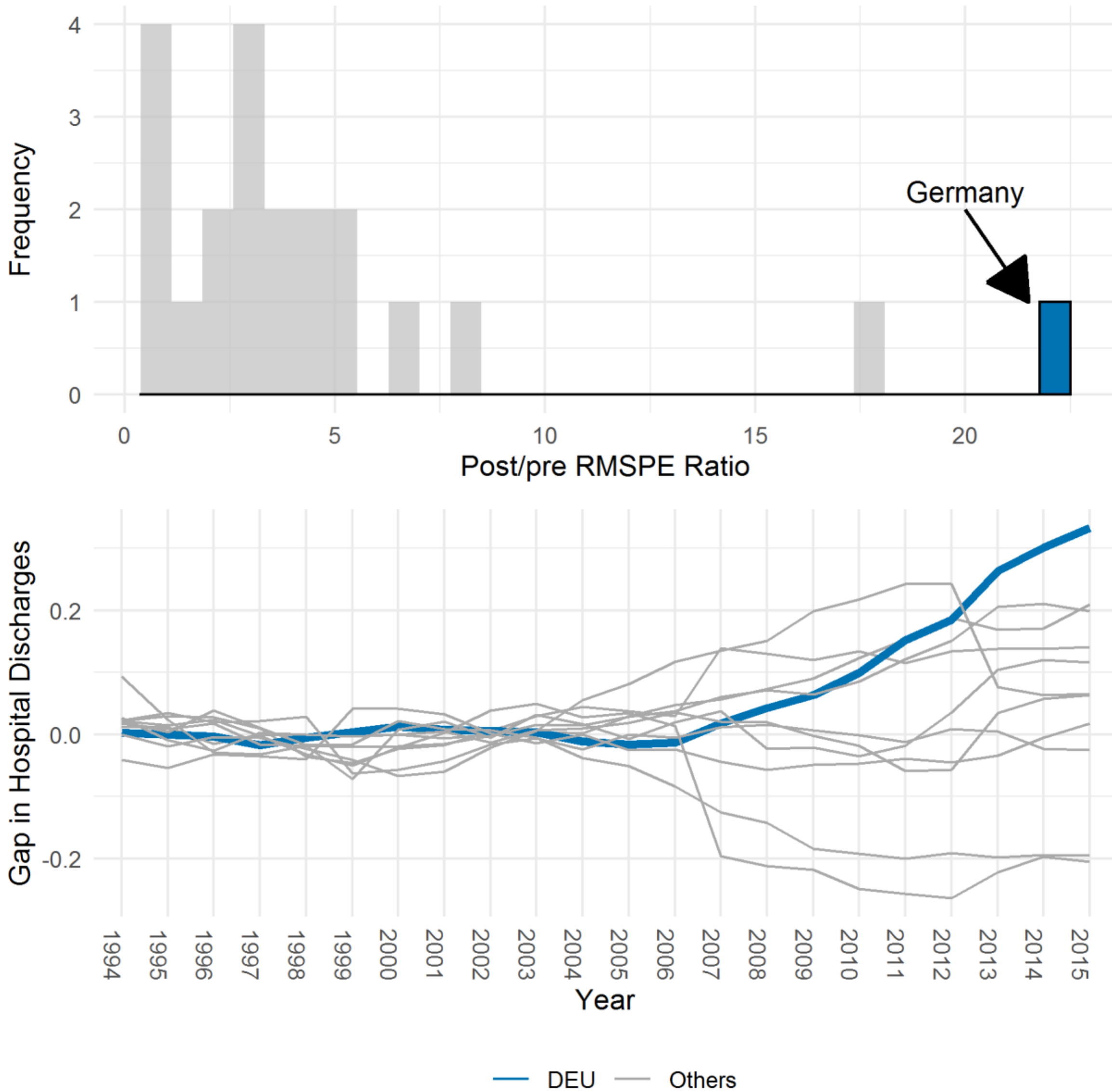


Figure A. 12: Placebo-in-space test for Hospital Discharges

Notes: Only countries with less than five times Germany's pre-treatment fit (RMSPE) are plotted below.

Placebo-in-time for Hospital Discharges

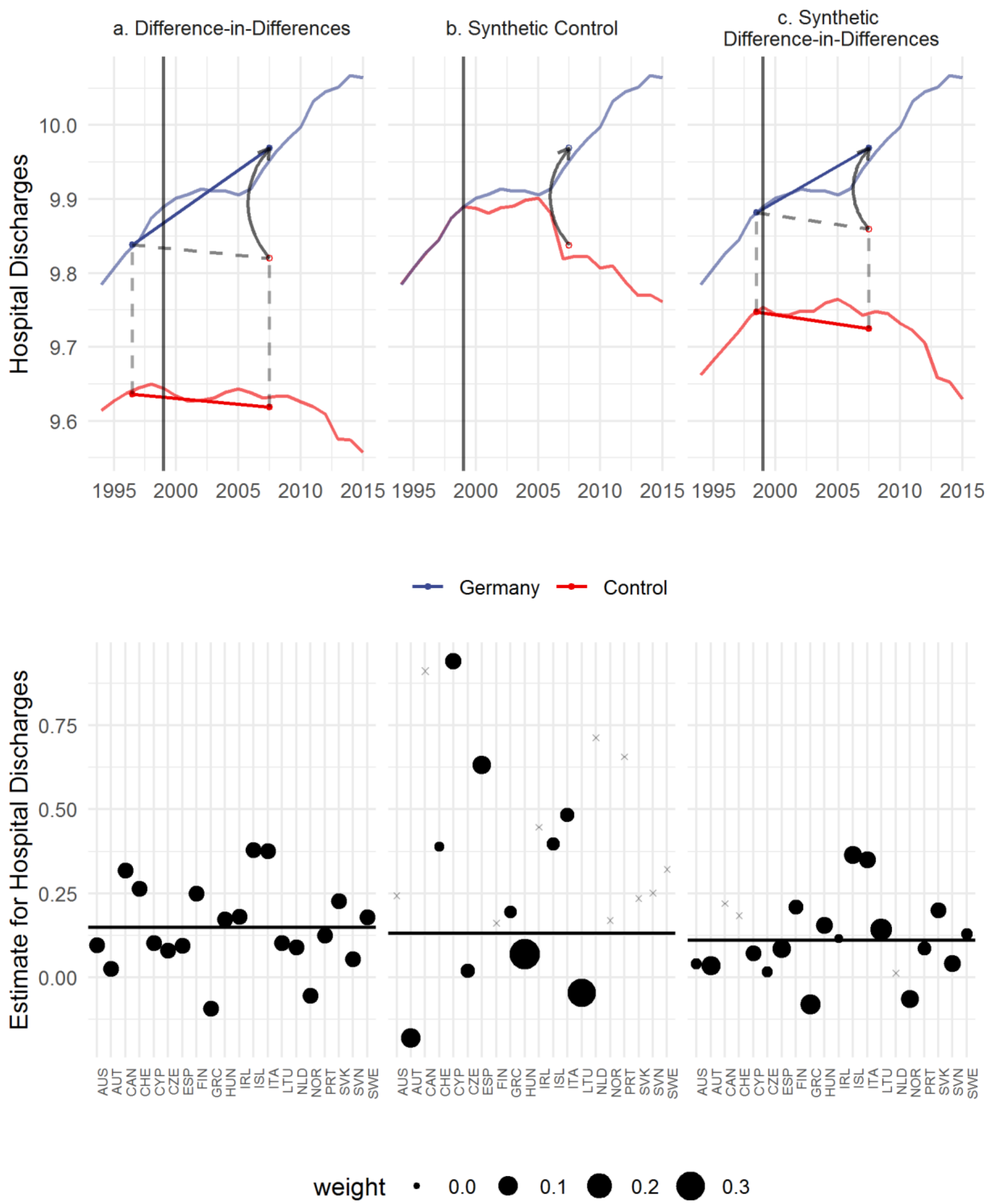


Figure A. 13: Placebo-in-time test for hospital discharges. Treatment assigned to t=2000.

Supporting Information

Hospital payment scheme classification

In order to investigate the effects of the gDRG reform, we constructed - in analogy to Moreno-Serra and Wagstaff (2010) and Wubulihasimu et al. (2016) - a data set describing the predominant hospital payment schemes for selected OECD and EU member states from 1994 to 2015. All European OECD countries, Australia, Canada and New Zealand as well as EU member states at the timing of the reform (2004) are included. Thereby our selection covers developed countries with in general comparable levels of healthcare to Germany. This appendix provides more details towards the classification and the underlying sources.

General classification

In a first step to classify hospital payment schemes, we created two basic categories with opposite incentive structures: Fixed budgets (FB) or activity based funding (ABF). A country is classified as FB in a given year if budgets or block grants are the predominant form of hospital funding. In this case, hospital revenues are mainly pre-determined depending on provider characteristics like hospital size or range of care. In contrast, a country was classified as activity based funding if hospitals are paid according to the characteristics of admitted patients or activities. This includes per case and DRG-based methods as well as fee-for-service.

A major obstacle is that in many countries hospital reimbursement schemes have been introduced only gradually or only partially. Thereby ABF often only affects a fraction of hospital budgets (e.g. Denmark), is limited to certain hospitals and regions (e.g. Finland and Sweden) or is used for budgeting but not for actual billing processes (e.g. Ireland). However, the incentive structure in payment systems with strict adherence to ABF differs from the structure in systems, where ABF is one among many payment methods. Within mixed systems, no clear incentive structure might be in place at all. This can be even intentional since countries can balance contradicting incentives this way.

To incorporate this line of thought into our classification, we introduced a mixed funding category. We narrowed the definition of ABF to systems where ABF is the predominant way of billing hospital care. This mode of payment must apply to the majority of acute care hospitals nationwide. Systems, which use ABF only partially or mainly for budgeting purposes, were classified as mixed funding. In contrast, a global healthcare or hospital care budget does not necessarily lead to a classification as mixed funding because the incentive structure at the hospital level should not be significantly affected. In e.g. Hungary DRGs are used for billing purposes but the cost weight is not fixed because of a global budget. But since an individual hospital has no influence over the activity of other hospitals, the theoretic incentives of increased activity and efficiency are still in place (Kroneman & Nagy, 2001).

Main source of information for the classification is the World Health Organization's Health System in Transition series (HiT) and the OECD Health Systems Characteristics Survey (Busse et al., 2011; OECD, 2016; Paris et al., 2010; WHO, 2021). We primarily screened the "Payment mechanisms" section of the HiT series and used additional sources where the HiT lacked information. Country specific literature is referenced in Table A. 14.

Our assignment is – as the classification as a whole – not straightforward. This payment scheme classification is at least subject to certain interpretations. Most countries use several methods in parallel. The proportion of the individual schemes is often unclear. Moreover, when payment scheme reforms are gradually implemented, there is ambiguity regarding the starting time of the reform. We therefore included alternative classifications for some countries for robustness checks.

Table A. 14: Hospital payment scheme classification

Country	Remarks and Additional Data Sources	Main Classification	In CG	Alternative classification	In CG
Australia	Hospitals were mainly financed through prospective global budgets. In 1993, Victoria and South Australia started adopting case-mix DRG systems for budget allocation. Nationwide implementation started in 2012. Since 2014 it is used for billing, before mainly for budgeting purposes. (Hilless & Healy, 2001; Parliamentary Library, 2016)	mix 1994 - 2013 ABF 2014 - 2015	x		
Austria	Hospitals were mainly financed through prospective global budgets. In 1997, a DRG-based system called 'Leistungsorientierte Krankenhausfinanzierung' (LKF) was introduced. Austria uses this DRG like system nationwide (for >70% of revenues), regional adjustments are possible (on average 30%) and the shares vary widely between the states from approx. 15-50 % leading substantial to price variation. (Theurl, 2015)	FB 1994-1996 ABF 1997-2015	x	FB 1994-1996 Mixed 1997-2015	x
Belgium	Hospitals are financed by complex budget allocations consisting of several items. Prior to 2001 this budget distribution was partly based on cost comparison. DRGs were introduced in 1994, but are used for some aspects of budget allocation only since 2002. There are no direct DRG Payments. (Gerkens & Merkur, 2010; Stephani et al., 2018)	FB 1994-2001 Mixed 2002-2015	x	FB 1994-2015	x
Canada	Payments are generally are based on the previous year's allocation adjusted for inflation and budget growth. 2010 ABF was partially introduced in British Columbia. The ABF program re-directed up to 20% of	FB 1994-2015	x	FB 1994-2009 Mixed 2010-2015	x

	global budget funding to a case mix adjusted per case payment, replacing a portion of global budget's funding. (Marchildon et al., 2020; Sutherland et al., 2016; Sutherland & Repin, 2014)				
Cyprus	Hospitals are financed through historic global budgets. FFS only in private sector; DRGs introduced in 2019. (Theodorou et al., 2012)	FB 1994-2015	x		
Czech Republic	From 1993 to mid-1997 a FFS system was partially used for hospital reimbursement. Afterwards a prospective global budget was the main source of financing for most years. The share of hospital care paid through DRG based case payments gradually increased from its introduction in 2007. In 2012 it was the main reimbursement mechanism for hospitals. In 2015, the Czech Republic stopped financing hospital care based on DRGs completely and returned temporarily to flat fees. (Bryndová et al., 2009; Kotherová et al., 2021)	Mixed 1994-1996 FB 1997-2011, 2015 ABF 2012-2014	x	Mixed 1994-2011 FB 2015 ABF 2012-2014	x
Denmark	In Denmark, hospital budgets are partially determined based on ABF, while they also depend on annually negotiated global budgets. In 1999 after the introduction of DRGs 10% of budget was allocated by ABF, in 2008 the mandatory share was set to 50%. Since 2015, value based procurement becomes more important. (Christiansen & Vrangbæk, 2018; Street, A., Vitikainen, K., Bjorvatn, A. and Hvenegaard, A., 2007) Not in control group for comprehensive hospital reform.	FB 1994-1999 Mixed 2000-2015		FB 1994-1999 Mixed 2000-2007 ABF 2008-2015	
Estonia	Fee-for-service and per diem system within a hard global budget was used until 2003. In 2003, the DRG system was introduced for case grouping. From 2004 on the NordDRG system became operational in combination with other payment methods. The proportion of DRG payment was gradually raised from 10% in 2004 to 70% in 2009. (Bredenkamp et al., 2020; Estonian Health Insurance Fund, 2009; Lai et al., 2013; Mathauer & Wittenbecher, 2013)	Mix 1994-2007 ABF 2008-2015			
Finland	Region specific budgets are used to finance hospitals. Often NordDRG based methods are used for invoicing to municipalities. (Keskimäki et al., 2019; Mikkola, 2003; Stig & Paulsson Lütz, 2013)	Mixed 1994-2015	x		

France	Global budgets were mainly used for hospital financing before 2004, at least for public hospitals. DRGs were introduced in 2004/2005. In 2010 56% of total hospital expenditures were covered by the DRG-based payment. (Chevreul et al., 2015; Stephani et al., 2018; van de Voorde et al., 2013)	FB 1994-2003 ABF 2004-2015		FB 1994-2003 Mixed 2004-2015	x
Germany	From 1993 on reimbursement with prospective fixed budgets (before full cost compensation). In 2004 change to DRG system. (Blümel et al., 2020)	FB 1994-2003 ABF 2004-2015			
Greece	Until 2010, hospitals were payed within a retrospective per diem system under a global budget. Government compensated deficits periodically. German DRG system was implemented from 2011 on. (Economou et al., 2017; Polyzos et al., 2013)	Mixed 1994-2010 ABF 2011-2015	x		
Hungary	DRGs were introduced as early as 1993. However, DRGs are not used as fixed payment units. Instead prices of DRGs vary with budget utilization. (Endrei et al., 2014; Gaál et al., 2011; Kroneman & Nagy, 2001)	ABF 1994-2015	x	Mixed 1994-2015	x
Iceland	Since the 1980s, Iceland uses mainly a fixed budget. The NordDRG System has been in use in Iceland since 2001 but DRGs are mainly used for hospital internal distribution of the budget to departments. (Stig & Paulsson Lütz, 2013)	FB 1994-2015	x		
Ireland	Ireland predominately uses historic budgets for hospital financing. In 2001, a full phased implementation of ABF for financing was announced. However, in 2021 budget allocation was still only partially based on ABF for some hospitals. (Health Service Executive, 2021; McDaid et al., 2009; McElroy & Murphy, 2014)	FB 1994-2015	x	FB 1994-2000 ABF 2001-2015	
Italy	Hospitals were predominately reimbursed by fixed budgets. DRGs were introduced in 1995, but the degree of use varies strongly in regions. Moreover, budgets for specific care services (e.g. emergency services, oncology treatments, hospital-teaching activities) are not based on DRG tariffs but are paid on the basis of global budgets. Focus on ABF has faded recently. (Anessi-Pessina et al., 2019; Cantù et al., 2011; Ferré et al., 2014)	FB 1994 Mixed 1995-2015	x	FB 1994 ABF 1995-2015	

Latvia	Until 2010 Latvia used a mix of case-based payments and per diem payment, afterwards fixed budgets were implemented. DRGs were implemented starting 2014, but the role is limited. (Behmane et al., 2019; Dubas-Jakóbczyk et al., 2020)	Mixed 1994-2009, 2015 FB 2010-2014	x		
Lithuania	Hospitals were paid for admitted patients according to major specialty. DRG-based payments were introduced in 2012. (Murauskiene et al., 2013)	Mixed 1995-2011 ABF 2012-2015	x		
Luxembourg	Hospital are funded by global budgets. (Berthet et al., 2015)	FB 1994-2015	x		
New Zealand	Public hospitals receive an annual budget. (Cumming et al., 2014)	FB 1994-2015	x		
Netherlands	Hospitals were predominately reimbursed by fixed budgets. In 2001 change to FFS and in 2005 first adoption of DRG-like system (DBCs); by 2008 about 20% of hospital costs were reimbursed by DBCs payment. 2012 fundamental changes and stronger emphasis on DBCs. (Krabbe-Alkemade et al., 2017; Kroneman et al., 2016; Schut & van de Ven, 2005)	FB 1994-2000 Mixed 2001-2011 ABF 2012-2015	x	FB 1994-2000 Mixed 2001-2004 ABF 2005-2015	
Norway	Hospital financing changed in 1997 from fixed budgets to a mixed financing system with budgets and case-based payments. (Sperre Saunes et al., 2020; Stig & Paulsson Lütz, 2013).	FB 1994-1996 Mixed 1997-2015	x		
Poland	Until 1999 historical budgets were used for hospital funding, afterwards hospital payment was based on an activity based catalog. DRGs were implemented as nationwide hospital financing system in 2009. However, for some parts of hospital care they are only used to calculate biannual lump sum payments. (Czach et al., 2011; Moreno-Serra & Wagstaff, 2010; Sowada et al., 2019)	FB 1994-1999 Mixed 2000-2008 ABF 2009-2015		FB 1994-1999 Mixed 2000-2015	x
Portugal	The NHS funds inpatient care through global budgets. Between 1997 and 2002 funding through DRGs gradually increased. Since 2003 it represents around 75–85% of NHS hospitals' inpatient budget. DRGs are used for budget allocation, not payments. (Simões et al., 2017)	FB 1994-1996 Mixed 1997-2015	x	FB 1994-1996 ABF 1997-2015	
Slovak Republic	From 1993 to 2000, several distinct payment mechanism followed in short order. Since 2001, hospitals are funded through several channels, case payments grouped by departments and hospitals,	Mixed 1994-1999 FB 2000-2001 Mixed 2002-2015	x		

	Fee-for-service and per diem payments. DRG implementation started 2016. (Smatana et al., 2016)				
Slovenia	Hospital payment is based on provider budgets. Since 2003, DRG case payments were gradually introduced into budget negotiations. (Albreht et al., 2021)	FB 1994-2002 Mixed 2003-2015	x	FB 1994-2002 ABF 2003-2015	
Spain	Hospitals are primarily financed by global budgets. Regional (especially Catalonia) use of DRG for budget allocation. DRGs are also used for regional balance payments. (Bernal-Delgado et al., 2018)	FB 1994-2015	x		
Sweden	Before 2002, Sweden used fixed budgets as main hospital payment scheme. In 2002, hospitals introduced case payments based on DRGs as partial aspect of payment. Budget allocation methods differ between regions and global budgets still have a major role. (Anell et al., 2012)	FB 1994-2001 Mixed 2002-2015	x	FB 1994-2001 ABF 2002-2015	
Switzerland	Before reform in 2012 hospitals received per diem payments combined with fixed budgets. In general, Swiss hospital funding is subject to large regional variations. Nationwide Swiss-DRG started in 2012, before there were regional DRG implementations starting in 2003 (mainly related to budget only). (Busato & Below, 2010; Pietro et al., 2015)	Mixed 1994-2004 ABF 2012-2015	x		
United Kingdom (England)	Before 2003, hospitals were mainly paid by annual fixed budgets. A DRG system was introduced in 2003 and became operational in 2005. (Cylus et al., 2015; van de Voorde et al., 2013)	FB 1994-2004 ABF 2005-2015			

CG = Control Group

Data collection process

We use unbalanced country-level panel data from OECD sources, complemented by data from Eurostat (Eurostat, 2021; OECD, 2021) and for some economic indicators from the World Bank (World Bank, 2021).

Our main outcomes of interest concern the hospital activity and efficiency. Due to the limited nature of aggregate data, we use hospital discharges and the average length of stay as outcome variables. Secondary outcome variables concern hospital resources and expenditures as well as population health status. We use inpatient expenditures but the data availability and quality is poor for the considered period. We further use the number of physicians and nurses employed by hospital for some additional analysis since it is an intensive point of discussion in Germany. However, the data quality and availability is rather poor as well. As indicators for the overall population health status we use standardized all cause death rates, life expectancy and potential years of life lost (PYLL).

For some of our variables multiple closely related definitions exist. They differ e.g. with respect to the types of considered hospitals (curative vs. rehabilitative) or in the unit of measurement (full time equivalents vs. head counts). We chose a main definition for each variable. When this definition was not available for a country or when it covered only a small period, we complemented data using closely related definitions with richer data. E.g. instead of curative care discharges we used inpatient discharges for some countries. See Table A. 15 for an overview of the definitions and the countries where additional data definitions were used. Data for Cyprus was added from Eurostat.

Due to our described data collection process, data definitions can slightly differ between countries but they do not differ over time within a country. This approach is suitable for two reasons. First, country level data regularly contain country-specific deviations from the main definition. Thus, the additional variation induced by using closely related data definitions seems acceptable. Second, since the focus of our analyses are differences over time, varying data definitions – across countries but not over time – should not bias the results.

Table A. 15: Data definitions

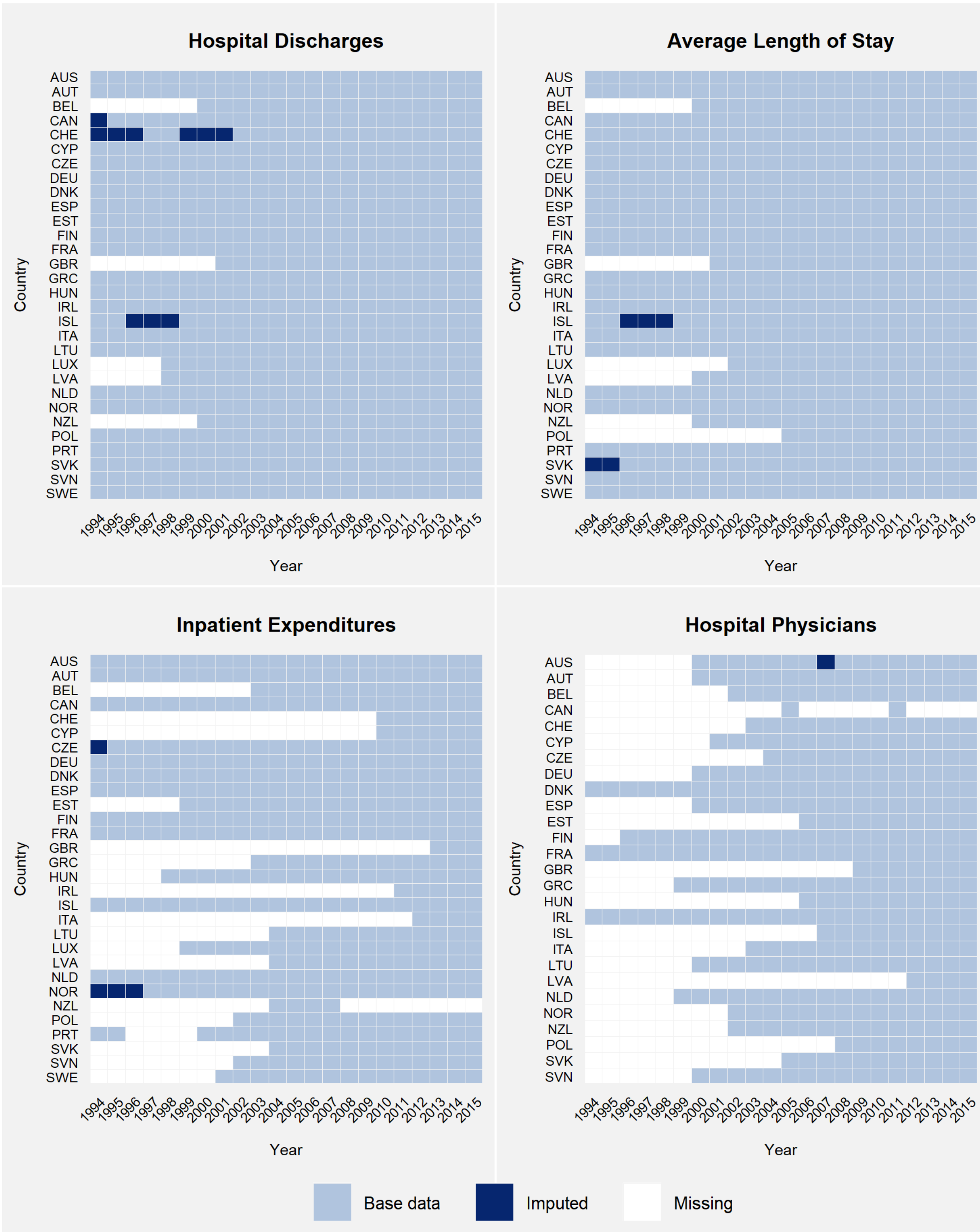
Variable	Data definition
Hospital discharges per 100 000 inhabitants	Main: • Curative (acute) care discharges *(HEALTH_PROC)* Additional: • Inpatient care discharges (All hospitals) (CZE, DNK, LUX, POL) *(HEALTH_PROC)* • All causes discharges - Hospital discharges by diagnostic categories (AUS, NLD) *(HEALTH_PROC)* • Hospital discharges for curative care (Eurostat)
Average length of stay	Main: • Curative care average length of stay *(HEALTH_PROC)* Additional: • Inpatient care average length of stay (All hospitals) - (AUS, CZE, DNK, ISL, NOR) *(HEALTH_PROC)* • Average length of stay for inpatient care (Eurostat)
Share of population 65+	Main: • Population 65 years old and over as percent of total population *(HEALTH_DEMR)* Additional: • Proportion of population aged 65 years and more (Eurostat)
GDP per capita	Main: • GDP per capita, PPP (constant 2011 international $) (World Bank) Additional: • None
Inpatient Expenditures	Main: • Inpatient curative and rehabilitative care expenditures (HC11HC21) for all healthcare providers (HPTOT) per capita, constant prices, constant PPPs, OECD base year *(SHA)* Additional: • Inpatient curative and rehabilitative care expenditures (HC11_21) for all healthcare providers (TOTAL) PPS per inhabitant (Eurostat)
Outpatient Expenditures	Main: • Outpatient curative and rehabilitative care expenditures (HC13HC23) for all providers (HPTOT) per capita, constant prices, constant PPPs, OECD base year *(SHA)* Additional: • Outpatient curative and rehabilitative care expenditures (HC13_23) for all providers (TOTAL) PPS per inhabitant (Eurostat)

Hospital Expenditures	Main: • All healthcare expenditure (HCTOT) for all hospitals (HP1) per capita, constant prices, constant PPPs, OECD base year *(SHA)* Additional: • All healthcare expenditure (TOT_HC) for all hospitals (HP1) PPS per inhabitant (Eurostat)
Healthcare Expenditures	Main: • All healthcare expenditure (HCTOT) for all providers (HPTOT) per capita, constant prices, constant PPPs, OECD base year *(SHA)* Additional: • All healthcare expenditure (TOT_HC) for all providers (TOTAL) PPS per inhabitant (Eurostat)
Hospital physicians and nursing professional (including associates) per 1 000 inhabitants	Main: • Total hospital employment in full time equivalents *(HEALTH_REAC)* Additional: • Total hospital employment in head counts (CAN, DNK, ESP, GRC, ITA, LVA, NLD, POL, POR, SVN) *(HEALTH_REAC)* • Health personnel employed in hospital in head counts (Eurostat)
Hospital beds per 1 000 inhabitants	Main: • Curative (acute) care beds *(HEALTH_REAC)* Additional: • Total hospital beds: (AUS, GBR, GRC, POR) • Available beds in hospitals (Eurostat)
Private hospital beds per 1 000 inhabitants	Main: • Beds in for profit privately owned hospitals *(HEALTH_REAC)* Additional: • Hospital beds by hospital ownership – for profit private ownership (Eurostat)
Bed occupancy rate and idle bed capacity	Main: • Curative (acute) care occupancy rate *(HEALTH_PROC)* Additional: • Curative care bed occupancy rate (Eurostat) *Idle bed capacity = [1- bed occupancy rate] * hospital beds*
Life Expectancy at birth / at 65	Main: • Life expectancy total population at birth / at 65 (*HEALTH_STAT*) Additional: • Life expectancy in the age class "less than one year" / 65 (Eurostat)
Stand. Death Rates per 100 000 inhabitants	Main: • All causes of death per 100 000 population (standardized rates) (*HEALTH_STAT*) Additional: • Standardized death rate for the total of ages (Eurostat)

Potential years of life lost per 100 000 inhabitants	Main: • Potential years of life lost - years lost per 100 000 population aged 0 to 75 (*HEALTH_STAT*) ○ All causes Additional: • None
Unemployment rate	Main: • Unemployment, total (% of total labor force) (modeled ILO estimate) (World Bank) Additional: • None
Outpatient Consultations	Main: • Doctors consultations in all settings per capita [all settings do not include inpatient] *(HEALTH_PROC)* Additional: • Consultation of a medical doctor (in private practice or as outpatient) per inhabitant (Eurostat)

Missing data was - where justifiable - imputed using time series imputation (Moritz & Bartz-Beielstein, 2017). We considered a period of maximum three missing years to be justifiable for imputation. The following figures exemplify data availability and imputed data points.

Figure A. 14: Data availability for outcomes – Part 1



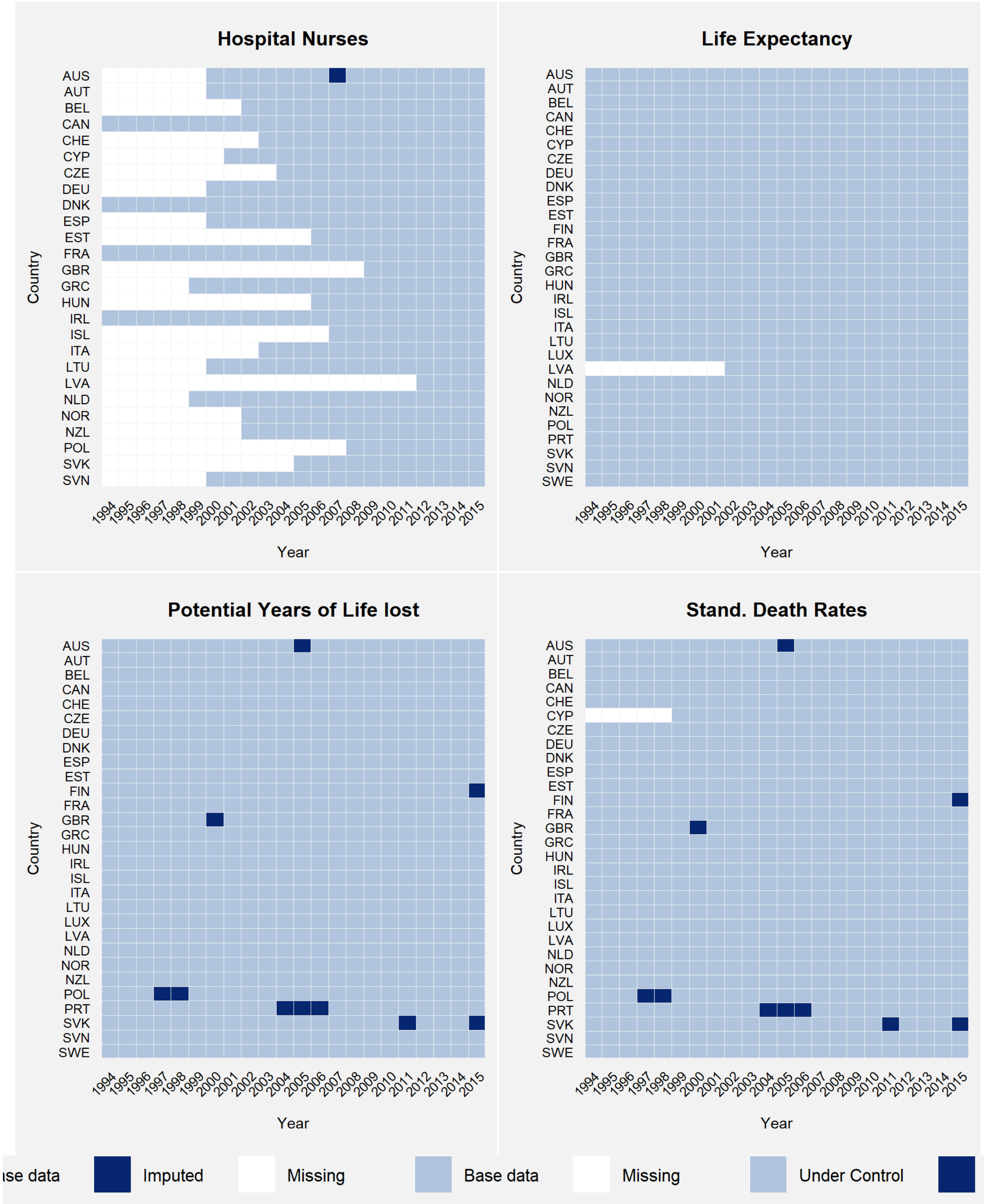


Figure A. 15: Data availability for outcomes – Part 2

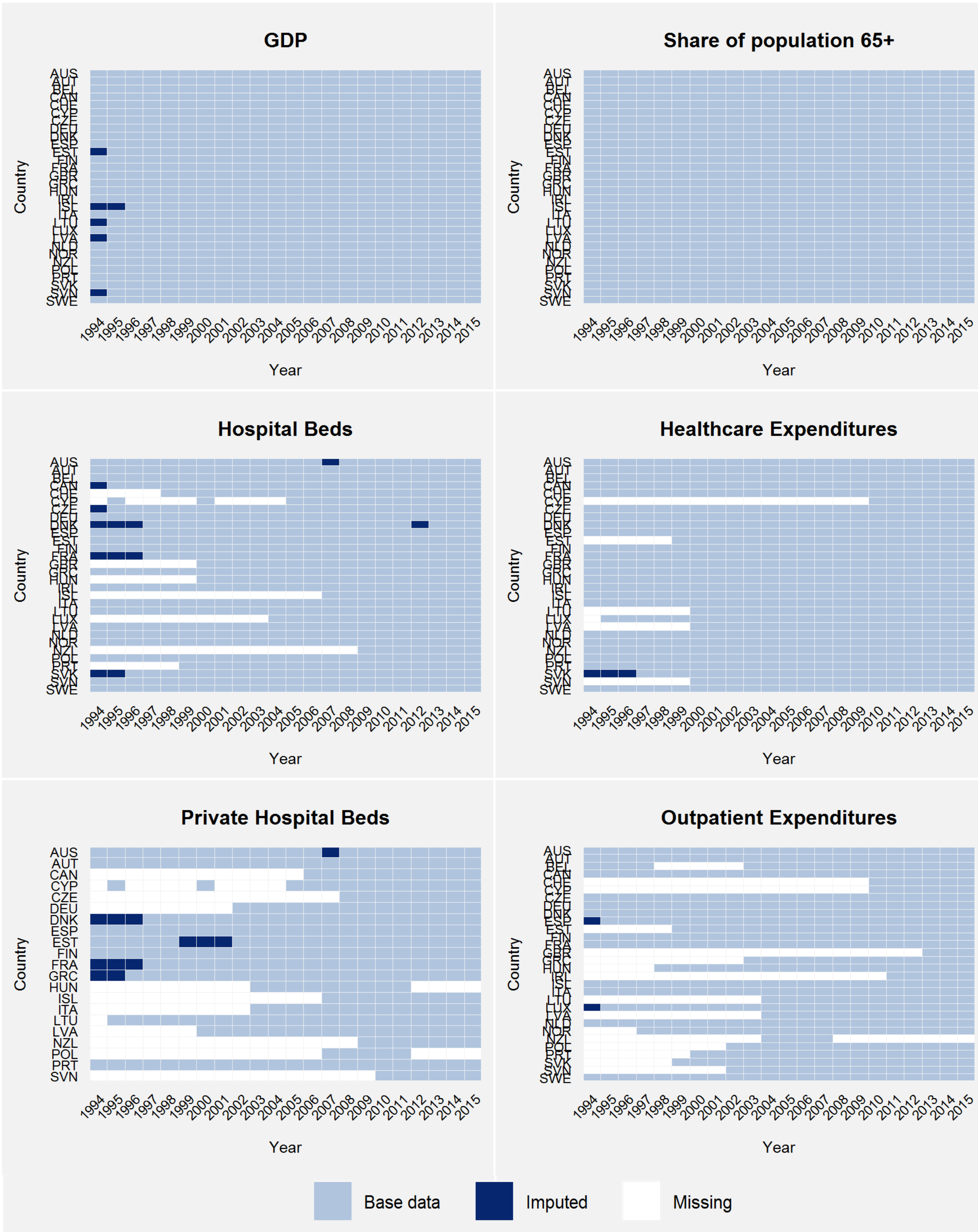


Figure A. 16: Data availability for control variables

Supporting Information References

Albreht, T., Polin, K., Brinovec, R. P., Kuhar, M., Poldrugovac, M., Rehberger, P. O., Rupel, V. P., & Vracko, P. (2021). Slovenia: health system review. *Health Systems in Transition*, *23*(1).

Anell, A., Glenngård, A. H., & Merkur, S. (2012). Sweden: Health system review. *Health Systems in Transition*, *14*(5).

Anessi-Pessina, E., Nieddu, L., & Rizzo, M. G. (2019). Does DRG funding encourage hospital specialization? Evidence from the Italian National Health Service. *The International Journal of Health Planning and Management*, *34*(2), 534–552. https://doi.org/10.1002/hpm.2715

Azzopardi-Muscat, N., Buttigieg, S., Calleja, N., & Merkur, S. (2017). Malta: Health system review. *Health Care Systems in Transition*, *19*(1).

Behmane, D., Dudele, A., Villerusa, A., Misins, J., Klavina, K., Mozgis, D., & Scarpetti, G. (2019). Latvia: health system review. *Health Systems in Transition*, *21*(4).

Bernal-Delgado, E., Garcia-Armesto, S., Oliva, J., Sanchez Martinez, F. I., Repullo, J. R., Pena-Longobardo, L. M., Ridao-Lopez, M., & Hernandez-Quevedo, C. (2018). Spain: Health System Review. *Health Systems in Transition*, *20*(2), 1–179.

Berthet, F., Calteux, A., Wolter, M., Weber, L., van Ginneken, E., & Spranger, A. (2015). Luxembourg: HiT in brief. *Health Systems in Transition*.

Blümel, M., Spranger, A., Achstetter, K., Maresso, A., & Busse, R. (2020). Germany: health system review. *Health Systems in Transition*, *22*(6).

Bredenkamp, C., Bales, S., & Kahur, K. (Eds.). (2020). *International development in focus*. *Transition to diagnosis-related group (DRG) payments for health: Lessons from case studies*. World Bank Group. https://doi.org/10.1596/978-1-4648-1521-8

Bryndová, L., Pavloková, K., Roubal, T., Rokosová, M., Gaskins, M., & van Ginneken, E. (2009). Czech Republic: health system review. *Health Systems in Transition*, *11*(1).

Busato, A., & Below, G. von (2010). The implementation of DRG-based hospital reimbursement in Switzerland: A population-based perspective. *Health Research Policy and Systems*, *8*, 31. https://doi.org/10.1186/1478-4505-8-31

Busse, R., Geissler, A., Quentin, W., & Wiley, M. M. (Eds.). (2011). *European Observatory on Health Systems and Policies series*. *Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. Open University Press.

Cantù, E., Carbone, C., & Anessi-Pessina, E. (2011). 21. Do Italian regions effectively use DRG funding to steer provider behaviour? *Policy, Performance and Management in Governance and Intergovernmental Relations: Transatlantic Perspectives*, 341.

Chevreul, K., Berg Brigham, K., Durand-Zaleski, I., & Hernández-Quevedo, C. (2015). France: health system review. *Health Systems in Transition*, *17*(3).

Christiansen, T., & Vrangbæk, K. (2018). Hospital centralization and performance in Denmark-Ten years on. *Health Policy*, *122*(4), 321–328. https://doi.org/10.1016/j.healthpol.2017.12.009

Cumming, J., McDonald, J., Barr, C., Martin, G., Gerring, Z., & Daubé, J. (2014). New Zealand health system review. *Health Systems in Transition*, *4*(2).

Cylus, J., Richardson, E., Findley, L., Longley, M., O'Neill, C., & Steel, D. (2015). United Kingdom: health system review. *Health Systems in Transition*, *17*(5).

Czach, K., Klonowska, K., Swiderek, M., & Wiktorza, K. (2011). Poland: The Jednorodne Grupy Pacjentów-Polish experiences with DRGs. In R. Busse, A. Geissler, W. Quentin, & M. M. Wiley (Eds.), *European Observatory on Health Systems and Policies series. Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals* (pp. 359–380). Open University Press.

Dubas-Jakóbczyk, K., Albreht, T., Behmane, D., Bryndova, L., Dimova, A., Džakula, A., Habicht, T., Murauskiene, L., Scîntee, S. G., Smatana, M., Velkey, Z., & Quentin, W. (2020). Hospital reforms in 11 Central and Eastern European countries between 2008 and 2019: A comparative analysis. *Health Policy (Amsterdam, Netherlands)*, *124*(4), 368–379. https://doi.org/10.1016/j.healthpol.2020.02.003

Economou, C., Kaitelidou, D., Karanikolos, M., & Maresso, A. (2017). Greece: health system review. *Health Systems in Transition*, *19*(5).

Endrei, D., Zemplényi, A., Molics, B., Agoston, I., & Boncz, I. (2014). The effect of performance-volume limit on the DRG based acute care hospital financing in Hungary. *Health Policy (Amsterdam, Netherlands)*, *115*(2-3), 152–156. https://doi.org/10.1016/j.healthpol.2013.12.005

Estonian Health Insurance Fund. (2009). *Overview of Estonian experiences with DRG system: Department of Health Economics.*

Eurostat. (2021). *Health Database*. https://ec.europa.eu/eurostat/web/health/data/database

Ferré, F., Belvis, A. G. de, Valerio, L., Longhi, S., Lazzari, A., Fattore, G., Ricciardi, W., & Maresso, A. (2014). Italy: health system review. *Health Systems in Transition*, *16*(4).

Gaál, P., Szigeti, S., Csere, M., Gaskins, M., & Panteli, D. (2011). Hungary: health system review. *Health Systems in Transition*, *13*(5).

Gerkens, S., & Merkur, S. (2010). Belgium: Health system review. *Health Systems in Transition*, *12*(5).

Health Service Executive. (2021). *Activity Based Funding (ABF) Programme Implementation Plan 2021 to 2023*. https://www.hse.ie/eng/services/publications/activity-based-funding-abf-programme-implementation-plan-2021-2023.pdf

Hilless, M., & Healy, J. (2001). Australia: Health system review. *Health Care Systems in Transition*, *13*(3).

Keskimäki, I., Tynkkynen, L.-K., Reissell, E., Koivusalo, M., Syrjä, V., Vuorenkoski, L., Rechel, B., & Karanikolos, M. (2019). Finland: health system review. *Health Systems in Transition*, *21*(2).

Kotherová, Z., Caithamlová, M., Nemec, J., & Dolejšová, K. (2021). The Use of Diagnosis-Related Group-Based Reimbursement in the Czech Hospital Care System. *International Journal of Environmental Research and Public Health*, *18*(10), 5463. https://doi.org/10.3390/ijerph18105463

Krabbe-Alkemade, Y. J. F. M., Groot, T. L. C. M., & Lindeboom, M. (2017). Competition in the Dutch hospital sector: An analysis of health care volume and cost. *The European Journal of Health Economics*, *18*(2), 139–153. https://doi.org/10.1007/s10198-016-0762-9

Kroneman, M., Boerma, W. G., van den Berg, M., Groenewegen, P., Jong, J. de, & van Ginneken, E. (2016). Netherlands: health system review. *Health Systems in Transition*, *18*(2).

Kroneman, M., & Nagy, J. (2001). Introducing DRG-based financing in Hungary: a study into the relationship between supply of hospital beds and use of these beds under changing institutional circumstances. *Health Policy*, *55*(1), 19–36. https://doi.org/10.1016/S0168-8510(00)00118-4

Lai, T., Habicht, T., Kahur, K., Reinap, M., Kiivet, R., & van Ginneken, E. (2013). Estonia: health system review. *Health Systems in Transition*, *15*(6).

Marchildon, G. P., Allin, S., & Merkur, S. (2020). *Canada: Health system review* (Vol. 22). World Health Organization. Regional Office for Europe.

Mathauer, I., & Wittenbecher, F. (2013). Hospital payment systems based on diagnosis-related groups: Experiences in low- and middle-income countries. *Bulletin of the World Health Organization*, *91*(10), 746-756A. https://doi.org/10.2471/BLT.12.115931

McDaid, D., Wiley, M., Maresso, A., & Mossialos, E. (2009). Ireland: health system review. *Health Systems in Transition*, *11*(4).

McElroy, B., & Murphy, A. (2014). An economic analysis of money follows the patient. *Irish Journal of Medical Science*, *183*(1), 15–22. https://doi.org/10.1007/s11845-013-1050-7

Mikkola, H. (2003). Hospital Pricing Reform in the Public Health Care System: An Empirical Case Study from Finland. *International Journal of Health Care Finance and Economics*, *3*(4).

Moreno-Serra, R., & Wagstaff, A. (2010). System-wide impacts of hospital payment reforms: Evidence from Central and Eastern Europe and Central Asia. *Journal of Health Economics*, *29*(4), 585–602. https://doi.org/10.1016/j.jhealeco.2010.05.007

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, *9*(1), 207. https://doi.org/10.32614/RJ-2017-009

Murauskiene, L., Janoniene, R., Veniute, M., van Ginneken, E., & Karanikolos, M. (2013). Lithuania: health system review. *Health Systems in Transition*, *15*(2).

OECD (Ed.). (2016). *Health Systems Characteristics Survey*. https://qdd.oecd.org/Home/

OECD. (2021). *Health statistics*. https://doi.org/10.1787/data-00900-en
Paris, V., Devaux, M., & Wei, L. (2010). *Health Systems Institutional Characteristics: A Survey of 29 OECD Countries: OECD Health Working Papers No. 50.* https://doi.org/10.1787/5kmfxfq9qbnr-en
Parliamentary Library. (2016). *Budget Review 2016-17: Research paper Series, 2015-16*. https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/rp/BudgetReview201617/Hospital
Pietro, C. de, Camenzind, P., Sturny, I., Crivelli, L., Edwards-Garavoglia, S., Spranger, A., Wittenbecher, F., & Quentin, W. (2015). Switzerland: health system review. *Health Systems in Transition*, *17*(4).
Polyzos, N., Karanikas, H., Thireos, E., Kastanioti, C., & Kontodimopoulos, N. (2013). Reforming reimbursement of public hospitals in Greece during the economic crisis: Implementation of a DRG system. *Health Policy (Amsterdam, Netherlands)*, *109*(1), 14–22. https://doi.org/10.1016/j.healthpol.2012.09.011
Schut, F. T., & van de Ven, W. P. M. M. (2005). Rationing and competition in the Dutch health-care system. *Health Economics*, *14*(Suppl 1), S59-74. https://doi.org/10.1002/hec.1036
Simões, J., Augusto, G. F., Fronteira, I., & Hernández-Quevedo, C. (2017). Portugal: Health system review. *Health Systems in Transition*, *19*(2), 1–184.
Smatana, M., Pažitný, P., Kandilaki, D., Laktišová, M., Sedláková, D., Palušková, M., van Ginneken, E., & Spranger, A. (2016). Slovakia: health system review. *Health Systems in Transition*, *18*(6).
Sowada, C., Sagan, A., Kowalska-Bobko, I., Badora-Musiał, K., Bochenek, T., Domagała, A., Dubas-Jakóbczyk, K., Kocot, E., Mrożek-Gąsiorowska, M., Sitko, S. J., Szetela, A. M., Szetela, P., Tambor, M., Więckowska, B., & Zabdyr-Jamróz, M. (2019). Poland: health system review. *Health Systems in Transition*, *21*(1).
Sperre Saunes, I., Karanikolos, M., & Sagan, A. (2020). Norway: health system review. *Health Systems in Transition*, *22*(1).
Stephani, V., Quentin, W., van den Heede, K., van de Voorde, C., & Geissler, A. (2018). Payment methods for hospital stays with a large variability in the care process: Health Services Research (HSR) Brussels: Belgian Health Care Knowledge Centre (KCE) 302.
Stig, K., & Paulsson Lütz, I. (2013). *Financing of Health Care in the Nordic Countries.*
Street, A., Vitikainen, K., Bjorvatn, A. and Hvenegaard, A. (2007). Introducing activity-based financing: : a review of experience in Australia, Denmark, Norway and Sweden. *CHE Research Paper*(30).
Sutherland, J., Liu, G., Crump, R. T., & Law, M. (2016). Paying for volume: British Columbia's experiment with funding hospitals based on activity. *Health Policy (Amsterdam, Netherlands)*, *120*(11), 1322–1328. https://doi.org/10.1016/j.healthpol.2016.09.010

Sutherland, J., & Repin, N. (2014). *Current Hospital Funding in Canada: Policy Brief. Vancouver: UBC Centre for Health Services and Policy Research*. www.healthcarefunding.ca

Theodorou, M., Charalambous, C., Petrou, C., & Cylus, J. (2012). Cyprus: health system review. *Health Care Systems in Transition*, *14*(6).

Theurl, E. (2015). Reform of hospital financing in Austria: Successes, failures, and the way forward. *The European Journal of Health Economics*, *16*(3), 229–234. https://doi.org/10.1007/s10198-014-0641-1

van de Voorde, C., Gerkens, S., van den Heede, K., & Swartenbroekx, N. (2013). *A comparative analysis of hospital care payments in five countries* (KCE Reports No. 207). Brussels. Belgian Health Care Knowledge Centre (KCE).

WHO. (2021). *European Observatory on Health System and Policies*. https://eurohealthobservatory.who.int/

World Bank (Ed.). (2021). *World Bank Open Data*. https://data.worldbank.org/

Wubulihasimu, P., Brouwer, W., & van Baal, P. (2016). The Impact of Hospital Payment Schemes on Healthcare and Mortality: Evidence from Hospital Payment Reforms in OECD Countries. *Health Economics*, *25*(8), 1005–1019. https://doi.org/10.1002/hec.3205

hche Research Paper Series, ISSN 2191-6233 (Print), ISSN 2192-2519 (Internet)

2011/1 Mathias Kifmann and Kerstin Roeder, Premium Subsidies and Social Insurance: Substitutes or Complements? March 2011

2011/2 Oliver Tiemann and Jonas Schreyögg, Changes in Hospital Efficiency after Privatization, June 2011

2011/3 Kathrin Roll, Tom Stargardt and Jonas Schreyögg, Effect of Type of Insurance and Income on Waiting Time for Outpatient Care, July 2011

2012/4 Tom Stargardt, Jonas Schreyögg and Ivan Kondofersky, Measuring the Relationship between Costs and Outcomes: the Example of Acute Myocardial Infarction in German Hospitals, August 2012

2012/5 Vera Hinz, Florian Drevs, Jürgen Wehner, Electronic Word of Mouth about Medical Services, September 2012

2013/6 Mathias Kifmann, Martin Nell, Fairer Systemwettbewerb zwischen gesetzlicher und privater Krankenversicherung, July 2013

2013/7 Mareike Heimeshoff, Jonas Schreyögg, Estimation of a physician practise cost function, August 2013

2014/8 Mathias Kifmann, Luigi Siciliani, Average-cost Pricing and Dynamic Selection Incentives in the Hospital Sector, October 2014

2015/9 Ricarda Milstein, Jonas Schreyögg, A review of pay-for-performance programs in the inpatient sector in OECD countries, December 2015

2016/10 Florian Bleibler, Hans-Helmut König, Cost-effectiveness of intravenous 5 mg zoledronic acid to prevent subsequent clinical fractures in postmenopausal women after hip fracture: a model-based analysis, January 2016

2016/11 Yauheniya Varabyova, Rudolf Blankart, Jonas Schreyögg, Using Nonparametric Conditional Approach to Integrate Quality into Efficiency Analysis: Empirical Evidence from Cardiology Departments, May 2016

2016/12 Christine Blome Ph.D., Prof. Dr. Matthias Augustin, Measuring change in subjective well-being: Methods to quantify recall bias and recalibration response shift, 2016

2016/13 Michael Bahrs, Mathias Schumann, Unlucky to be Young? The Long-Term Effects of School Starting Age on Smoking Behaviour and Health, August 2016

2017/14 Konrad Himmel, Udo Schneider, Ambulatory Care at the End of a Billing Period, March 2017

2017/15 Philipp Bach, Helmut Farbmacher, Martin Spindler, Semiparametric Count Data Modeling with an Application to Health Service Demand, September 2017

2018/16 Michael Kvasnicka, Thomas Siedler, Nicolas R. Ziebarth, The Health Effects of Smoking Bans: Evidence from German Hospitalization Data, June 2018

2019/17 Jakob Everding, Heterogeneous Spillover Effects of Children's Education on Parental Mental Health, July 2019

2019/18 Esra Eren Bayindir, Hospital Ownership Type and Service Provision, a Structural Approach, November 2019

2019/20 Shushanik Margaryan, Low Emission Zones and Population Health, December 2019

2020/21 Barbara Boggiano, Long-ter effects of the Paraguayan War (1864-1870): from male scarcity to intimate partner violence, May 2020

2020/22 Matthias Bäuml, Christian Kümpel, Hospital Responses to the Introduction of Reimbursements by Treatment Intensity in a (Presumably Lump Sum) DRG System, May 2020

2020/23 Philipp Bach, Victor Chernozhukov, Martin Spindler, Insights from optimal pandemic shielding in a multi-group SEIR framework, November 2020

2021/24 Florian Hofer, Benjamin Birkner, Martin Spindler, Power of machine learning algorithms for predicting dropouts from a German telemonitoring program using standardized claims data, June 2021

2021/25 Malte Griebenow, Mathias Kifmann, Diagnostics and Treatment: On the Division of Labor between Primary Care Physicians and Specialists, November 2021

2022/26 Robert Messerle, Jonas Schreyögg, System-wide Effects of Hospital Payment Scheme Reforms: The German Introduction of Diagnosis-Related Groups, March 2022

The Hamburg Center for Health Economics is a joint center of Universität Hamburg and the University Medical Center Hamburg-Eppendorf (UKE).

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

UKE
HAMBURG

hche | Hamburg Center for Health Economics

Esplanade 36
20354 Hamburg
Germany
Tel: +49 (0) 42838-9515/9516
Fax: +49 (0) 42838-8043
Email: info@hche.de
http://www.hche.de

HCHE Research Papers are indexed in RePEc and SSRN.
Papers can be downloaded free of charge from http://www.hche.de.