Universität Bayreuth

Rechts- und Wirtschaftswissenschaftliche Fakultät

Juniorprofessur Gesundheitsmanagement

Prof. Dr. Andreas Schmid

Master Thesis

Does the introduction of a prospective payment system influence the length of stay in hospitals?

A quantitative analysis using aggregated data from Germany

Abstract

This paper investigates whether the introduction of a prospective payment system in the German inpatient sector in 2004 had an effect on the average length of hospitalizations. A depiction of the features of the new reimbursement system and its predecessor is followed by a discussion of the relevance and development of the average length of hospital stays. Providing a short overview of relevant aspects of principal-agent theory rounds of the introductory chapters. An extensive literature review on the theoretical and empirical findings on the impact of the introduction of prospective payment systems builds the foundation on which an economic model for hospitals under prospective payment is constructed. According to the results of the review and the implications of the economic model, the hypothesis was formulated that the DRG introduction led to the politically intended decrease of length of stay in German hospitals. The analysis continues with summarizing key elements of the aggregated data set on the full sample of German hospitals from 2000 to 2012, utilized to test this hypothesis. The econometric approach and the used truncated Poisson model are described in detail before the general results are presented and discussed. These results allow the conclusion that there has been a small reduction of the length of hospitalizations due to the DRG introduction. However, several limitations, mainly attributed to the used data, have to be considered when interpreting the results. The depiction of these limitations and a subsumption of the results in a broader context complete this paper.

Number of characters: 110,965

Contents

Index of Figures

Index of tables

Index of Acronyms and Abbreviations

AOP	Ambulatory Conducted Procedures and Surgeries
Destatis	Statistisches Bundesamt
DRG	Diagnosis Related Groups
e.g.	exempli gratia
et al.	et alii
GM	German modification
i.e.	id estl
ICD	International Classification of Diseases
InEK	Institute for the Hospital Remuneration System

KHG	Krankenhausentgeltgesetz
LOS	length of stay
MeSH	Medical Subject Headings
MHOW	Ministry of Health and Welfare
OECD	Organization for Economic Cooperation and Development
p.	page
pp.	pages
PPS	prospective payment system
U.S.	United States
vdek	Verband der Ersatzkassen

1 Introduction

1.1 Motivation and scope of the analysis

The introduction of the DRG system in Germany in 2003 was accompanied by compulsory supporting scientific research, which was meant to quantify the effects of the reform on a variety of parameters such as treatment quality and service delivery (§ 17b (8) KHG). One of the main goals of the reform was to abolish incentives to prolong hospital stays (Deutscher Bundestag, 2001, p. 26). In the 598 pages of the final report of the accompanying research drafted by the IGES Institute only ten pages are explicitly dedicated to examine the length of stay (LOS) development after the reform. On the basis of a purely descriptive analysis, this short disquisition concluded that a distinct change of the general LOS trend after the DRG introduction “cannot be identified” (Fürstenberg et al., 2013, p. 168). Both scope and manner of this inquiry represent an unsatisfactory and deficient analysis of one of the central objectives of the allegedly most influential hospital reform in the last decades. As a consequence, there is need for a thorough and well-grounded analysis of this matter even over ten years after the actual reform. It is important to understand the impacts of hospital reforms especially in a time, in which the frequency of policy interventions in the inpatient sector appears to continuously increase. Hence, the work at hand will provide an exhaustive and elaborate analysis of the impact of the DRG reform on the LOS in German hospitals on the basis of an aggregated data set, attempting to achieve the level of scrutiny this kind of discussion deserves. In the course of this analysis, following questions will be answered: Did the DRG introduction change the financial incentives concerning the length of hospitalizations, and is there empirical evidence of a LOS change due to the reform?

1.2 Approach

To achieve the mentioned scope and to answer these questions, the following approach has been chosen for the analysis, which roughly can be split into three parts:

First, chapters Two to Four set the stage for the subsequent methodological examination of the matter at hand. The first of these chapters starts out with a short overview of the historical development of hospital reimbursement in Germany and then presents facts about the background and development of the reform before discussing key features of the German DRG system itself. Chapter Three discloses, why policy makers should be and are interested in the length of hospitalizations, before illustrating the national and international development of LOS itself in the last decades. The following chapter contains a discussion about which economic theory and framework must be applied, in the context of examining a change of incentives in the inpatient sector.

The second part, chapters Five to Seven, builds the foundation of the statistical analysis. Chapter Five presents methods and results of an exhaustive literature review on the theoretical and empirical impact of a shift to a prospective payment system (PPS) on LOS. Both national and international findings are included in the review. On the basis of these findings and the general framework, set in the first chapters, an economic model is developed in chapter Six. This model explicitly illustrates the change of incentives associated with and intended by the introduction of a PPS. Chapter Seven condenses the heretofore established findings and conclusions by formulating explicit hypothesis about the impact of the reform on LOS.

The last, but most important part, from chapters Eight to Eleven, contains the necessary steps of a sound statistical analysis of the aggregated data set. Chapter 8 provides a discussion of the key features of the data used in the subsequent calculations. The following chapter starts out by illustrating the general econometric approach, before presenting the dependent variable of the analysis. Considerations concerning model selection and the inclusion of covariates into the final econometric model build the endpoint of the methods section. Chapter Ten presents descriptive statistics and detailed regression results before concluding with presenting several robustness checks. Chapter Eleven builds the endpoint of the analysis itself by summarizing and discussing the findings of the statistical analysis and presenting several limitations of the chosen approach. The last chapter concludes this paper with a brief subsumption of the general findings of the analysis in its entirety.

2 DRG reform in Germany

2.1 Historical development of hospital reimbursement

Before being able to assess the impact of the DRG reform, it is necessary to understand the G-DRG system itself and the way hospitals in Germany were financed before the reform. Hospital reimbursement in Germany can roughly be divided into three phases, as depicted in figure 1.

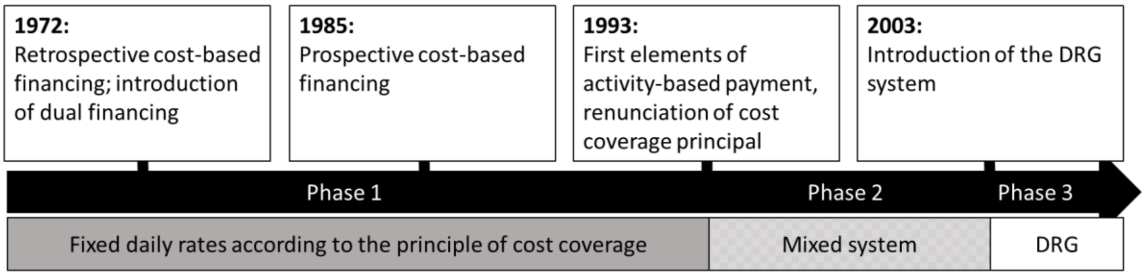

Figure 1: Phases of hospital reimbursement
Source: Own illustration after (Fleßa, 2013, pp. 134–137).

In the first decades after the war, hospitals in the former western part of Germany were reimbursed almost exclusively according to hospital-specific daily fixed rates (Fleßa, 2013, pp. 134–135). This slightly changed with the introduction of the first elements of activity-based payment in 1993. This year also marked the end of the principle of cost coverage, where hospitals profited from loss compensation (Rau et al., 2009, p. 9). As the systems before, the mixed system was not able to reduce the growth of hospital costs (Hilgers, 2011, pp. 24-25). The use of a per-diem payment system especially incentivized hospitals to unreasonably prolong hospital stays (Hilgers, 2011, pp. 38–40). The probable result of this unintended consequence can be seen in figure 2, which plots the LOS averages for different OECD countries in the year 2000.

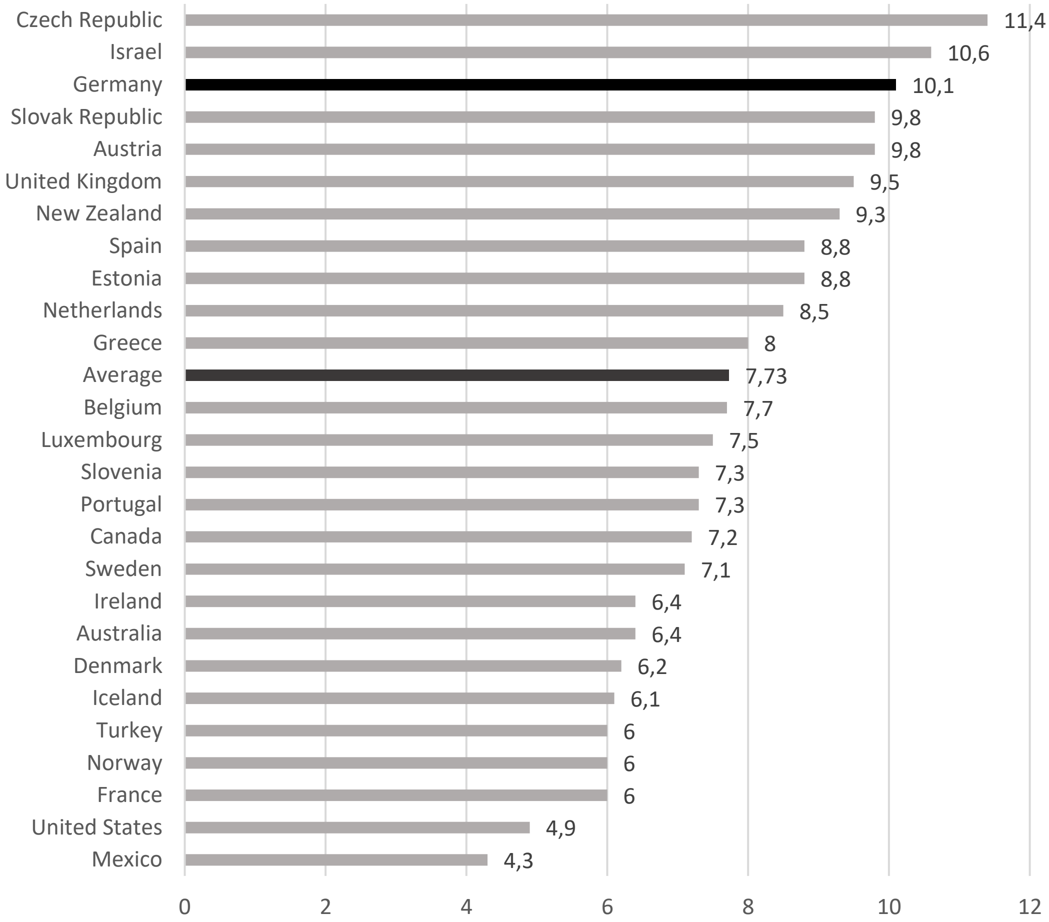

Figure 2: Average length of stay (ALOS) in selected OECD countries in 2000
Source: Own illustration, data from OECD (2016).

As daily rates were still predominantly negotiated individually for each hospital, competition and incentives for an efficient provision of care were still quite weak (Debatin, 2008, p. 393). Therefore, German policy makers were forced to consider a paradigm change.

2.2 DRG reform: Background and development

The "GKV-Gesundheitsreformgesetz 2000" introduced an activity-based prospective payment system, with the goal to enhance efficiency and quality in the German health care sector (Deutscher Bundestag, 2003, p. 1). The former mentioned fixed daily rates were substituted by diagnose-oriented lump sums, the so called "diagnosis related groups" or DRG (Fleßa, 2013, p. 138-139). These were originally developed by Yale University in the late 1970s and were first applied for the reimbursement of Medicare patients in the 1980s. Since then, numerous different DRG systems were introduced in various countries (Kobel et al., 2011, pp. 38–40). After extensive groundwork, e.g. by

Fischer (2000), German policy makers selected the Australian Refined DRG Version 4.1 as basis for the development of a German DRG system (Rau et al., 2009, pp. 9–10). The German Institute for Medical Documentation and Information (DIMDI) then converted the Australian procedure and diagnostic codes to the German procedure classification codes (OPS) and the International Classification of Diseases (ICD), 10th revision, German modification (GM) diagnostic codes. Relative cost weights were calculated by the newly founded Institute for the Hospital Remuneration System (InEK) on the basis of the cost data of approximately 100 German hospitals. The first version of the German DRG (G-DRG) system with 664 DRGs was available by the end of 2002 (Geissler et al., 2011, p. 248). The year 2003 marked the starting point of the G-DRG system. However, the application of the DRG systematic in this year was still on a voluntary basis and budget neutral - a chance for the hospitals to get accustomed to the process and a test run for the governing bodies. Starting January 2004, accounting via DRG was mandatory for all inpatient treatments provided by hospitals contracting with the statutory health insurance. However, by the end of 2004, only 1580 of more than 2000 hospitals actually applied the DRG systematic. Practically full compliance was reached in 2005 (Friedrich et al., 2010, p. 138).

2.3 The G-DRG system

2.3.1 Patient classification

Fleßa (2013, pp. 144–163) and Geissler et al. (2011, pp. 247–255) provide a thorough description of the functionality of the German DRG system. The following expositions present a summary of the most important features discussed in their publications.

The core of any such system is a mechanism for classifying treatment cases into economical homogenous groups. Grouping software employs the inpatient hospital discharge dataset for that purpose, which contains major diagnosis and other diagnoses, medical procedures, patient characteristics (e.g. age, gender), LOS, duration of ventilation, reason for hospital discharge, and type of admission. ICD-10-GM and OPS are used to code diagnoses and procedures.

In general, the first step in the classification system is to assign the case into one of 25 major diagnostic categories (MDC) according to the main diagnosis that was causal for the hospital admission. Afterwards, the type of procedures conducted define the "partitions" (surgical, other and medical). In combination with the MDC, this forms the "base-DRG". In 2016, 590 different base-DRGs could be distinguished (Fischer, 2015, article 68). In the last step co-morbidities, additional procedures, patient characteristics and cost data are used to group cases into different categories of treatment/resource intensity. The total number of different DRGs has amounted to 1220 in 2016 (Fischer, 2015, article 68).

2.3.2 DRG reimbursement

DRGs have to be linked to financial compensation by assigning cost weights to all DRGs. The weights represent the economical effort of treating a patient relative to a base case The product of these case weights and a price factor, the base rate, eventually build the monetary compensation for treating a specific patient (Fleßa, 2013, p. 154-155). The base rates were calculated hospital specific in the first two years of the DRG system. Case mix adjusted hospital individual average costs per case in 2003 were the basis of the hospital specific base rates in 2004.[1] They varied between 2200 € and 3200 € depending on region and hospital characteristics (Geissler et al., 2011, pp. 249–251). Starting 2005, the hospital specific base rates were gradually converging to binding state-wide base rates forcing high cost hospitals to reduce costs. Possible budget reductions were confined to a certain yearly percentage between 2005 and 2009 (1-3 %). From 2010 onwards, all inpatient cases were reimbursed according to a binding state-wide base rate, which are yearly adjusted by the InEK (Geissler et al., 2011, p. 250).

The state rates differed significantly, ranging from 2.585 € to 3.000 € in 2005 (vdek, 2016). The reasons for those differences are still not entirely clear and are discussed in (Augurzky and Schmitz, 2013). In the years after the phase of convergence, the state-wide base rates were supposed to be consolidated into one federal base rate to establish one single price for DRG services. Up to the end of 2014, state rates should have converged to a corridor of 2.5 % above and 1.25 % below the federal base rate. Starting 2016, a fully convergence could still not be achieved. The base rates move between 3.264,35 € and 3.465,02 € around the federal base rate of 3,311.98 € (vdek, 2016).

[1] The case mix is the aggregate measure for a hospital's case severity.

Under certain circumstances, in order to increase the adequacy of the reimbursement system, the actual reimbursement deviated from the principal of average costs per care episode. Special arrangements had to be made for day outliers, to shape the incentives in a way, that hospitals would not inappropriately prolong or shorten hospital stays. To account for cost outliers, additional charges for certain products and services for a defined group of patients were introduced. Typical examples are dialysis, extremely expensive cancer medications and elaborate implants. Further special arrangements had to be designed, concerning the treatment of transferals and readmissions, to prevent hospitals to split cases and bill them separately (Fleßa, 2013, pp. 156–161).

3 Length of stay: Relevance and development

3.1 Relevance of length of stay

Why are managers and policy-makers interested in the LOS? German lawmakers apparently deemed it important enough to explicitly mention the LOS development, when formulating the goals for the "GKV-Gesundheitsreformgesetz 2000" (Deutscher Bundestag, 2001, p. 26).

From a patient's perspective, it is clear that an adequate LOS would be desirable. Most of them want to get home as soon as possible, others might want to stay longer to fully recover. However, hospitalizations are also connected to adverse events. A review by Vries et al. (2008) states, that almost one out of ten patients experiences an adverse event (Vries et al., 2008, p. 219). Another study estimated, that each additional day in a hospital increases the probability of an adverse drug event by 0.5%, an infection by 1.6%, and an ulcer by 0.5% (Hauck and Zhao, 2011, p. 1072). Therefore a reduction in LOS in itself can be seen beneficial.

Moreover, the ALOS is a verifiable hospital input measure for quality and performance (Chalkley and Malcomson, 2000, p. 853). Thomas et al. (1997) showed, that LOS can be used as an indicator for quality of care (Thomas et al., 1997, pp. 489–491). But, as otherwise similar patients can experience significant variations in the length of their hospitalization, LOS foremost, can be used as a measure for efficiency in hospitals (OECD, 2015, p. 108). LOS could either be applied as a modified index (Kourie, 1976, pp. 959–960), or directly as an approximation for the average resource usage per case to try to determine inefficiencies when comparing hospitals (Herwartz and Strumann, 2014, p. 179). The appeal of LOS as an efficiency measure is that the concept is quite plausible and changes in LOS can be measured effortlessly by the hospital management or other stakeholders (Clarke, 1996, p. 172). The crucial assumption of the viability of LOS as an efficiency indicator is, that reductions in the length of hospitalizations are achieved through increased productivity and efficiency, without compromising the quality of care and patient's outcomes.

Additionally, curtailing hospital stays not only reduces the cost per patient episode within the inpatient sector, but also has the potential to reduce costs for the whole continuum of care, as care is shifted towards less costly outpatient facilities (OECD, 2015, p. 108). LOS therefore can be viewed as an indicator for the efficiency of a health system in its entirety. It is noteworthy however, that reductions in LOS do not necessarily imply overall cost reductions, as the number of patients treated might rise, since hospitals might try to maintain the same operating grade (Clarke, 1996, p. 176).[2] Or invoking Roemer's law: A built bed is a filled bed (Roemer, 1961, pp. 36–42). This phenomenon was discussed by Augurzky et al. (2012) in the context of the German hospital system.

In conclusion, under the right circumstances, LOS can be used as a feasible measure for hospital efficiency. It is important, however, to be able to disentangle whether the change in LOS comes from real efficiency and productivity gains or from quality reductions and cost shifting. In order to achieve that, a high level of transparency and comprehensive quality assurance mechanisms are necessary. This was one of the reasons, why the introduction of the DRG system in Germany was convoyed by multiple quality assurance mechanisms (see for example Hilgers, 2011, pp. 51–54).

3.2 Development of length of stay

Figure 3 depicts the development of the LOS in Germany and comparable countries from 1994 to 2010.[3] Although there are huge variations, a general downward trend can be identified with steep reductions in the 1990s especially in countries with a high level of

[2] Some evidence for increases in hospital cases are summarized in Böcking et al. (2005, p. 133)

[3] It is noteworthy, that the point estimates for the LOS in Germany differ between the two used data sources, Statistisches Bundesamt and OECD. In- and exclusion criteria, the way deaths and one day cases are treated and whether the day of discharge is included, might be responsible for this difference.

LOS. The reasons for this downward trend are manifold and might lie in the use of less invasive surgical procedures, changes in hospital payment methods, and the expansion of early discharge programs (OECD, 2015, p. 108). An analysis by Sloan and Valvona (1986, pp. 71-72) identified technological change as one of the main driver of the LOS reduction in surgical cases. Improvements in surgical techniques, instruments and material made surgery less invasive, enabling a quicker recovery.

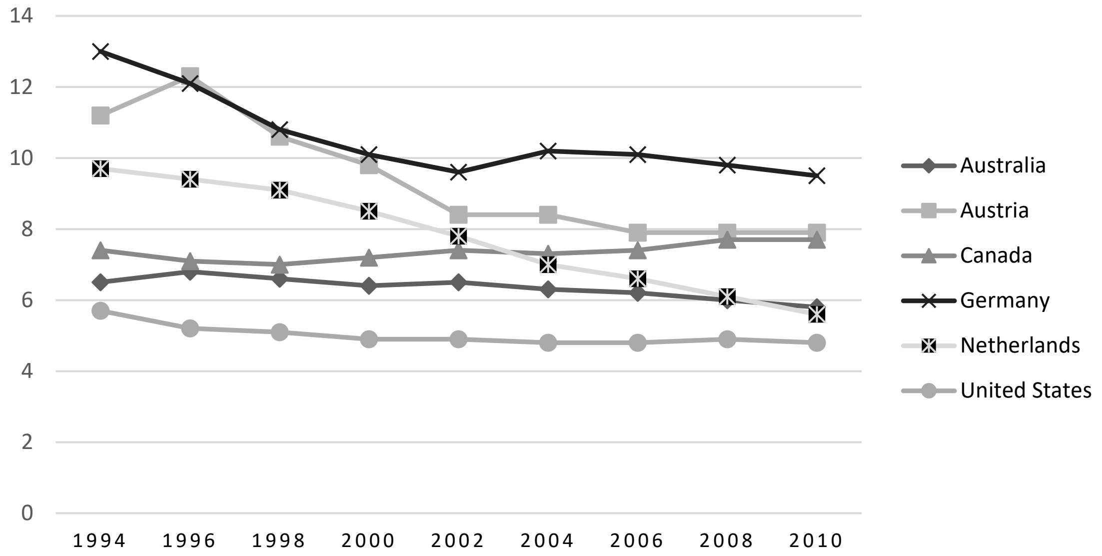

Figure 3: International development of LOS
Source: Own illustration, data from OECD (2016).

The only country where LOS did not drop during that period is Canada. The LOS level was already quite low in 1994, but countries like the United States (U.S.) and Australia still managed to decrease LOS significantly. One of the reasons might be that Canada adhered to its hospital financing system through global budgets for the longest time (McKillop et al., 2001, pp. 43–56), whereas the other countries at some point introduced activity based funding (Kobel et al., 2011, p. 39). Whether activity based funding did indeed play a role in the reduction of LOS in Germany in particular, depicted in figure 4, will be established in the course of this paper.

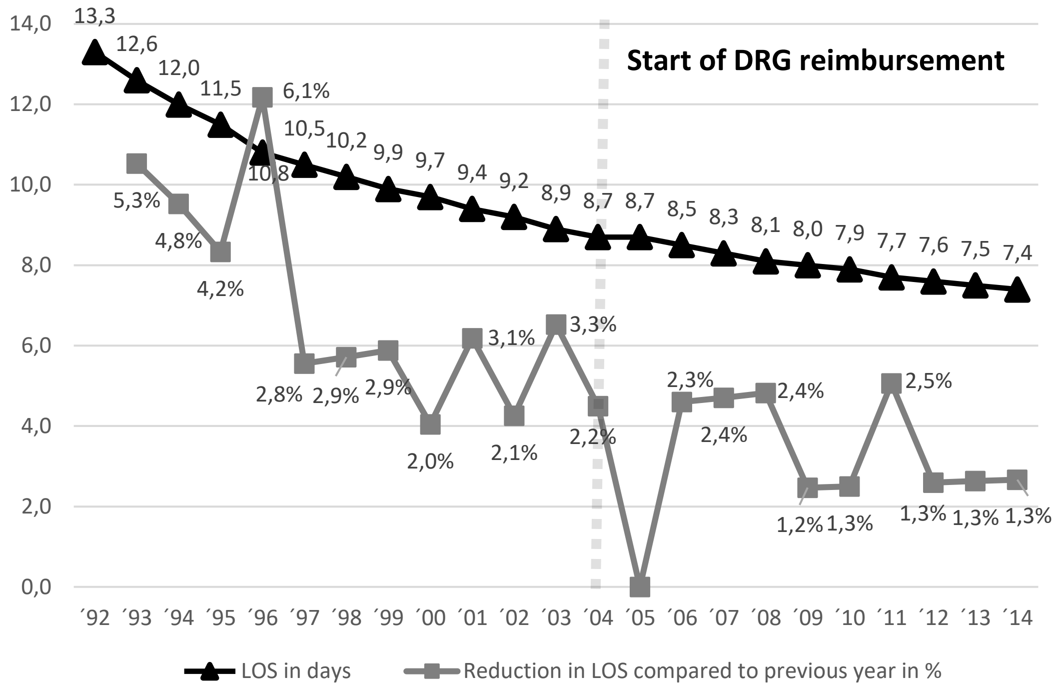

Figure 4: LOS development in Germany
Source: Own illustration, data from Destatis, 2016c.

LOS has been declining from 1992 to 2003 from 13.3 days to 8.9 days and the reduction rate varied between 6.1% and 2.0%. In the years after the reimbursement reform came into effect, the reduction rate was noticeable smaller, never exceeding 2.5% and even dropped to 0% in 2005. LOS can be influenced by a plurality of factors. Trends from either the demand or the supply-side can superimpose a possible impact of the reform. It will be crucial for this analysis to adequately account for that.

4 Theoretical framework

4.1 Special characteristics of health care

It is important for the course of this analysis to understand how health care markets and especially hospitals can be depicted in an economical framework. Basic microeconomic theory is a good starting point, but has to be modified by some characteristics (Folland et al., 2013, p. 12). The most relevant deviations are firstly: the production and consumption of medical care occurs simultaneously, making it impossible for the consumer to check the quality of the product before utilization. Trust and ethical restrictions therefore play an important role. Additionally, the supply side is heavily regulated and there are significant market access barriers, such as licensing, which limits the provider choice drastically. Further deviations are unusual reimbursement mechanisms and the virtually non-existent price competition.

Most important for the course of this analysis is, however, the heavily pronounced product uncertainty. Besides a general high level of uncertainty about parameters like incidence and probability of recovery, there is another special feature to it in health care markets: an unparalleled level of asymmetry of information, particularly about the possible outcomes of the consumption of a certain product (Arrow, 1963, pp. 948–954).

4.2 Asymmetry of information and principal-agent theory

Of special interest for this analysis is the relationship between the physician, or respectively the hospital and the patient. The latter is often poorly informed about his medical condition, available treatments, outcomes, and prices in the market (Folland et al., 2013, p. 197). Asymmetric information within a business relationship is both the main assumption and the starting point of principal-agent theory, a part of the economics of organizations, analyzing agency relationships within institutions, as well as firms and bureaucracies (Buchanan, 1988, p. 317; Furubotn and Richter, 2000, pp. 181–182).

An agency relationship can be defined as a contract under which a principal (patient) delegates decision making authority to another party, the agent (physician/hospital), or engages the agent to perform some service on their behalf (Jensen and Meckling, 1976, p. 308; Folland et al., 2013, p. 202). The principal's motive to delegate certain tasks is, that he wants to utilize the specialized knowledge and skills of the agent, as the necessary effort to be able to self-produce is prohibitive (Sappington, 1991, p. 45). Asymmetry of information therefore is by construction a part of an agency relationship and just the other side of the medal of utilizing the superior knowledge of the agent (Buchanan, 1988, p. 318). Figure 5 depicts the general case of the principal-agent relationship.

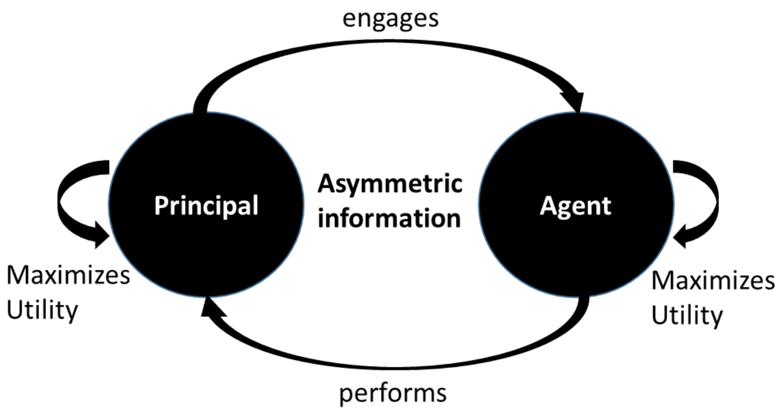

Figure 5: Principal-agent Relationship
Source: Own illustration, after Jensen and Meckling (1976, p. 308) and Buchanan (1988, p. 318).

Principal and agent are both assumed to be utility maximizers. This raises the possibility that the agent might not always act in the best interest of the principal, deviating from perfect agency (Buchanan, 1988, p. 318). In a world with complete and symmetric information, this could be solved by drafting a complete contract (Milgrom and Roberts, 1992, p. 127). In reality , however, as the principal cannot directly observe the activities of the agent, it is necessary to establishing appropriate incentives and monitoring mechanisms, to limit possible deviations from the principal's interests (Jensen and Meckling, 1976, p. 308). Depending on whether the asymmetry of information occurs before or after the formation of the contract, two types of principal-agent problems can be identified and are depicted in Figure 6: Adverse selection (pre contractual) and moral hazard (post contractual).

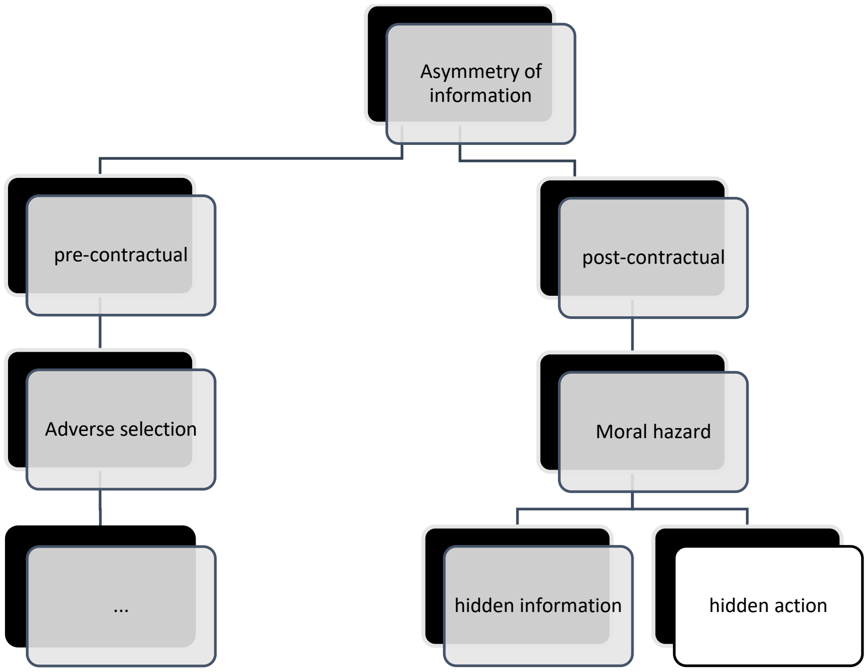

Figure 6: Classification of information asymmetry
Source: Own illustrations, based on (Furubotn and Richter, 2000, pp. 181–182).

Adverse selection implies, that the agent might try to engage in opportunistic behavior before (ex-ante) a contract is forged (Furubotn and Richter, 2000, pp. 181–182). The problem of adverse selection is most famously described by Akerlof in his paper about the used-car market (Akerlof, 1970). His lemons principle is widely applied to the health insurance market, where asymmetry of information leads to inefficiencies in the market (Folland et al., 2013, pp. 199–201). Opportunistic behavior on the grounds of ex post information advantage can be categorized in hidden information and hidden action moral hazard. The latter implies that the efforts and actions of the agent cannot be observed by the principal (Furubotn and Richter, 2000, pp. 181–182).

4.3 Relevance of principal-agent theory in the analysis

In the physician/hospital-patient relationship, considered in this analysis, patients act as principal, giving away authority about the treatment decisions, like duration of the hospitalization. The physician, respectively the hospital, acts as the agent, performing services and deciding about the treatment on behalf of the patient. As the analysis concerns patients already contracting with hospitals, the principal-agent problem occurs post contractually. The relevant opportunistic behavior lies in hidden actions, as the principal cannot observe whether the decision by the agent, about the LOS was in his best interest, or not. If the LOS turns out to be dependent on the financial incentives set by the regulator, it becomes apparent, that the decisions, about keeping or discharging patients are not solely based on the patient's objective need for health care and that hospitals are not acting as perfect agents and are maximizing their own utility function.

5 Literature Review of prospective payment system introductions

5.1 Methods

The following chapters present methodology and results of the extensive literature review on the theoretical and empirical implications of the introduction of an activity based payment system. A two-step identification process was applied for the review.

First, relevant databases were searched for theoretical and empirical studies on the effect of the introduction of a PPS. LOS was of particular interest. Second, the results of the search were used to identify possible further research, by analyzing the reference lists. As the topic at hand contains both medical and economical aspects, it was necessary to use databases from each discipline to detect the whole range of the relevant literature. On the grounds of accessibility, standing in the scientific community and search engine mechanisms, PubMed and EconLit were chosen for the review.

To identify relevant key terms or MeSH-terms, generic search strings were developed in the beginning. Different strings revealed additional suitable search terms, starting an iterative process, which lead to the inclusion of 20 search terms. These can be categorized in supply side, reimbursement and theoretical terms, summarized in table 1. Using German terminology and definitions did not produce meaningful results, since both search engines are more or less restricted to the English language. As 95% of the economics literature from Germany is published in English (Toepfer, 2011, p. 247), and the better part of the medical literature at least provides an abstract in English, the chosen search strategy should also cover the relevant literature from Germany.

Supply side	Reimbursement	Theory
Hospital	DRG	Incentives
Clinic	Diagnosis related groups	Principal-agent theory
LOS	Prospective payment system	Principal-agent relationship
Hospital efficiency	Reimbursement mechanism	Principal-agent model
Hospital reform	Prospective reimbursement	Information asymmetry
Healthcare reform	Activity based funding	Efficiency
	Lump sum	Effect

Table 1: Search terms

The terminology concerning prospective hospital reimbursement mechanisms is not homogenous, therefore it was necessary to embed different terms in the search strings. As seen in chapter 4, principal-agent theory depicts a suitable way to model patient-physician or hospital-patient relationships, providing "principal-agent theory" as the key component of the theoretical search strings.

The identified terms were iteratively combined by using boolean operators AND, OR, and NOT. If possible, MeSH terms were used in the PubMed search. Overall, this resulted in five search strings in PubMed and ten search strings in EconLit. The latter were intended to identify most of the relevant theoretical paper and therefore included more economical key words than the PubMed search strings. Since the identification of the key search terms was an ongoing process, going hand in hand with the screening of the results of the first strings, it was not feasible to incorporate the search terms in one single search string.[4] The review was supplemented by one Google Scholar search (principal-agent theory AND hospital from 1970 to 2000), to check whether there were any important papers missing from the early discussion of principal-agent theory in healthcare. The database search in PubMed and EconLit itself was not restricted to any kind of regional or chronological boundaries. Figure 7 presents the flow of acquisition of relevant studies.

[4] Appendix A contains a list of the different search strings and their results.

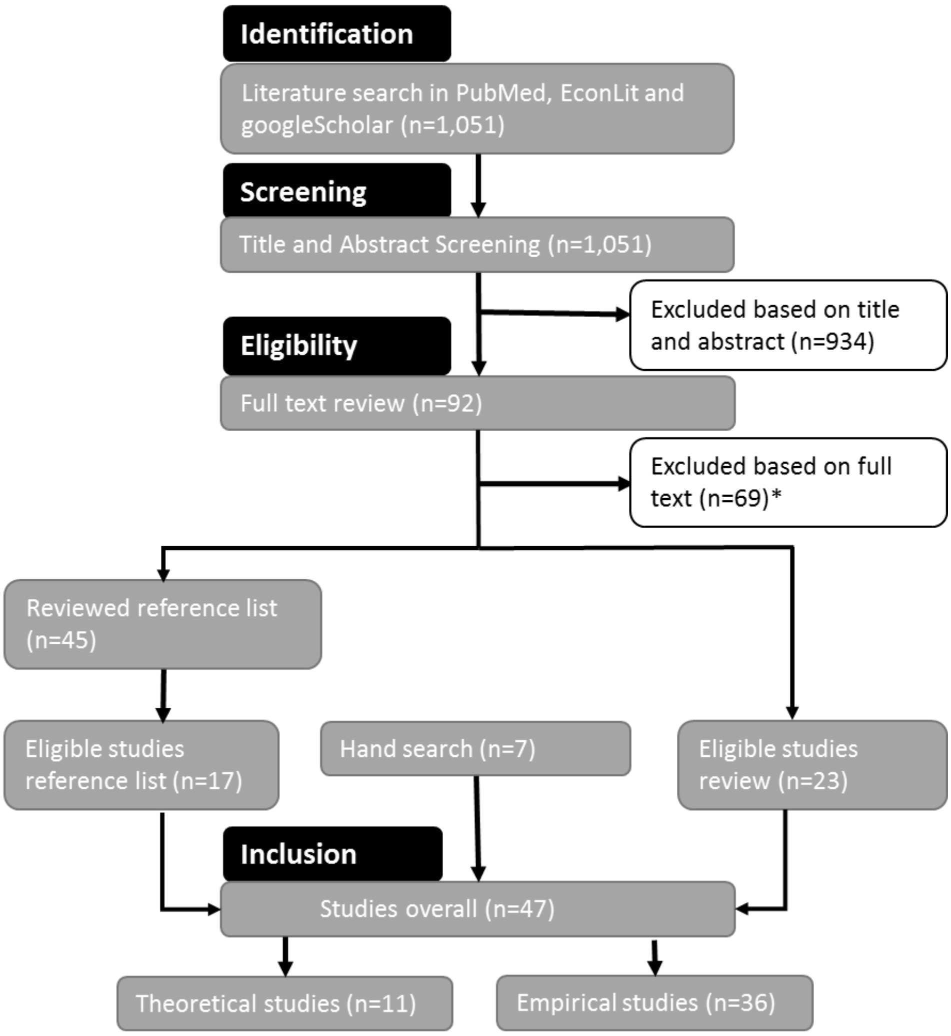

Figure 7: Flow chart of literature search
*Exclusion criteria are summarized in table 2.

The review was conducted in January 2016 and the initial search strings identified 1,051 citations, 272 from PubMed, 628 from EconLit and 151 from Google Scholar. After screening titles and abstracts according to their relevance for the discussion of PPSs and their implications, 92 papers remained eligible for inclusion in the review. Of these eligible articles, it was not possible to obtain the full texts for four. The full text analysis lead to the inclusion of 23 and exclusion of 69 studies (Figure 7). Exclusion criteria are summarized in table 2.

Exclusion Criteria
Overall
- Full text not available
- Duplicate of already included paper
- Paper contains analysis of different payment system
- Lacking quality, assessed by journal, institution and methods used
Theoretical paper
- Theoretical model only applicable to outpatient sector
- No discussion of any kind of PPS
- Models which are not applicable in the German context
- Policy paper lacking a thorough theoretical discussion
Empirical paper
- No empirical result for LOS
- Only descriptive or qualitative assessment
- Analysis of too different payment systems

Table 2: Exclusion Criteria for international review

In the next step, the reference lists of the included full texts were analyzed, which produced 45 additional papers for the analysis. After employing the same exclusion criteria, 17 of them were finally included after full text analysis. Overall, 47 relevant paper on the theoretical and empirical implications of a PPS were identified and included in the analysis. Thirty-six of those provided empirical results for the LOS, whereas eleven provided a thorough theoretical discussion of the effects of the introduction of a PPS.

5.2 Review of the theoretical discussion of prospective payment systems

5.2.1 General remarks

The following chapter will summarize core concepts and key findings of this theoretical discussion. It is noteworthy, that the bulk of the theoretical research was conducted in the U.S. in the late 1980s and early 1990s in the wake of the introduction of Medicare's PPS.

It is important to clarify what is meant by the term PPS. Dowling was one of the first authors to describe and discuss such a system According to him, prospective reimbursement is a method of payment wherein the level of the payments are defined in advance for the coming period and the hospital is reimbursed by the predetermined rates, regardless of the actual costs of treatment. Therefore, prospective payment is an instrument that partially shifts the cost-risk from payers to providers, such as hospitals (Dowling, 1974, p. 163). PPSs are known by an array of different denominations. PPS is most commonly used in the U.S. context. Activity-based funding, case-mix funding and service-based funding are other terms, which describe a system, where reimbursement rates are determined in advance (Duckett et al., 2013, p. 6). In the following, the term PPS will be used synonymously for all of these denominations.

5.2.2 Findings in the literature

In one of the first discussions of the concept of a PPS, Dowling (1974) gave a broad overview of types of possible prospective reimbursement mechanisms, according to the basis of payment, and their respective effects on parameter like costs, efficiency, quality and LOS. Overall, a PPS provides an incentive for hospitals to increase cost-consciousness (Dowling, 1974, p. 164). Considering episode of illness as the payment unit, as a DRG system does, Dowling expects a decrease in number of treated cases, lengths of hospitalizations, intensity of service and overall quality level, but an increase in hospital efficiency (Dowling, 1974, p. 166). Using these findings as a guide, the following discussion aims to summarize relevant literature on the effects of PPS on costs and efficiency, as well as its impacts on the other parameters, with a focus on quality, treatment intensity and LOS.

5.2.2.1 Cost decreasing

A widely known and respected paper in this area is "A theory of yardstick competition" by Shleifer (1985). He started his analysis by stating that in order to set appropriate prices in a regulated industry, the regulator needs an accessible benchmark to evaluate the performance of a firm. According to him, present and past firm values are no suitable candidates. He suggests comparing similar regulated firms with each other and to use the average costs of those firms as a "yardstick" for the level of costs, which have to be attained by the firms. Shleifer uses a framework, where he compares identical firms, operating as monopolists in different markets, and expected to be able to reduce costs at the same rate. The incentive for efficient production lies in the fact that if one of the firms reduces its costs compared to the "yardstick", it profits while the firms that do not lower costs will incur losses. The same intuition lies behind the introduction of Medicare's PPS, where similar hospitals are reimbursed according to average costs induced by treating similar groups of patients (Shleifer, 1985, pp. 319–320). Shleifer concludes, that this kind of “yardstick competition” can generate a situation where the regulator’s and the firm’s preferences for cost reduction coincide (Shleifer, 1985, p. 323), which leads to a socially efficient allocation of resources. The paper identified collusive behavior and premature discharge as potential problems, which can arise when applying such a system in the hospital sector (Shleifer, 1985, pp. 325–327).

Dranove (1987) incorporated local market competition into Shleifer’s model. According to Dranove, PPSs provide another channel to reduce aggregate hospital expenditure, besides through the “yardstick competition”. As hospitals face average costs as prices, they are incentivized to concentrate on areas where their expected costs are smallest in relation to the DRG rate. Extensive specialization and cost savings through specialization should be observable after the introduction of a PPS (Dranove, 1987, pp. 417–418). Dranove further modified Shleifer’s model by incorporating patient characteristics alongside hospital characteristics in the calculation of the treatment costs of a patient within a certain DRG. Dranove found, that a PPS produces the best results if the cost variation within the payment unit, e.g. DRG, is minimized (Dranove, 1987, pp. 426).

5.2.2.2 Impact on other hospital parameters

The first comprehensive theoretical discussion of the effect of a PPS on parameters other than costs was presented by Seidman and Frank in 1985. A novel characteristic of their approach was to allow hospital preferences to not only be dependent on profit maximization, but also on the quality of care provided (Seidman and Frank, 1985, p. 166). They showed that per-diem payment incentivizes longer LOS, whereas per case PPS encourages hospitals to discharge patients sooner, in order to maximize their profits (Seidman and Frank, 1985, p. 164). Another way for hospitals to maximize their profits lies in reducing the cost per patient day, which could be achieved by efficiency improvements.

However, hospitals might protect themselves against the financial risk of the PPS, to not be able to reduce their costs below the “yardstick” through increased efficiency, by employing other cost-saving, but not welfare-improving means (Seidman and Frank, 1985, p. 162). Seidman and Frank found that hospitals might decrease treatment intensity by providing fewer and lower quality services (Seidman and Frank, 1985, p.

165). Besides this undertreatment, other possible actions include trying to attract less costly patients and decline costly patients (cream skimming) and splitting cases by prematurely discharging patients with the purpose of re-admitting and creating a new DRG (Seidman and Frank, 1985, pp. 169–170). Seidman and Frank highly doubted, that per case PPS would produce welfare gains – i.e. cost reductions without infringing on treatment quality (Seidman and Frank, 1985, p. 173).

A paper by Ellis and McGuire (1986) examines how well prospective payment increases efficiency by reducing the quantity of hospital services used and, how to ensure that the level of supply response is desirable. In contrast to Seidman and Frank, where the hospitals utility function is used to explain the effects of PPS, Ellis and McGuire look at the physicians, as main decision-makers who select the level of services to be provided in the hospital, and their incentives. In their model, physicians act as agents for two principals, trading of benefits to the patients against benefits to the hospital (Ellis and McGuire, 1986, pp. 130–131). As prospective payment implies that the hospital bears the entirety of marginal costs and the patient the equivalent of marginal benefits, the physician has to weigh benefits to the hospital and the patient equally, thus acting as a perfect agent, in order to make the most efficient and desirable quantity choice (Ellis and McGuire, 1986, pp. 135–137). However, Ellis and McGuire argue that physicians most likely will value the hospital's financial interests higher, because of the high bargaining power hospitals possess compared to the physicians. If the physician does in fact not act as a perfect agent for the patient, Ellis and McGuire show that prospective payment leads to undersupply of medical needs (Ellis and McGuire, 1986, p. 138).

Hodgkin and McGuire in 1994 formulated a simple model, which combined the models of Ellis and McGuire and Dranove and incorporated treatment intensity, alongside profits in the hospitals objective function. Intensity is defined as the volume and intricacy of resources used for one patient episode, e.g. LOS or technical sophistication (Hodgkin and McGuire, 1994, pp. 3–4). Hodgkin and McGuire found, that a change in hospital reimbursement might affect outcomes through two different channels, possibly working in opposite directions: the marginal level of reimbursement, i.e. the share of reimbursed costs induced by an additional service per patient, and the average reimbursement level. The latter also has a significant effect on hospital decisions

(Hodgkin and McGuire, 1994, pp. 4–6). Another finding of the paper was, that altogether, the introduction of a PPS should have a negative effect on the treatment intensity (Hodgkin and McGuire, 1994, p. 24).

Another important parameter concerning hospital production is treatment quality. There is some literature theoretically discussing the impact of the introduction of a PPS on treatment quality (Pope, 1989; Rogerson, 1993; Ma, 1994). A thorough discussion however, would be not be constructive concerning the original purpose of this analysis. Thus, the findings are shortly summarized in the following:

Pope (1989) examines the influence of increased competition on hospital quality. According to his findings, treatment quality decreases after PPS introduction. Ma (1994) states that demand for hospital care depends on quality and that a PPS only results in societal benefits if hospitals are not permitted to deny treatment to patients. Rogerson (1993) concludes in his paper that the introduction of a PPS will lead to higher treatment quality only if quality does in fact have an influence on the demand.

Although PPS are getting implemented in more and more countries, there is a clear consensus in the literature, that pure PPS and yardstick competition cannot be seen as the optimal payment system for hospital treatment. Mixed systems with different cost sharing mechanisms are believed to lead to more optimal outcomes (Ma, 1994; Ellis and McGuire, 1996; Chalkley and Malcomson, 2002; Levaggi, 2005). It is noted, however, that adequate cost sharing schemes are hard to implement and require intensive monitoring (Chalkley and Malcomson, 2002, pp. 245–246).

5.3 Empirical Studies

5.3.1 Studies in the United States

Whereas the general approach is quite similar across most theoretical studies discussed here, the identified empirical publications on the LOS development after the introduction of a PPS vary greatly concerning their setting, estimation techniques, data sets and results. The following will discuss this briefly in the U.S. context. Subsequent chapters will then present results of international studies findings from Germany.[5]

[5] For detailed information about the settings, data sources, methods and possible limitations of empirical studies, see appendix B.

The review identified ten studies, which quantified the LOS change after the introduction of a PPS in the U.S. context. Two of those studies examined the shift to Medicare's PPS in 1983 and found a first year reduction of 10.3% (Hadley et al., 1989, p. 358) and an overall reduction by 24% (Kahn et al., 1990). However, the research design by Kahn et al. (1990), using a sampled pre-post design, might overestimate the effect, since they do not account for a time trend effect.

Four additional publications try to estimate the effect of a policy change on LOS within particular states in the U.S. By using Difference-in Differences and Fixed Effects methods, they found a variety of results, ranging from no significant effect (Salkever and Steinwachs, 1988; Norton et al., 2002) to a trend change of about -2 % (Hsiao and Dunn, 1987), as well as a reduction of LOS by 3.91 days (Gilman, 2000). The generalizability of these results towards the German context are at best questionable, since the studies were either restricted to very narrow indications (Gilman, 2000; Norton et al., 2002) or the policy change was not as revolutionary as the DRG introduction in the German hospital sector (Hsiao and Dunn, 1987; Salkever and Steinwachs, 1988).

Two additional publications exploited the fact, that hospital payment schemes for Medicaid patients varied substantially between states in the early 1980s and used this as a setting for a quasi-natural experiment (Frank and Lave, 1989; Lave and Frank, 1990). Frank et al. (1989) estimated, that applying a PPS leads to a 17% reduction of LOS, compared to cost based reimbursement for psychiatric patients. Lave and Frank (1990) distinguished between medical, psychological and surgical cases, as well as between public and non-public facilities. They calculated effects between -2% for medical cases in non-public hospitals up to -17.9% for psychological patients in public facilities.

Two remaining studies examined the effect of the implementation of a PPS for Medicare patients in inpatient rehabilitation facilities. Whereas Sood et al. (2008) estimated reductions of LOS for patients with joint replacements (-4.7%), strokes (-10.5%) and hip fractures (-6.0%), McCue and Thompson (2006) observed an increase of LOS in inpatient rehabilitation facilities due to PPS introduction by 1.03 days. However, their simplistic Difference-in-Differences approach did not account for differences in patient characteristics. Figure 8 illustrates the changes of LOS due to a PPS introduction in the U.S. context.

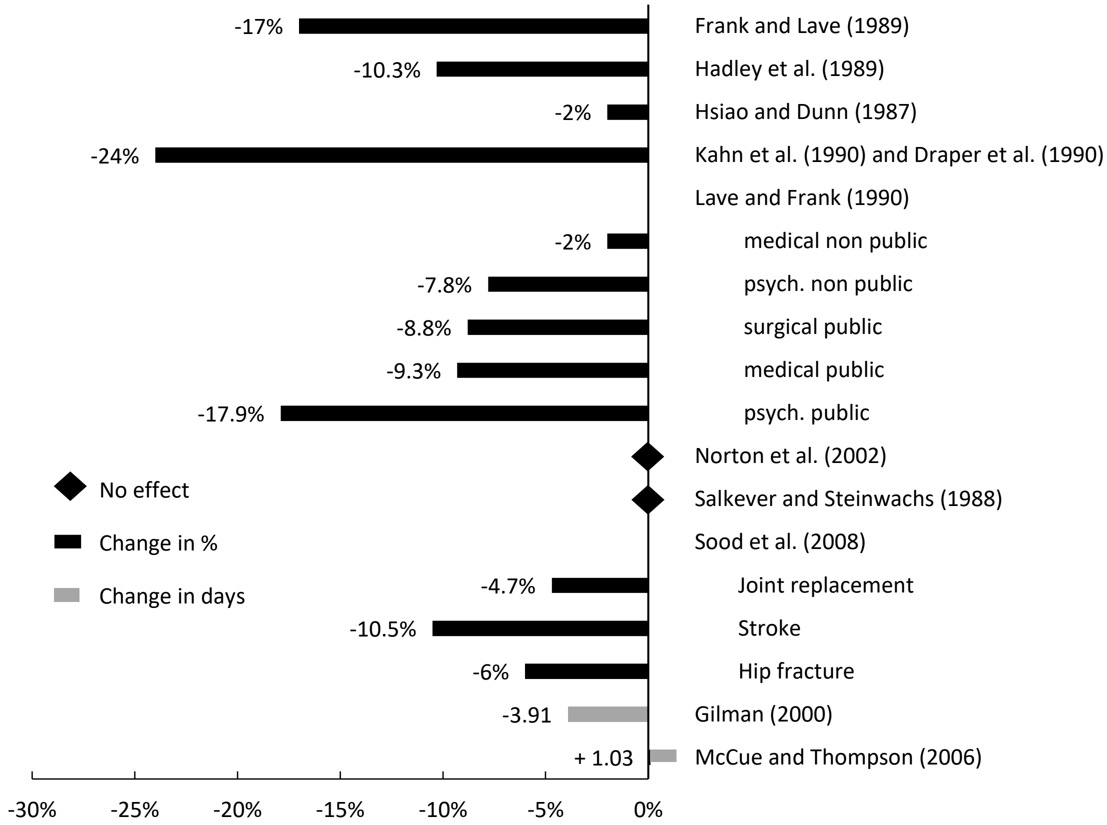

Figure 8: Effects of PPS introduction in the U.S. context in % and days
Most papers provided only estimates in %.

It is noteworthy that most studies in the U.S. examined policy changes in Medicaid and Medicare programs. People eligible for those programs are in general different from the rest of the population, as they are either poor or retired. This selection effect possibly affects the impact a PPS introduction has on LOS.

5.3.2 International studies

In light of the review of the relevant literature, 18 studies were found which quantified the effect of a PPS introduction on LOS in various countries outside of the U.S. and Germany. As was the case for the U.S. studies, their approaches and general results vary significantly. As figure 9 depicts, the estimated effects vary between +0.66 and -3 days.[6]

[6] Appendix B contains key features of the different studies.

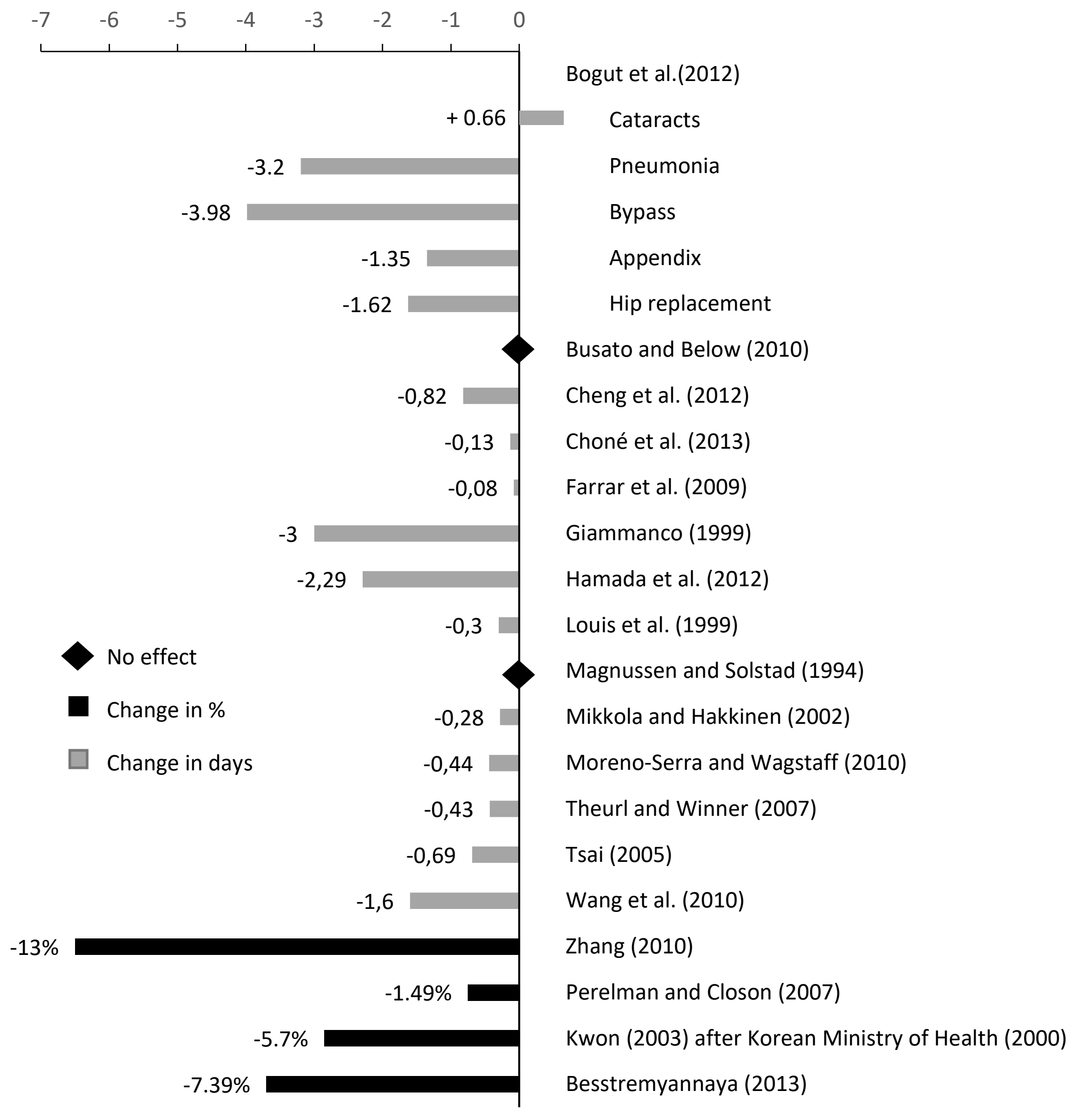

Figure 9: Change in LOS due to PPS in days and %, International studies
Most papers provided estimates only for change in days.

The highest reduction in LOS concerning a general population was observed by Giammanco (1999) in Italy with a three day decline after the policy change. High reductions were also observed in Japan by Hamada et al. (2012), Besstremyannaya (2013) and Wang et al. (2010). It does not come as a surprise, however, that the highest reductions are observed in Japan, where the ALOS by far exceeded other OECD countries (OECD, 2015, p. 109).

One study by Zhang (2010) evaluated a short term PPS introduction in Shanghai and estimated that the reform reduced the LOS by 13%. Yet cost shifting and quality reductions might be a more serious problem in the context of developing countries with a higher uninsured population and weak quality assurance systems.

As in the U.S. context, there are also studies that could not find a significant impact of the introduction of a PPS on the LOS (Magnussen and Solstad, 1994; Busato and Below, 2010). While the study design and the results of Magnussen and Solstad (1994) with a sample of just four Norwegian pilot hospitals is questionable, the results of Busato and Below (2010), who based their estimation on a spatial Difference-in-Differences model using complete hospital discharge data in Switzerland, seem to be quite convincing.

A particularly interesting study was conducted by Theurl and Winner (2007), who analyzed the replacement of the traditional per-diem-based reimbursement by a per case-based reimbursement system in Austria at the turn of the millennium. Germany and Austria are in general quite comparable, a notion which seems to uphold in this context as well, as LOS developed similarly in Austria and Germany until the introductions of the respective payment systems.[7] Theurl and Winner (2007) estimated a fixed effects regression based on aggregated state level data and found a -0.43 day reduction due to the policy change.

5.3.3 Evidence from Germany

As mentioned in chapter 1, the introduction of the DRG system in Germany in 2003 was accompanied by compulsory supporting scientific research, which was supposed to quantify the effects of the reform on a variety of parameters. The official investigation concerning impact on LOS however was superficial and purely descriptive. This also applies to most independent studies of the impact of DRG reform on LOS in Germany, as will be seen in the following.[8]

In an overview of the effects of the reform, Böcking et al. (2005) list several early stage results, including some estimations about the LOS development. However those results were purely descriptive and not published in a scientific journal (Böcking et al., 2005, p. 132). Schreyögg et al. (2005) quantified the LOS reduction a year after the reform with 4 %, but added that a falling trend already existed before the reform and that it was not yet possible to relate that to the DRG introduction (Schreyögg et al., 2005, p. 5). Further descriptive early stage results, also suggest that LOS declined after the reform (Müller,

[7] See figure 3.
[8] There is a lack of evidence on the impact of the DRG reform in Germany in general, which is attributed to limited data availability and access (Schreyögg et al., 2014, p. 117)

2007; Hensen et al., 2008), and that the PPS lead to premature discharges into IRFs (Rutz, 2006). Qualitative assessments of the impact of the reform reached similar conclusions (Buhr and Klinke, 2006; Flintrop, 2006).

The review could identify only one publication, which tried to empirically quantify the effect of the DRG introduction on the ALOS in Germany. The analysis conducted by Hilgers (2011) employed a regression approach based on an unpublished benchmark model developed by Feess et al. in 2007. Anonymized hospital claims data from a small health insurer from 2001-2008 were used for the study, accompanied by data on hospital characteristics from official hospital statistics (Hilgers, 2011, pp. 98–99). The base case results of the analysis quantified the effect of the DRG introduction on the LOS in German hospitals to be a highly significant reduction of 0.45 days. It is also mentioned, that this estimate most likely undervalues the true effect (Hilgers, 2011, pp. 114–115). In general, this effect lies within the range of LOS reductions, observed after PPS implementations in other countries.[9] However, as the functionality of the estimation strategy is not presented in a transparent and thorough fashion, it is difficult to evaluate the overall quality of the findings.

5.4 Summary of general findings in the literature

All in all, the review could identify 36 studies, which tried to quantify the impact of a PPS introduction on LOS. Only one empirical study was found for the German context. Although the settings for PPS introductions varied substantially between the different studies, a significantly negative effect was observed in almost all studies, establishing confidence in expecting a negative effect in this analysis. Predominantly used data sets contained either hospital discharge data, claims data or official (aggregated) hospital statistics. Most common estimation methods used in the analysis were Difference-in-Differences techniques, Fixed-Effects regressions and basic Ordinary Least Squares (OLS).

Nevertheless, many of the studies are confined by too narrow medical indications, small samples and methodological issues. Additionally, the results of many international studies cannot be compared and generalized to the German context, as the systems

[9] See chapter 5.3.1 and 5.3.2.

themselves, the policy change and the pre-reform LOS are too different. One particular exception is the work of Theurl and Winner (2007) for the Austrian context. Their results (-0.43 days) lies close to the result of Hilgers (2011, p. 114), who calculated an effect of -0.45 days for the DRG introduction in Germany. It will be part of the further analysis to establish, whether the magnitude of this reduction can be confirmed.

When reviewing the theoretical literature, following trends and conclusions become apparent. Yardstick competition (Shleifer, 1985) provides strong incentives to increase efficiency and specialization (Dranove, 1987), both possibly resulting in cost reductions. Additionally, it was shown that a PPS also incentivizes hospitals to reduce the quantity of services provided per patient episode (Seidman and Frank, 1985; Ellis and McGuire, 1986). This reduction could either be the result of the intended efficiency gains or the unintended deteriorating of otherwise beneficial hospital services. Whether reimbursement according to a PPS really leads to an undersupply of hospital services depends on whether the physicians act as perfect agents for the patients (Ellis and McGuire, 1986). Hodgkin and McGuire (1994) concluded that the change of reimbursement system not only has an effect through the change of marginal reimbursement, but also through changes in average payment levels. The overall consensus in the literature was that PPS will most likely lead to a reduction in the treatment intensity and therefore a reduction in LOS.

6 Economic Model

6.1 General outline

The following chapter will present an economic model, which tries to explain why replacing a per-diem payment system by a DRG system has a negative effect on LOS, as observed in the literature. The aforesaid model for this analysis is based on a model built by Hodgkin and McGuire (1994, pp. 4-7). Their paper “Payment levels and hospital response to prospective Payment“, examines the effect of the implementation of Medicare’s PPS on the hospital’s choice of treatment intensity. The setting of their model is quite simple, as they only consider a market with one hospital, one payer and one type of discharge. In contrast to Ellis and McGuire (1986), they did not separate between physicians and hospitals and combined their preferences into one utility function. Hospitals objective function is determined by net profits (**π**) and the treatment intensity (***I***).

$$U = U(\pi, I) \tag{1}$$

Whereas profit maximization as a goal for hospitals is straightforward, intensity needs some explanations: Intensity here is seen as the number and complexity of patient care resources, or intermediate outputs, used in producing a patient care episode, like services per admission, technical sophistication or especially LOS. Hospitals are interested in increasing intensity, since it implies additional quality and benefits for the patients. It is also argued, that high intensity signals technological sophistication and therefore adds to the prestige of the hospital (Hodgkin and McGuire, 1994, p. 4).

$$\pi = R - TC \tag{2}$$

Net profits are the sum of patient revenues (***R***) minus the total costs of patient treatment (***TC***).

$$R = p * X \tag{3}$$

Patient revenues themselves are a function of reimbursement per admission (*p*) times the volume of admissions (***X***). In contrast to the original model of Hodgkin and McGuire, ***X*** is exogenous and not a function of treatment intensity. This simplification was conducted as this analysis only examines the effect of a change in marginal reimbursement and not average reimbursement level.

$$p = \alpha + \beta c \tag{4}$$

Reimbursement per admission is the sum of a fixed per case payment (**α**) and a share of actual costs (**β**). In a fully cost based system, **β** would be one and **α** zero. For a per-diem reimbursement scheme one can assume, that after an initial phase of higher costs, any additional day is reimbursed according to the incurred costs (or even more). Therefore **β** is 1 and **α** zero. In a plain PPS however, **β** is zero, as the reimbursement level is independent from the actual expenses of the hospital. The provider payment solely consists of the fixed per case payment α > 0.

$$TC = c * X \tag{5}$$

Total costs are treatment costs per patient episode (***c***) multiplied by the exogenous number of patients treated in a certain period (***X***).

$$c = c(I) \quad c' > 0 \tag{6}$$

Treatment costs per patient episode depend on treatment intensity. Every additional service, day in the hospital or more complex procedure increases the costs per patient episode.
Inserting equations (3) through (7) into equation (2) produces the following profit function for hospitals:

$$\pi = [\alpha + \beta c(I)]X - c(I)X \tag{7}$$

Deriving (7) with respect to intensity produces following equation:

$$\pi_I = Xc'(\beta - 1) \tag{8}$$

The hospital's utility function in this model not only contains profit but also treatment intensity. The derivation of the utility function with respect to intensity therefore is the sum of the partial derivatives and takes the following form:

$$Xc'(\beta - 1) + \frac{U_I}{U_\pi} = 0 \tag{9}$$

The key term of the first-order condition for utility maximization $Xc'(\beta - 1)$ is referred to as the "Moral Hazard Effect" by Hodgkin and McGuire (1994, p. 6). As $\boldsymbol{\beta}$, the share of actual costs reimbursed, only takes on values between zero and one, this expression is never positive. This means that increasing intensity always has a negative effect on the utility and that decreasing intensity has a positive effect on utility. The only exception is when $\beta = 1$, which is the case in any kind of cost based reimbursement system. Then the whole term equals zero and there is no such "Moral Hazard Effect".

Hodgkin and McGuire (1994, p. 6) speak of a "Moral Hazard Effect", as in a PPS, where $\beta = 0$, the hospital faces the incentive to diminish treatment intensity per care episode regardless of the patient's objective need. The presence of asymmetric information prevents payers and patients from evaluating whether the chosen level of intensity is adequate or not. The second term of the first-order condition of the utility function (U_I/U_π) is positive, as long as treatment intensity in itself is valued by the hospital. The magnitude of this positive utility effect depends on the physicians' bargaining power relative to the hospital administration because physicians are assumed to value treatment quality and patient benefits higher than the administration.

One can argue, that the relative bargaining power of the physicians determines the trade-off between profit maximization and setting of an adequate level of treatment intensity (Ellis and McGuire, 1986, pp. 131–132). One can also assume, that for-profit hospitals, which face high pressure from shareholders, are more profit-oriented than non-private hospitals, which still are eligible for cross-subsidies from municipalities (Ärzteblatt, 2016). If there is too much weight on profit maximization, LOS reductions may not represent efficiency gains, but a drop to deficient levels of treatment intensities, infringing on the overall treatment quality.

6.2 Reimbursement change in Germany

As described in chapter 2.1, a per-diem payment scheme was used to reimburse hospitals in Germany before the introduction of the DRG system in 2004. The daily rates were calculated individually and on the basis of historical costs. Every additional day of a hospitalization was reimbursed by the same rate. Reimbursement per admission, in general, could accordingly be described by the following:

$$p_{PD} = \beta c \tag{10}$$

Comprehensive fixed per case payments occurred only in rare cases, letting α drop out of the equation compared to (4). As the cost for an additional day is reimbursed, $\boldsymbol{\beta}$ is assumed to be one, equivalent to no supply side cost sharing. The "Moral Hazard Effect" described by Hodgkin and McGuire (1994, p. 6) vanishes, as $Xc'(\beta - 1)$ converges to zero if $\beta = 1$. Hospitals therefore had no incentive to reduce treatment intensity. Introducing the DRG system changed the incentive structure fundamentally. Reimbursement per admission now was only constituted by one parameter:

$$p_{DRG} = \alpha \tag{11}$$

The cost sharing parameter $\boldsymbol{\beta}$ in a DRG system in general is zero. The payment is solely oriented on the per-case payment $\boldsymbol{\alpha}$. There now is a strong "Moral Hazard Effect", as $\beta - 1 = -1$ and $Xc'(\beta - 1) = -Xc'$. Hospitals subsequently have a strong incentive to reduce the treatment intensity or LOS. Figure 10 depicts the different incentive structures concerning LOS:

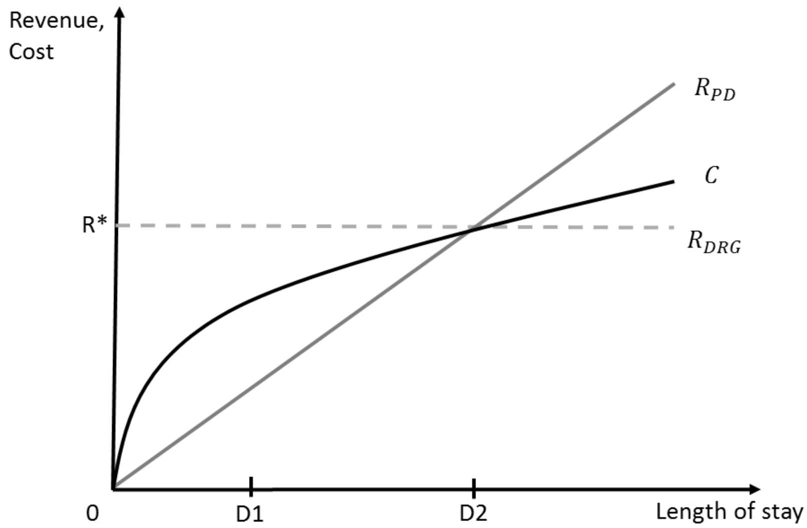

Figure 10: Reimbursement incentives for LOS
Source: Seidman and Frank, 1985, p. 164.

In a per-diem payment scheme the level of reimbursement $\boldsymbol{R_{PD}}$ increases constantly with every additional day. The cost curve is characterized by higher treatment intensity and costs in the first days of hospitalization, and a subsequent reduction of treatment costs per day, eventually falling below the fixed daily rate. After a certain LOS is reached (D2), any additional day bears profits. In a DRG system, the payment level $\boldsymbol{R_{DRG}}$ is fixed, independent of the actual length of hospitalization. The hospital only makes profits if it can discharge the patient before D2 and incurs losses for every additional day after that point.

As demonstrated, LOS reductions not only can be the manifestation of efficiency gains, but also the result of deteriorated treatment quality. This is true if hospitals insufficiently value treatment intensity in itself, which depends on the bargaining power of the quality-oriented physicians working in the hospital. The bargaining power of physicians in German hospitals is limited, because most physicians are employees of the hospital (Porter and Guth, 2012, p. 110). In addition, most leading physicians have contracts, which guarantee them bonuses for reaching certain financially relevant targets in their wards (Blum et al., 2010, p. 36). Besides budget specifications and case-mix standards, LOS targets can be part of these contracts (Klages, 2010, p. 104). Therefore it is highly likely that possible reductions of LOS in German hospitals are not only the results of reduced slack, but also quality reductions.

7 Main hypothesis

Considering the results of the review and the implications of the theoretical model, the following hypothesis are formulated and will be tested in the subsequent statistical analysis:

Hospital LOS response to Change from PD to DRG		
	Funding body	
	Not for profit*	Private for profit
Budget neutral optional year 2003	0	0
Mandatory use of DRG starting in 2004	↓	↓↓
First year of statewide base rates	↓↓	↓↓↓

Table 3: Expected response to reimbursement change
*Private non-profit are assumed to behave similar to public hospitals.

First, the voluntary and budget neutral application of the DRG system in 2003 is expected to have no significant effect on LOS. For one thing, the dissemination of the new payment system was limited in 2003. Only 294 hospitals used the DRG classification throughout the year, and merely half of the German hospitals adopted the system until the end of 2003 (Friedrich et al., 2010, p. 138). Additionally, the actual reimbursement for 2003 was budget neutral, set based on the budget from 2002. Therefore hospitals did not have an incentive to reduce intensity, or LOS.

For the first year of mandatory use, hospitals are expected to respond with only slight reductions of LOS, since compliance was not complete and the reimbursement was still based on hospital specific rates, which did not yet converge towards statewide base rates as a yardstick. As mentioned in chapter 6.1, private for profit hospitals are in general more profit oriented and therefore are expected to respond with higher LOS reductions than their not-for-profit counterparts in general. For 2005, the first year the DRG based yardstick competition truly came into effect (Geissler et al., 2011, p. 250), higher LOS reductions than in 2004 are anticipated.

Overall, it is expected that the following calculations confirm that the DRG introduction lead to a significant decrease of LOS. It will be interesting to see, whether that reduction lies reasonably close to the estimated 0.45 days of Hilgers (2011, p. 114).

8 Data

Researchers predominantly used hospital level data to analyze the impact of a hospital reform on LOS.[10] Routinely gathered administrative insurance claims and hospital discharge data are the basis of most conducted calculations. Cost data and annual hospital survey data were also consulted at times (e.g. McCue and Thompson (2006) and Hadley et al. (1989)). Aggregated country data, respectively state level were only used by Moreno-Serra and Wagstaff (2010) and Theurl and Winner (2007).

According to Grobe (2005, pp. 79–83), there are two main options to obtain comprehensive hospital level data containing both patient and hospital characteristics in Germany. One possible source is routinely gathered claims data from a public health insurance company. Hilgers (2011) was able to obtain such data and used discharge data of all inpatient episodes from a small health insurer for her analysis.

Another feasible data set would be the micro-level data of the German "Krankenhausstatistik". Every German hospital is required by law to provide key-data on their performance and structure to their respective state statistics bureau on a yearly basis. The big advantage of these data would be that they contain the full sample of German hospitals, making sampling considerations obsolete. The hospital level data sets are aggregated by state and federal statistics bureaus to generate publicly available reports on costs, base data and diagnosis data of German hospitals (Destatis, 2016a).

[10] See appendix B.

Unfortunately, it was not possible to gain access to either one of the two mentioned sources of hospital level data.

Therefore, the primary data source for the analysis is the publicly available, state level aggregated figures of the “Krankenhausstatistik”. Appendix D lists the retrieved variables and their sources. Cost data were retrieved from the “Kostennachweis”. Information on numbers of hospitals, number of beds, degree of utilization (in %), occupancy days, medical and non-medical full time equivalents and on the hospitals’ funding bodies were obtained from the “Grunddaten” report. Data on the number of cases, share of female cases, age distribution of the patients and the LOS, were compiled using the diagnosis data of the “Krankenhausstatistik”.

It was possible to access these data on the level of the 20 ICD10 categories the inpatient cases are classified in. The data of the “Krankenhausstatistik” was augmented by state level information on long-term care capacities and the share of the population in need for long-term care, retrieved from the “Pflegestatistik”. The data set was rounded out by including additional state level data on the number of outpatient physicians and available income per household. With the exception of the last two mentioned variables, the entirety of the data set was compiled using adjustable tables, made available by the “Gesundheitsberichterstattung”.

Data on all relevant variables could be retrieved for the years 2000 to 2012. With the mandatory application of the DRG systematic starting in 2004, this implies that four pre-reform and nine post-reform periods can be used in the regression analysis. Additional pre-reform periods would have increased the power of the calculations, but were not feasible to obtain, as some of the variables were not available before the year 2000 and the diagnosis data was classified according to ICD 9 and not ICD10 until 1999 (Destatis, 2013, p. 3).

The raw data set was assembled and presorted using Microsoft Excel 2013. Data on the share of population in need for long-term care and on long-term care capacities were manipulated using the mean imputation method, as the “Pflegestatistik” is generated only every two years. This kind of imputation method possibly entails several problems for the validity of the calculations (Gelman and Hill, 2007, pp. 532–533). Therefore this issue will be addressed in a robustness check. Any further data manipulations, like variable generating, category building or rescaling, as well as the calculations and their visualizations were conducted using STATA 12.

The data set contains exclusively aggregated data based on the full German sample, making sampling considerations obsolete. Moreover, "Krankenhausstatistik" and "Pflegestatistik" undergo extensive test procedures to ensure the validity of the data (Destatis, 2016a, 2016b). Therefore, the overall confidence in the quality of the assembled data set is high.

9 Methods

9.1 General Approach

Information on the ALOS in German hospitals for the 20 ICD-GM categories, aggregated on the 16 German states for the years 2000 to 2012 is used to evaluate whether the introduction of the DRG system in Germany did have a causal negative effect on the LOS in German hospitals. Differentiating between the ICD10 categories, instead of just using the state aggregates, has the advantage that the number of observations is expanded, and that the existing differences between disease groups concerning LOS and its impacting factors will not explicitly have to be accounted for in the regression analysis. The same approach has already been applied by Theurl and Winner (2007) in the Austrian context.

Altogether, the analysis is based on a short, balanced panel with 4160 observations (20 ICDs, 16 states, 12 time periods).[11] The general estimation strategy is based on a classical treatment effect approach, firstly used in economics for the evaluation of subsidized training programs (for example by Ashenfelter, 1978). The treatment effect of an exogenous shock, like a policy change, is measured by including a dummy variable into the regression. This variable is set as equal to one if the observation underwent treatment and zero otherwise. Thus a dummy variable of the following form will be included in the subsequent regressions:

$$t = \begin{cases} 1 \; if \; year >= 2004 \\ 0 \quad otherwise \end{cases} \tag{12}$$

[11] See Wooldridge (2009, pp. 10-12; 466) and Cameron and Trivedi (2005, p. 700) for definitions.

The coefficient of the dummy variable $\boldsymbol{t}$ will then quantify the causal effect of the DRG introduction on LOS.

9.2 Dependent variable

The dependent variable of this analysis is the ALOS, measured in days, for the 20 ICD-GM groups in the 16 German states.[12] Figure 11 illustrates the LOS development in the full sample for the thirteen time periods in the data set. As seen in chapter 3.2, the LOS in German hospitals has declined significantly in the 2000s.

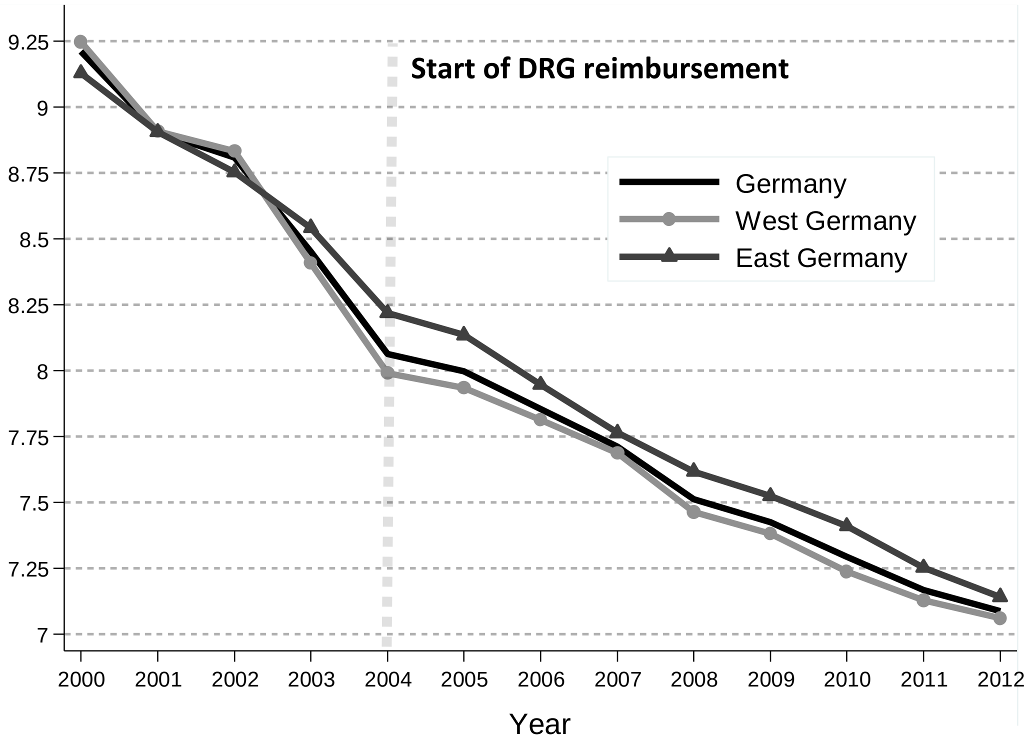

Figure 11: Development of LOS in sample

The ALOS decreased from 9.21 days in 2000 to 8.45 days in 2003. The reduction of 0.39 days to 8.06 days in the reform year of 2004 appears to be greater than the reduction seen in other years. After 2004, LOS declined at a constant rate to 7.09 days in 2012. LOS in the eastern part of Germany, without Berlin, tends to be slightly higher in the last years. Figure 12 plots the differences between the German states for all observations between 2000 and 2012. Figure 13 is the equivalent for the 20 ICD10-GM disease groups in the data set. Although not very pronounced, there is some variation between the German states concerning LOS. Especially the range of the bulk of LOS observations alters from state to state. As could have been expected, the variation between the

[12] Appendix E lists the used ICD10-GM groups.

disease groups are vast with peaks for psychiatric (F00-F99) and perinatal cases (P00-P96).

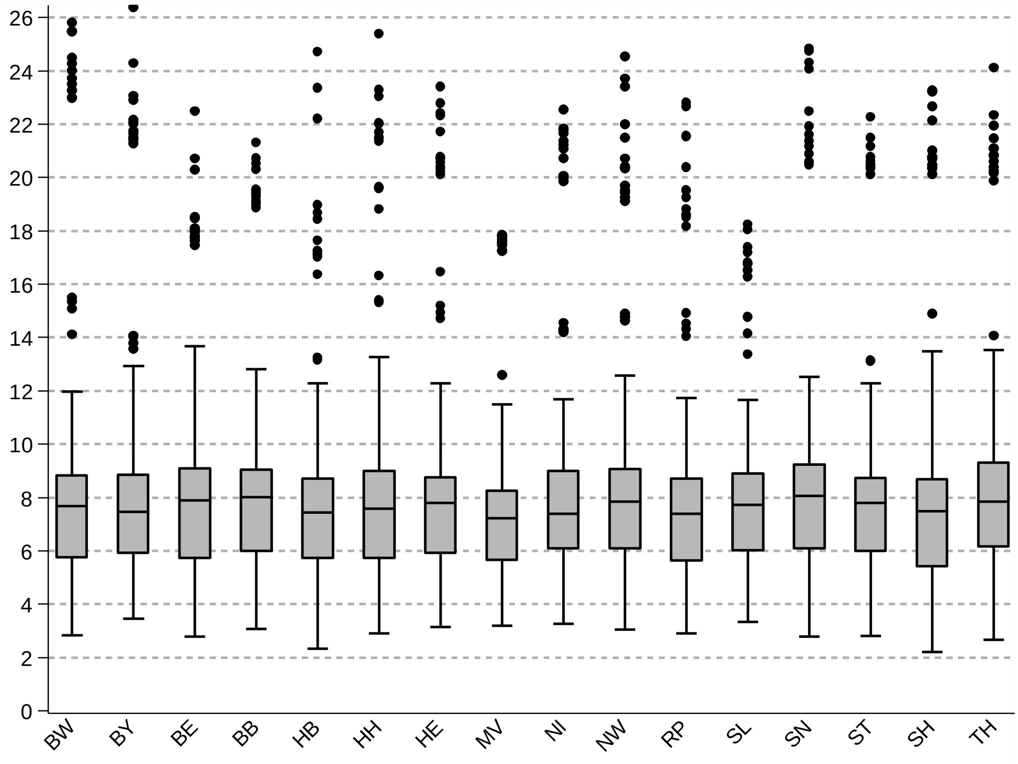

Figure 12: ALOS for the 16 German states
*A list for the abbreviations can be found in appendix E.

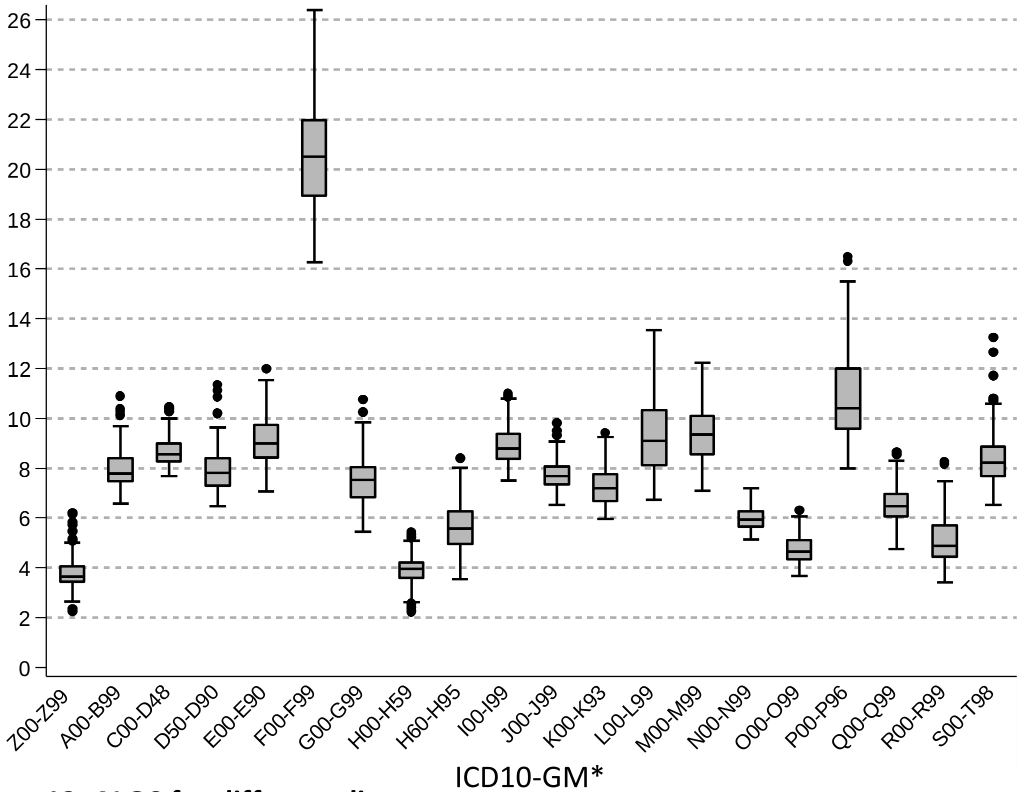

Figure 13: ALOS for different disease groups
*A list of the full denominations can be found in appendix E.

When looking at histogram and kernel density plot of the LOS observations in figure 14, it becomes apparent that the distribution of the data points is positively skewed with a long right tail and a maximum of mass points around 8 days.

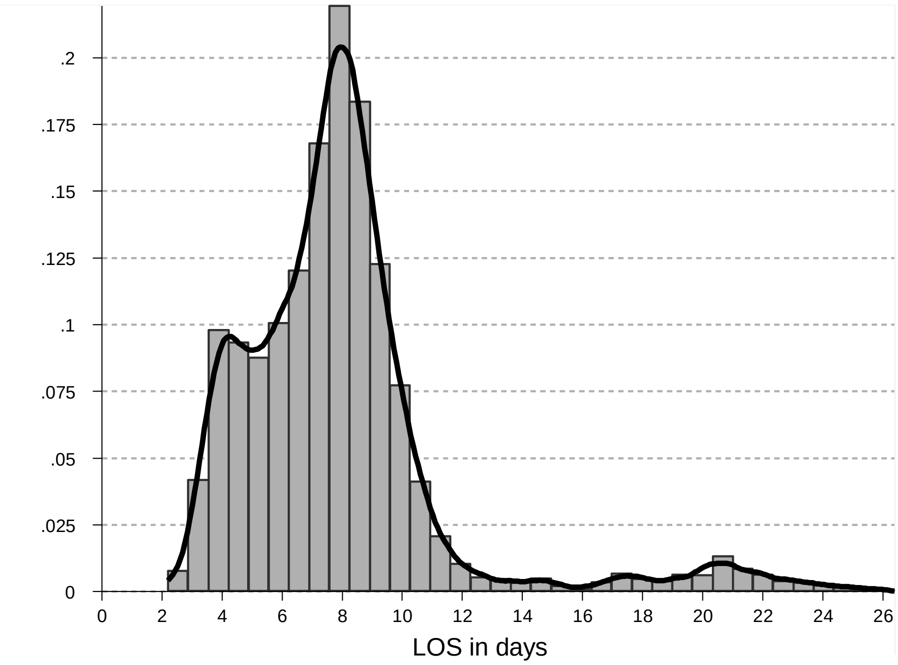

Figure 14: Distribution of LOS observations

The overall mean of LOS in the data set is 7.96 with a standard deviation of 3.61 days and a variance of 13.06. First quartile, median and third quartile lie at 5.84, 7.65 and 8.9 days, respectively. The data points range from 2.20 to 26.38 days. With skewness and kurtosis factors of 2.15 and 9.21, the distribution is in fact highly skewed and leptokurtic, significantly deviating from a normal distribution.[13] Nonnegative, positively skewed data concentrated on small values are symptomatic for count data (Cameron and Trivedi, 2001, pp. 331–332). The LOS observations in the data set are strictly speaking no classical count data, as they are disease group specific averages taking on non-integer values. Yet, as the underlying individual durations of hospitalizations are true count data and the averages seem to have maintained their characteristics, the following analysis will treat the LOS observations as counts.

9.3 Model selection

The use of basic linear estimation methods like ordinary least squares (OLS) would be problematic in this context for several reasons. As count data are intrinsically

[13]See Bulmer (1979, pp. 61–65) for definitions of kurtosis and skewness

heteroscedastic, since the variance is by definition a function of the regressors (Cameron and Trivedi, 2001, pp. 331-332), one of the Gauss-Markov assumptions is violated and OLS can lead to biased and inconsistent estimates (Wooldridge, 2009, p. 94; 102). Moreover, statistical inference might be erroneous, as the normality assumption needed for the validity of the t-statistics is often violated in count data (Wooldridge, 2009, pp. 117–118). Running Shapiro–Wilk and Shapiro–Francia tests for normality confirms non-normality of the LOS observations in the data (z-statistics = 16.043). Another argument against the use of OLS is that it is not restricted to the nonnegative-realm and might calculate negative predictions, which constitute no plausible values for count data.

Therefore a different kind of model has to be selected. According to Cameron and Trivedi (2005, p. 666), the Poisson model is the starting point for count data. Poisson regression is a nonlinear estimation technique which belongs to the class of the extremum estimators and is calculated via maximum-likelihood (Cameron and Trivedi, 2005, pp. 117–119; 124-125). The Poisson distribution has the following moments: $E[y] = \lambda$ and $V[y] = \lambda$. This equality of variance and mean is also called equidispersion. The Poisson regression model is usually specified as

$$\lambda = exp(X'\beta). \qquad (13)$$

As long as the conditional density of the dependent variable is correctly specified, Poisson maximum-likelihood consistently calculates the parameter vector **β** (Cameron and Trivedi, 2005, p. 121). The use of Poisson models for count data is often deficient, as the equidispersion property is usually not met. However, both deviance and Pearson statistics, as well as a likelihood-ratio test on the overdispersion parameter α did not lead to significant results hinting towards equidispersion.

Further considerations had to be made concerning the panel character of the data set at hand. General approaches for handling panel data are pooling cross sections over time and the use of fixed or random effects (Wooldridge, 2009, p. 445; 481). Both Hausman test comparing fixed effects and pooling the data, and Breusch-Pagan test comparing random effects and pooling the data lead to the rejection of the null hypothesis. Therefore using the pooled data set is not appropriate. Another Hausman test was conducted to determine whether fixed or random effects should be incorporated in the model. The null hypothesis that the differences in the fixed effects coefficients are not systematic was rejected and fixed effects will be used in the base model. Ignoring the existence of fixed effects would lead to inconsistent parameter estimates of pooled and random effects (Cameron and Trivedi, 2015, p. 244).[14]

Data on the LOS of hospitalizations are only observed for patients which had a LOS of at least one day. Therefore values for LOS can only range from one to infinity. This is called zero-truncation. As the basic Poisson model fits the data over the whole range off the response variable including zero, not accounting for truncation leads to inconsistent estimates (Cameron and Trivedi, 2005, pp. 679–680). Thus, the final base model of the following calculations is a zero-truncated Poisson fixed effects model, in which the fixed effects are included via 320 dummy variables for each individual disease category for a specific state. In contrast to the basic Poisson model, where the first and second moment are equal and solely defined by the rate parameter $\boldsymbol{\lambda}$ (13), mean and variance of the zero truncated Poisson fixed effects take following diverging forms:[15]

$$E[y_{it}|\, y_{it} > 0] = \frac{\alpha_i \lambda_{it}}{1 - e^{-\alpha_i \lambda_{it}}} \tag{14}$$

The expected value E of y, the LOS, for disease group $\boldsymbol{i}$ in year $\boldsymbol{t}$ conditioned on a non-zero LOS is now a function of the rate parameter$\boldsymbol{\lambda}$ and the individual specific effect $\boldsymbol{\alpha}$. The latter represents the 320 included time invariant disease group specific fixed effects.

$$V[y_{it}|\, y_{it} > 0] = \frac{\alpha_i \lambda_{it}}{1 - e^{-\alpha_i \lambda_{it}}} [1 - \frac{\lambda_i e^{-\alpha_i \lambda_{it}}}{1 - e^{-\alpha_i \lambda_{it}}}] \tag{15}$$

The conditional variance $\boldsymbol{V}$ of the LOS $\boldsymbol{y}$ clearly is not equal to the conditional mean, revoking the equidispersion characteristic of the basic Poisson model. This deviation stems from accounting for the truncation of $\boldsymbol{y}$.

[14] Running the Hausman tests was only feasible on the unclustered base model.

[15] Derivation of the moments, probability density function and likelihood functions was based on Cameron and Trivedi (1998, pp. 279–283); Cameron and Trivedi (1998, pp. 117–120); Winkelmann (2008, pp. 30–31) and Yeh et al. (2012, pp. 254–256).

The consequential conditional density function for disease group $\boldsymbol{i}$ in year $\boldsymbol{t}$ is given by:

$$\mathrm{P}(Y_{it} = y_{it} | \, Y_{it} > 0, \alpha_i, \lambda_{it}) = \frac{\alpha_i {\lambda_{it}}^{y_{it}}}{e^{\alpha_i \lambda_{it}} (1 - e^{-\alpha_i \lambda_{it}}) y_{it}!} \tag{16}$$

This leads to the following likelihood function, where the conditional density function has to be multiplied out over both unit and time dimensions $\boldsymbol{i}$ and $\boldsymbol{t}$:

$$L(\beta, \alpha) = \prod_{i=1}^{n} \prod_{t=1}^{T} \left[\frac{\alpha_i {\lambda_{it}}^{y_{it}}}{e^{\alpha_i \lambda_{it}} (1 - e^{-\alpha_i \lambda_{it}}) y_{it}!} \right] \tag{17}$$

The subsequent log-likelihood function, which has to be maximized by fitting parameter vectors **β** and $\boldsymbol{\alpha}$, takes the following form:

$$l(\beta, \alpha) = \sum_{i=1}^{n} \left[-\alpha_i \sum_{t=1}^{T} \lambda_{it} + ln\, \alpha_i \sum_{t=1}^{T} y_{it} + \sum_{t=1}^{T} y_{it} \ln \lambda_{it} - \sum_{t=1}^{T} \ln\,(1 - e^{\alpha_i \lambda_{it}}) \right. \\ \left. - \sum_{t=1}^{T} ln\, y_{it} \right] \text{ with } \lambda = exp(X'\beta) \tag{18}$$

The approach of including a dummy variable for each unit of observation is not computable for very large $\boldsymbol{n}$ as software packages often restrict the permissible number of variables. In this case, however, computation was feasible. Cameron and Trivedi (1998, pp. 281–282) additionally showed that there is no incidental parameter problem arising for the Poisson fixed effects model by including dummy variables for each individual. In other panel MLE models, a small $\boldsymbol{t}$ and a number of parameters going to infinity for large $\boldsymbol{n}$, would lead to inconsistent estimates.

Including fixed effects in the model does not fully account for the inter cluster correlation of the error terms between disease groups and states. Therefore, cluster-robust standard errors on the 310 yearly observations were calculated to enable valid statistical inference (Cameron and Miller, 2015, pp. 330–331).

9.4 Covariates

According to Lave and Frank (1990, pp. 332–334) and Ellis and McGuire (1996, p. 263), LOS is a function of patient and hospital (sector) characteristics but also depends on characteristics of the non-hospital health care system and the design of the reimbursement mechanism. Thus, the rate parameter λ of (13) can be schematically portrayed as the exponential function of the successive expression:

$$X_{it}'\beta = (\underbrace{\gamma * P_{it}}_{\text{Patient char.}} + \underbrace{\theta * H_{it}}_{\text{Hospital sector char.}} + \underbrace{\mu * D_{it}}_{\text{Health care delivery system}} + \underbrace{\delta * t_t)}_{\text{payment change 2004}} \quad (19)$$

Included patient characteristics are the share of female cases and the shares of different age groups in a disease group, as well as the average income per household member on the state level. Gender and age, the only variables besides LOS, which were available on the level of the disease groups, were added to the model as indicators for treatment intensity, which varies according to gender and increases with age. Treatment intensity is mainly influenced by the disease severity (Gilman, 2000, pp. 281–282) and the health status of the patient, which is expected to be the main driver of LOS (Lave and Frank, 1990, pp. 332–333). This is partly accounted for by regressing on the different disease categories. Another incorporated variable trying to capture the effect of case severity is costs per case on the state level. Income was included in the analysis as a proxy for social deprivation, which was found to vastly influence the LOS of patients in the French context (Yilmaz and Raynaud, 2013, pp. 249–250).[16]

Characteristics from the Inpatient sector included in the regression are for one cases and hospital beds per 100,000 inhabitants, as indicators for the demand for and supply of hospital services. In combination with the degree of bed utilization, these variables are supposed to pick-up the effect of the varying intensity of hospital competition, a factor, discovered to affect LOS in hospitals in a study by Choné et al. (2013, p. 18). Coming from the theoretical model in chapter 6, one can argue that a higher level of competition complicates making profits as the treatment quality has to be maintained on a higher level to attract patients. To compensate the additional costs, associated with

[16] This is an interesting finding, since it confirms the existence of altruism in hospitals and the assumption of the theoretical model, that hospitals not only derive utility from profits, but also from higher treatment intensity itself.

higher quality, hospital efforts to reduce LOS and thereby costs are higher in a competitive environment.[17] The share of private for profit hospitals was also included in the model since such hospitals might put a higher weight on the profits π in (1) and try to decrease LOS to a lower level. This connection has already been examined by Lindrooth and Weisbrod (2007) for hospice care in the U.S. Completing the list of inpatient covariates are the quantities of medical and non-medical personnel per 100,000 hospital cases since the number of involved caretakers might have an impact on the duration of the recovery or at least on the time span of the regular hospital production process from admitting a patient to conducting medical tests and procedures and the processing of the final discharge.

Another important category of variables according to Lave and Frank (1990, pp. 332–334) consists of non-hospital health care delivery system characteristics. These should have an effect on LOS since the existence of capable outpatient and long-term care facilities simplifies earlier hospital discharge. The base model of this analysis only includes the number of outpatient physicians per 100,000 inhabitants.[18] The final and most important covariate for this analysis is the treatment dummy variable t, introduced in chapter 9.1, indicating the change of payment system from per-diem to per case payment. A link test was performed on the base model, with a non-significant result, indicating that the inclusion of squared terms or interactions would not improve the model fit.

10 Results

10.1 Descriptive statistics

Table 4 on the following page provides descriptive statistics for the variables included in the base model. Column one lists if the variable at hand was available on the level of the disease groups. Column two and three contain the mean and standard deviation of the variables for the pre-reform periods and for the years following the mandatory use of the DRG systematic starting in 2004. As seen in chapter 9.2, LOS declined comparing pre- and post-reform periods. The share of patients in all age groups over 65 increased and

[17] However, hospitals need to prevent decreasing LOS to a level, where it infringes on treatment quality.
[18] Data on inpatient rehabilitation was not available for all states and long-term care data is only collected every two years. The latter was first included in the model but did not improve the model fit.

is 26% for the DRG reimbursement period. Income and costs per case both increased noticeable, likely a manifestation of overall economic growth and inflation. The degree of bed utilization and the number of beds declined simultaneously. Together with an increasing number of cases this is a first hint towards increases in productivity and possible LOS reductions.

	Available per disease group	From 2000 to 2003	From 2004 to 2010
LOS in days	Yes	8.84	7.57
		(3.89)	(3.41)
Share of inpatients younger than 65	Yes	0.69	0.64
		(0.18)	(0.20)
Share of inpatients between 65 and 75	Yes	0.15	0.16
		(0.08)	(0.09)
Share of inpatients between 75 and 85	Yes	0.12	0.14
		(0.08)	(0.09)
Share of inpatients 85 and older	Yes	0.05	0.06
		(0.04)	(0.04)
Share of female inpatients	Yes	0.55	0.54
		(0.12)	(0.12)
Income per household member in €1000	No	15.75	17.87
		(1.83)	(2.11)
Costs per case in €1000	No	3.14	3.60
		(0.38)	(0.37)
No. of beds per 100,000 inhabitants	No	677	639
		(78.7)	(71.1)
No. of cases per 100,000 inhabitants	No	21,349	22,514
		(2,578)	(2,898)
Share of for-profit hospital beds	No	0.09	0.20
		(0.08)	(0.15)
Non-medical personnel per 100,000 cases	No	4,162	3,763
		(379.0)	(294.2)
Medical personnel per 100,000 cases	No	656	728
		(98.3)	(82.7)
Bed utilization in %	No	80.56	78.02
		(2.52)	(3.05)
Outpatient physicians per 100,000 inhabitants	No	158	167
		(18.9)	(20.7)
N		1,280	2,880

Table 4: Descriptive statistics

Means with standard deviation in parenthesis; inpatient personnel in full time equivalents.

The strong increase in costs per case and cases per 100,000 inhabitants might hint towards rising morbidity and case severity as a result of the demographic change (see for example Oberender and Zerth, 2010, pp. 115–119). However, the increased number of cases cannot entirely be explained by increased morbidity in the population, but might be the product of hospitals responding to price changes (Augurzky et al., 2012, pp. 3–4). Opposing trends can be observed for the quantity of medical and non-medical personnel employed in German hospitals. Whereas the number of physicians per case has slightly increased, the quantity of nursing and other supporting staff in relation to the number of hospital cases has decreased by almost 10%. This might be a result of the DRG induced efficiency efforts depicted in chapter 5.2.2.1. The most striking difference between pre- and post-reform periods can be observed for the share of for-profit hospital beds, which more than doubled.

10.2 Regression results

The review in chapter 5 has shown that the vast majority of empirical findings suggest a negative effect of the change to a PPS on LOS in hospitals. Results varied greatly and ranged up to a reduction of 25% or four days. The only relevant study from the German context quantified the effect with -0.45 days (Hilgers, 2011, p. 114). Column (1) of Table 5 shows the average marginal effects of the base model, which was described in chapter 9.[19] State specific LOS values for the 20 ICD chapters were regressed on the covariates outlined in chapter 9.4. The coefficient of the treatment dummy variable is significant on the 5% level (*p*-value: 0.015), and the average marginal effect has a value of -0.172, indicating that the introduction of the G-DRG system has led to a decrease of LOS of about one sixth of a day. This confirms the tendency of the estimated effect of Hilgers (2011), but not the magnitude, which is nearly three times as big.

As expected, treating an older patient population is connected with an increased LOS. However, this only holds for the age categories over 75. If the share of patients over 85 increases by one percent, the ALOS increases by 0.05 days. The share of female cases apparently does not play a role, whereas income does have a large and significant effect on the LOS. Higher incomes as a proxy for social affluence seems to have the expected negative effect on LOS. As income was only observable on the state level, it is highly

[19] Several variables were rescaled in comparison to table 4 to facilitate interpretation.

likely that this variable captures the effect of a wide range of different potential socio-economic factors. Higher costs per case on the state level have a negative impact on LOS, possibly through the channel of higher technical sophistication. The degree of bed utilization and the number of beds relative to the size of the state's population both have an increasing effect on LOS, contradicting the hypothesis that increased supply and therefore competition leads to lower LOS. However, the causality could also be reversed for the degree of bed utilization, as hospitals with an exogenously sicker patient population have higher ALOS and as a result need fewer patients to have a high degree of bed utilization. The number of cases per 100,000 inhabitants have negative effect on LOS. The proportion of non-medical personal and the number of hospital beds relative to the population seems to positively impact the LOS.

However, this could also be the result of differences in the patient population. Patients with higher treatment intensity require additional hospital staff and are also more likely to have longer hospital stays. The included outpatient sector variable physicians per population has a surprising positive effect on the LOS, but as before, this relationship could be influenced by patient characteristics. The share of private for profit hospitals, which has increased in the period under review,[20] did not have the expected negative effect on LOS and an additional non-significant Wald test on the variable substantiates this finding.

[20] See table 4.

	(1) Base model	(2) Treat 2005	(3) Treat 2003
Treatment dummy	-0.172**	-0.0378	-0.0402
	(0.0706)	(0.0512)	(0.0533)
Share 65 to 75 in %	0.0128	0.0124	0.0136
	(0.0133)	(0.0133)	(0.0137)
Share 75 to 85 in %	0.0236*	0.0211	0.0190
	(0.0134)	(0.0135)	(1.321)
Share over 85 in %	0.0539***	0.0550***	0.0587***
	(0.0177)	(0.0176)	(0.0175)
Share female cases in %	-0.00662	-0.00545	-0.00557
	(0.0122)	(0.0122)	(0.0123)
Income HH member in €1,000	-0.169***	-0.167***	-0.164***
	(0.0413)	(0.0414)	(0.0400)
Costs per case in €1,000	-0.344***	-0.383***	-0.388***
	(0.131)	(0.128)	(0.128)
Beds per 100k P in 100	0.640***	0.648***	0.676***
	(0.123)	(0.130)	(0.00121)
Cases per 100k P in 1,000	-0.187***	-0.171***	-0.177***
	(0.0498)	(0.0499)	(0.0499)
Share for profit hospitals in %	0.00225	0.00189	0.00149
	(0.00328)	(0.00331)	(0.00326)
Non med per 100k cases in 100	0.0563***	0.0699***	0.0662***
	(0.0156)	(0.0144)	(0.0158)
Med per 100k cases in 100	-0.0606	-0.151*	-0.125
	(0.0933)	(0.0814)	(0.101)
Bed utilization in %	0.0482***	0.0526***	0.0565***
	(0.00911)	(0.0106)	(0.00883)
Outpatient phys. per 100k P	0.0137**	0.0155**	0.0156***
	(0.00617)	(0.00602)	(0.00602)
N	4,160	4,160	4,160
R^2 /pseudo R^2	0.2614	0.2613	0.2613
AIC	16,131.9	16,132.3	16,132.3
BIC	16,220.5	16,220.9	16,220.9

Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; Average marginal effects

Table 5: Marginal effects of base model and reform years 2003 and 2005

Putting these first results into context of the hypothesis formulated in chapter 7, one can conclude that the DRG introduction in Germany had the expected but small negative effect on the LOS, and that it could not be confirmed that the type of hospital funding body does influence the LOS development.[21]

The pseudo R-squared value of 0.2614 for the base model is a first indicator that the model has explanatory power. To underpin this notion, figure 15 contains the plotted density of the observed LOS, and the one predicted by the base model.

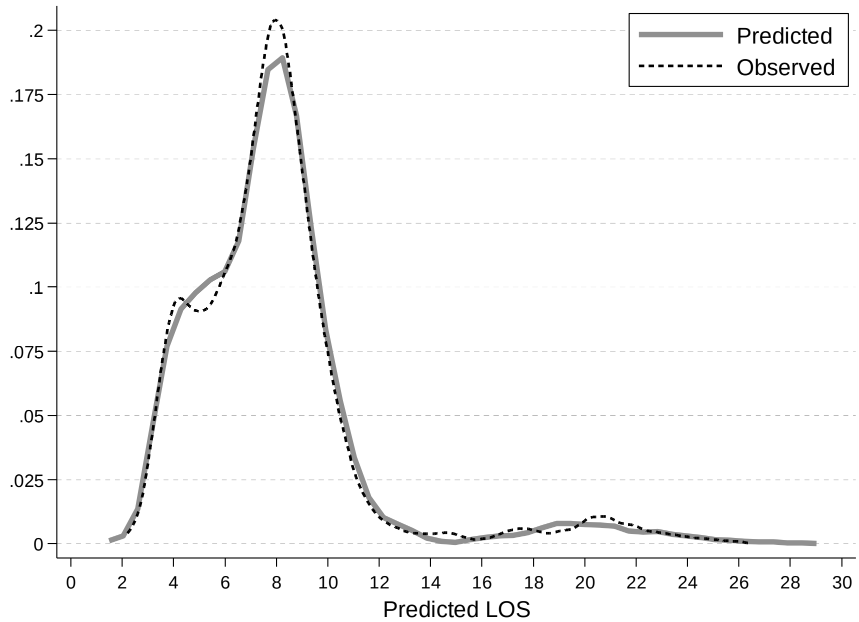

Figure 15: Predicted and observed LOS

It seems like the chosen model configuration fits the data quite well. With the exception of the slight deviations for the two spikes, the density of the predicted values follows the density of the observed LOS values. In contrast to a linear model, the Poisson specification is able to pick up the long right tail of the LOS distribution. Therefore, there is some confidence that the chosen model and its distribution are adequate for examining the LOS development in the data set, and that the estimates are consistently estimated.[22]

To address the hypothesis about the timing of the effect of the DRG introduction, two additional models were run, in which the treatment dummy was first altered to capture

[21] Running a linear and a not truncated specification reveals only small deviations from the base model, quantifying the reform effect with significant -0.197 and -0.168 days.
[22] See chapter 9.3 for the connection between distribution and consistency.

the effect of the voluntary use of DRGs in 2003. The treatment dummy in the second model tries to apprehend the effect of the first year, in which the DRGs were effectively impacting the hospital budgets (2005), as the hospital specific base rates started to converge towards the statewide base rates.[23] The average marginal effects of these two models are presented in column (2) and (3) of table 5. The coefficient of the treatment dummy in both models, depicted in column two and three, does not reach significance. This shows on the one hand that the hypothesis about no introduction effect in 2003 can be confirmed. Hospitals did not change their treatment practices in the year of voluntary use. On the other hand, the hypothesis about higher effects of the PPS introduction in 2005 have to be rejected.

In order to go beyond the mere consideration of the average marginal effect and to get a more differentiated view on the effect of the DRG introduction, figures 16 and 17 illustrate the treatment effect of the base model for the different states and disease categories.

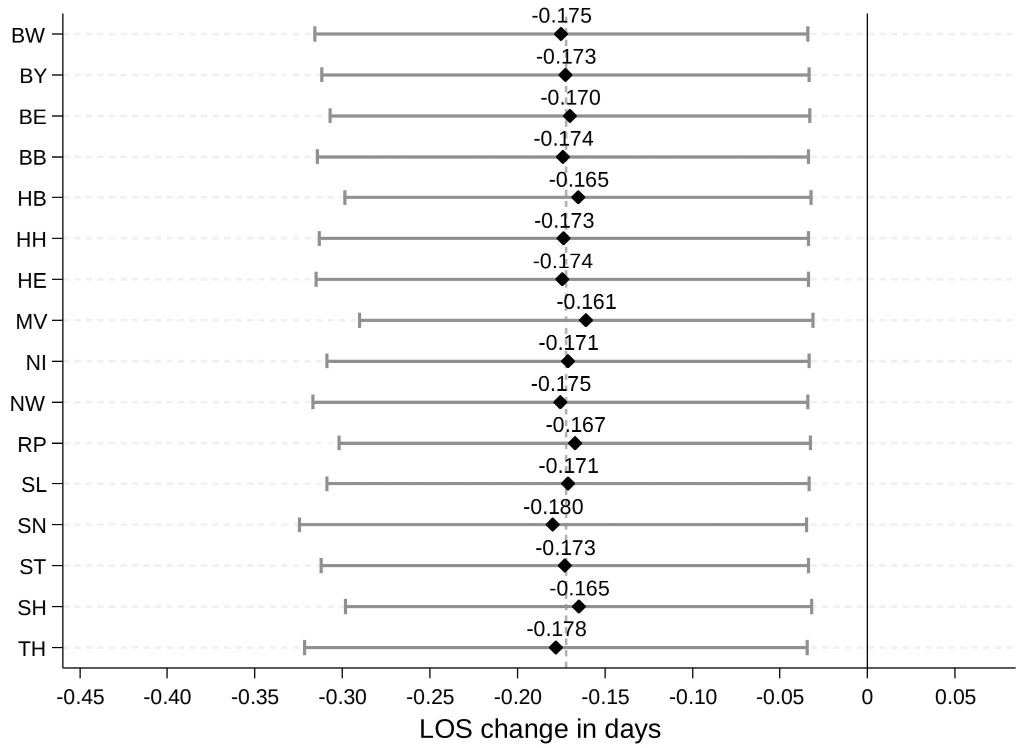

Figure 16: Marginal effects of DRG introduction for states
95% confidence intervals; *A list of the full denominations can be found in appendix E.

[23] See chapter 2.

By examining figure 16, it becomes apparent that the marginal effects did not vary to a great extent between the 16 German states and range close to the average marginal effect of -0.172 days. As the 95% confidence intervals indicate, the effect was significantly smaller than zero for all states.

A look at the marginal effects of the different disease groups in figure 17 shows that the reductions vary greatly between disease groups. As the overall levels of LOS in the different ICD chapters depicted in figure 13 are quite heterogeneous, these kind of results could have been expected.

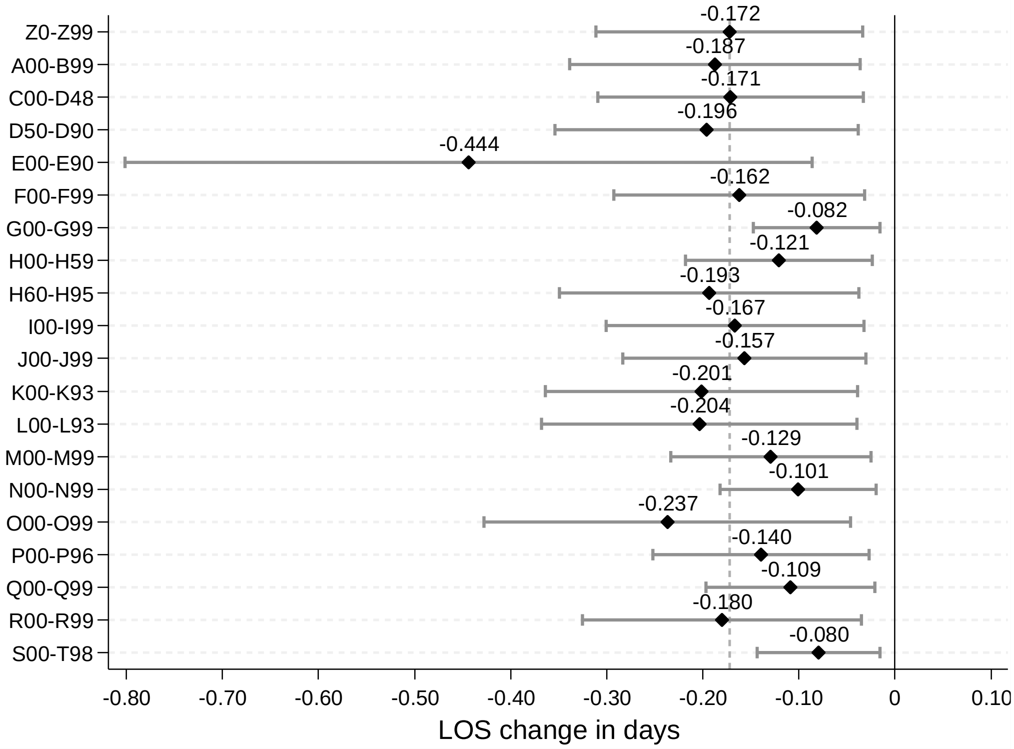

Figure 17: Marginal effects for disease groups
95% confidence intervals; *A list of the full denominations can be found in appendix E.

The biggest treatment effect can be observed for the ICD chapter E, which contains pregnancy and birth related hospitalizations. Hospitals reduced the LOS for such cases almost by half a day as a reaction of the changed incentive structure. However, the wide confidence interval indicates a fair amount of uncertainty concerning the magnitude of the impact. In other disease categories like G00-G99 (neurological cases), the effect was much smaller and barely significant.

10.3 Model alterations and robustness checks

To test the robustness and validity of the general result that the PPS introduction in Germany has led to a small but significant decrease of LOS, several model alterations were conducted. The marginal effects of the respective treatment dummies are depicted in figure 18.[24] First, following Theurl and Winner (2007, p. 382), yearly time trend dummies were included in the regression to approximate the effect of the technological change, which is believed to be an influential determinant of the LOS development.[25] The estimated marginal effect is more than four times as large as in the base model, but because of the huge standard error, the estimated coefficient is not significant anymore (*p*-value: 0.17). It is unclear, why the inclusion of the time dummies has inflated the standard errors to such an extent. As the year dummies possibly capture a lot of different effects, and as their inclusion did not improve the model fit (AIC 16152.9 vs. 16131.9), they were not part of the base model.

The second robustness check excluded the three German city states as they inherently differ from their urban counterparts, resulting in a just no longer significant estimate of -0.123 (*p*-value: 0.13).

[24] The complete regression tables can be found in appendix E.
[25] See for example Sloan and Valvona (1986).

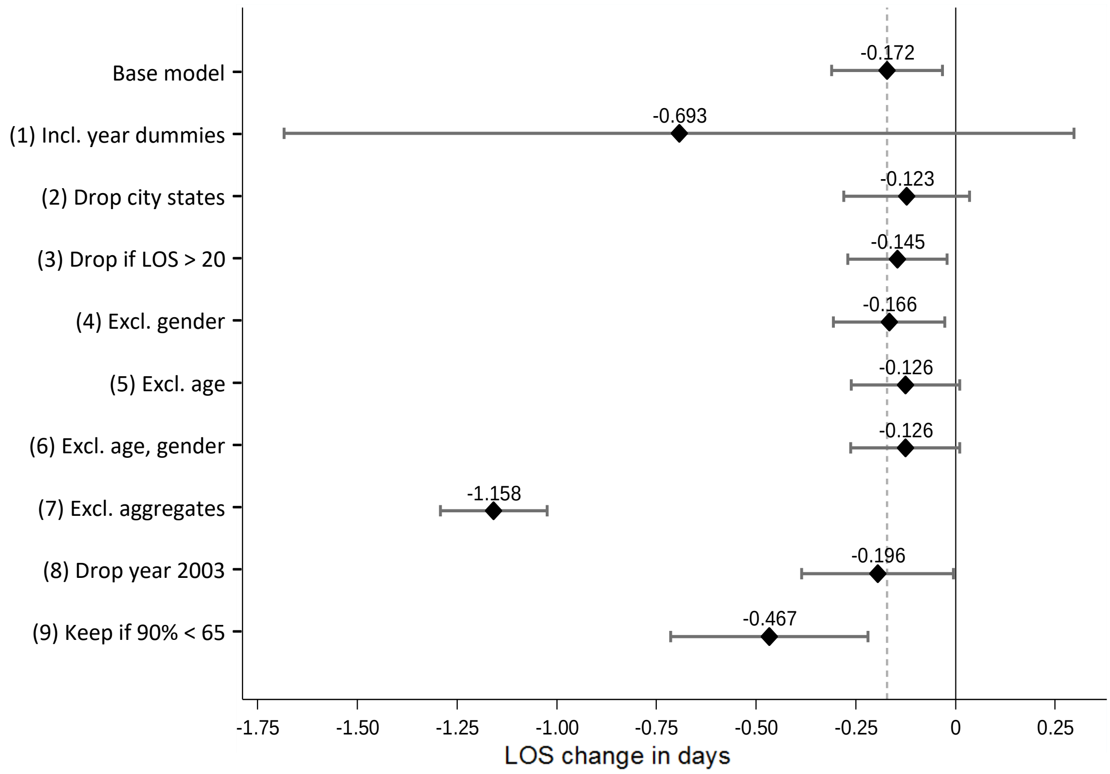

Figure 18: Treatment effect in robustness checks

Average marginal effects in days with 95% confidence intervals; Incl.: Inclusion in model; Excl.: Exclude variable from model; number in parenthesis refers to column of robustness check table in appendix E.

Dropping observations with a LOS of over 20 days and excluding gender from the model both produces slightly lower but significant estimates (p-values: 0.022 and 0.019). Excluding age or respectively age and gender, the two covariates which were available on the level of the disease group, both leads to a calculated reform effect of -0.126 days with a p-value of 0.07.

When running a specification without the aggregated state variables, the estimated reform effect is highly significant and amounts to -1.158 days. The variables on hospital and outpatient structure seem to mitigate the treatment effect. Dropping observations of the first year of DRG use in Germany (2003) brings about a slightly higher estimate than the base model (p-value: 0.04).

The last robustness check which was conducted considered only observations where the share of patients younger than 65 was at least 90%. The intention behind this alteration was firstly to examine whether the LOS reduction behaved differently according to age, and secondly to include only observations where the health status and morbidity can be assumed to have stayed constant for the period under review.

11 Discussion

11.1 Summary of findings

This analysis was conducted in order to determine whether the introduction of a PPS (the G-DRG system) incentivized German hospitals to change their practice style concerning the length of hospitalizations. The existing evidence on PPS introductions was gathered in an extensive literature review, followed by setting up an economical model illustrating the change of incentives for hospitals. Jointly, these steps enabled the formulation of several hypothesis concerning the LOS development as a result of the DRG introduction.[26] In a second step, the hypotheses were tested empirically on the basis of the aggregated data. The following table presents a synopsis of the main hypotheses compared to stylized effects, estimated in the regression analysis.

Hospital LOS response to Change from PD to DRG				
	Funding body			
	Not for profit		Private for profit	
	Prediction	Estimate	Prediction	Estimate
Budget neutral optional year 2003	0	0	0	0
Mandatory use of DRG starting in 2004	↓	↓	↓↓	↓
First year of statewide base rates	↓↓	0	↓↓↓	0

Table 6: Hypothesis compared to model results

As expected, no effect could be established for the year of the voluntary and budget neutral use of DRGs in 2003 for both for-profit and not for-profit hospitals. A negative significant but small effect could be confirmed for the first year of mandatory application of the DRG systematic in 2004. This implies that the new reimbursement system did in fact change the incentives according to the economic model established in chapter 6,

[26] See chapter 7.

forcing hospitals to improve their efficiency as they now have to lower their costs under the level of a “yardstick”.

As chapter 10 has shown, the effect varied considerably between different disease groups. The estimated negative significant effect was reasonably robust against model alterations. Most robustness checks confirmed the effect at least on the 10% significance level, and some alterations even lead to estimates two or three times the size of the base result of -0.172 days. This negative effect is in line with the findings from the international literature, where there is overwhelming evidence of LOS reductions as a result of the introduction of a PPS.[27]

This result is also in line with the findings of Hilgers (2011), to this date the only other empirical analysis of the effect of the DRG introduction in on LOS in Germany. However, her estimated effect of -0.45 days is more than twice as high as the base result of this analysis (Hilgers, 2011, p. 114). One possible explanation for this discrepancy is that Hilgers (2011) could directly control for the increasing morbidity in the population by including the PCCL codes in the regression. Using age and disease groups as proxies for that development might not be sufficient, and therefore this analysis potentially underestimates the effect on LOS. This difference could also be the result of sampling issues. Whereas this analysis is based on aggregated data of all German hospitals, Hilgers (2011) used hospitalization data from a small health insurance company with 270,000 insured persons (Hilgers, 2011, p. 98). As a result of the historical development of German statutory health insurance companies (Porter and Guth, 2012, pp. 72-74), it is very likely that socio economic status, as well as the regional distribution of the insured population, are highly correlated within this sample and are not representative for the German population.

This analysis has established that socio-economic status, approximated by income, has an effect on the LOS, and that slight regional variations for the reform effect do exist.[28]

Regarding the additional hypothesis about the influence of ownership on the magnitude of the LOS reductions, it has to be concluded that the predicted larger effect due to a

[27] See chapter 5.3.
[28] See chapter 10.2.

higher share of for-profit hospitals beds could not be observed in the econometric analysis. This partly contradicts the results of Hilgers (2011), where for-profit hospitals were associated with higher reductions than was the case with their counterparts. It is possible that the variable used in this analysis share of for-profit beds on the state level does not pick up the effect in its entirety. Hilgers was able to directly include the type of hospital ownership for every hospitalization. The hypothesis about a larger reform effect in 2005, the first year where hospital specific base rates were converging towards the statewide base rates, did not hold up in the empirical analysis. It seems like hospitals adjusted their practice styles and increased their efficiency within the course of the first year of mandatory use of the PPS systematic.

The results of robustness check 2, where city state observations were excluded from the regression, deserve further consideration, as they are hinting towards a widely discussed topic surrounding the DRG introduction: competition. Chapters 5 and 6 have established that in a DRG system hospitals compete against a “yardstick”, forcing them to improve their efficiency, partly by decreasing the LOS. By just including observations of the city states, the reform effect increases to a highly significant reduction of -0.46 days. Finding that city states behave differently suggests that the local competitive environment could also play a role in the efficiency decision of hospitals, as the three city states do have more beds per population than most territorial states in the data set. City states are different in many ways, so to support this claim, only the observations of the four states with the highest amount of beds per population, as proxy for hospital competition, were included in another regression. This resulted in an effect of -0.3 days, significant on the 1% level, confirming the hypothesis and also confirming the results of Choné et al. (2013), who discovered that higher competition is connected to higher LOS reduction as a response to the introduction of a PPS.

Overall, there is high confidence in the estimated small but significant decreasing effect of the DRG reform on LOS, confirming the results of Hilgers (2011). Additionally, this analysis is an important contribution to the literature itself, as it is the first work to examine the effect of the DRG reform on the basis of data on the full sample of German hospital cases and hospitals. Although the information used was aggregated on the disease group and state level, this work represents the first step of closing this gap in the literature.

11.2 Limitations

11.2.1 Aggregated data

As presented in chapter 8, the used data set contained aggregated data on either disease group level or state level. The use of such data is connected to several problems. First, inference of aggregated data is only valid as long as the aggregates are representative for the true population (Hellerstein, 1991, 865-866). This is given in this context, as the aggregates are based on the full sample of hospital cases and hospitals. Another problem of using aggregated data arises when the micro level observation and the aggregate measure different constructs and the aggregate has a different effect on the dependent variable than the micro observation (Firebaugh, 1978, pp. 559-560). This is the case, for example, for the income variable in the model. The aggregate captures a multitude of additional effects, such as economic status, social deprivation or education. Hence inference on some of the variables that were only available on the state level might not be valid.

The general results, however, should not be influenced, as the variable of interest, the treatment dummy, is not aggregated. Working with aggregated data has the additional drawback that it prevents differentiated inference. It would have been interesting to see whether specific patient populations were more affected by the reform, or what types of hospital, concerning size or degree of specialization, were responding stronger to the new incentives.

11.2.2 Econometric approach

There are also some aspects of the econometric approach itself which require a critical assessment. For one thing, the analysis employed a Poisson regression and treated the LOS observations as count data, although they are averages of count data, taking up also non-integer values. However, as was discussed in chapter 9.2, the shape of the distribution of the dependent variable is still typical for count data. It is assumed that this allows the use of count data models, as their condition for consistency is purely based on specifying the conditional distribution correctly (Cameron and Trivedi, 2005, p. 121), which has to be approximated by the distribution of the variable itself.

Figure 15, which depicts the observed and predicted values for LOS, shows that the chosen model fits the data quite well. In combination with more or less robust estimates, which lie within an acceptable range of results from the literature, there is some confidence that treating the LOS observations as counts was appropriate.

Another aspect worth discussing is the matter of clustering standard errors. As described and implemented, this kind of individual effects model requires clustering standard errors on the observational unit (Cameron and Miller, 2005, pp. 330-331), here the disease groups per state. If the model would only contain variables on that level, this would be sufficient. However, the majority of covariates are aggregates on the state level. Therefore, it might be necessary to cluster standard errors also on the state level. Multi-way clustering, as discussed for example by Cameron et al. (2011), is a feasible solution in this case. Nevertheless, to this day, there is no STATA or user written command allowing multi-way clustering for Poisson regression.[29] Not accounting for the second cluster still leads to consistent estimates, but the calculated standard errors and p-values might be under- or overestimated (Cameron et al. 2011, p. 238).

Concerning the overall estimation strategy, it needs to be said that the truncated Poisson model with fixed effects was implemented, as the author believes it to be the most straightforward method of obtaining consistent estimate for the treatment effect of the DRG reform. Nevertheless, other, more sophisticated methods like quantile regression or GMM estimators are considered to be the state of the art for panel count data by now (Trivedi, 2010, p. 11; 74).

11.3 Confounding trends

As depicted in chapter 9.4, LOS can be influenced by many different factors and developments. Unfortunately, this analysis cannot control for all of them, due to the lack of appropriate data. Hence, the most relevant potential confounders which had to be omitted in the calculations are discussed in the following.

One issue, which has to be addressed here, is the effect, the technological progress has on the LOS development. As discussed in chapter 3.2, LOS has been declining long before the introduction of the DRG system in Germany. One driver is believed to be medical advances like less invasive surgeries (Sloan and Valvona, 1986). Assuming that the same trend holds for the periods under review means that not accounting for that pre-existing trend could overestimate the treatment effect in the year 2004. There is no perfect way to measure or capture technological progress. For example including year dummies, as done by Hilgers (2011) and Theurl and Winner (2007), greatly inflated the standard errors. The model alterations in chapter 10.2 have shown, that the treatment dummy did not capture any effect in the years 2003 and 2005. This shows, that either the treatment dummy does not capture the effect of the technological progress, or that there is no significant impact of medical advances in the respective period. Either way, overall results and inference should not be greatly influenced by not accounting for technological progress.

Another possibly confounding trend is the increasing number of ambulatory conducted procedures and surgeries (AOP), one manifestation of the progressively importance of the principle of precedence of outpatient care over inpatient care (§ 39(1) SGB V). Although possible since 1993, AOP gained importance also with the "GKV-Gesundheitsreformgesetz" and the introduction of the § 115b SGB V, as the whole concept was reformed, and the catalogue of such hospitalization replacing procedures was extended significantly (Roeder et al., 2014, p. 119). The number of such operations increased from under 600,000 in 2002 to over 1,800,000 in 2010, with the biggest increase from 2003 to 2004 (Bölt and Graf, 2012, p. 120). As a consequence, a lot of low intensity cases dropped out of the pool of hospitalizations in the same time period when the DRGs came into effect. Therefore, average case severity increased. This trend potentially leads to underestimating the effect of the DRG introduction.

A further development, which was only partly accounted for and which is likely to lead to underestimating the impact of the reform, is the demographic transition and the subsequent increases of morbidity and co-morbidity in the patient population (see for example Schnoor et al., 2011). Age was included in the regression, but it is fair to assume that the variable does not capture the increased morbidity and therefore case severity to its full extent. The analysis of Hilgers (2011) was able to account for that by including the PCCL in the regression and came to a larger estimate of the reform effect. The results of robustness check 9, where morbidity is assumed to stay constant, also hints in that direction, as the reform effect is quantified with a significant reduction of almost half a day.

12 Conclusion

The work at hand has provided a review of the effect of PPS introductions on the LOS in hospitals, and has shown that financial incentives change as a result of introducing a PPS. On the basis of an aggregated data set, it was calculated whether such an effect could also be observed for the DRG introduction in Germany in the year 2004. There is overwhelming evidence in the international literature that LOS declined as a consequence of the introduction of a PPS. The calculations of this analysis revealed that this is also true for Germany. However, because of several mentioned limitations, it is very likely that the estimated reform impact of -0.172 days constitutes a conservative estimate and that the true effect is notably larger. The conclusion of the report of the "DRG-Begleitforschung", mentioned in chapter 1, that there is no evidence that the DRG introduction has led to a change in the LOS development, has to be rephrased in the following way: "There is evidence for at least slight reductions of LOS as a consequence of the reform." The author intends to conduct further research, based on micro level data of all German hospitals, in order to provide a more valid and detailed quantification.

Nevertheless, by merely discussing the change in LOS, it cannot be inferred whether the DRG reform achieved its overall goal, to increase efficiency and reduce costs in the German inpatient sector. As was mentioned, hospitals likely increased the number of cases as a reaction to the new reimbursement system, which is at least partially responsible for the observed overall increase in hospital cases.[30] Therefore it is at least doubtful whether the DRG reform led to a more cost-effective delivery of inpatient care in Germany. Furthermore, the new reimbursement system is connected to quality deteriorations, ever-increasing work intensity, cost shifting, and a high burden of documentation and coding (Kunze and Nagel, 2010, pp. 263-266). Hence, it is possible that the "GKV-Gesundheitsreformgesetz 2000" sent the German inpatient sector down a meander. German policy makers should move away from such isolated reform efforts,

[30] See for example table 4.

towards a comprehensive and integrated approach to health care, with instruments like bundled payments or accountable care organizations, in which incentives can be stretched out over the whole continuum of care.

Appendix

Appendix A: Review search strings

PubMed search strings

String	Terms						Hits	Abs.	Inc.
#1	effect	hospital OR clinic	LOS	reimbursement OR prospective reimbursement OR reimbursement mechanism	diagnosis related groups	reform OR healthcare reform act	167	32	5
	(((((((hospital) AND LOS[MeSH Terms]) AND prospective reimbursement[MeSH Terms]) OR reimbursement mechanism[MeSH Terms]) AND diagnosis related group[MeSH Terms]) AND health care reform[MeSH Terms]))								
#1b	effect	hospital OR clinic	LOS	reimbursement OR prospective reimbursement OR reimbursement mechanism	diagnosis related groups	reform OR healthcare reform act	81	18	3
	((((((((hospital) AND LOS) AND prospective reimbursement) OR reimbursement mechanism) AND diagnosis related group) AND health care reform)))								
#2		hospital OR clinic	LOS		diagnosis related groups		11	2	0
	Search (Search (((LOS[MeSH Terms]) AND diagnosis related group[MeSH Terms]) NOT ((((((((hospital) AND LOS[MeSH Terms]) AND prospective reimbursement[MeSH Terms]) OR reimbursement mechanism[MeSH Terms]) AND diagnosis related group[MeSH Terms]) AND health care reform[MeSH Terms])))) NOT (((((((((hospital) AND LOS) AND prospective reimbursement) OR reimbursement mechanism) AND diagnosis related group) AND health care reform)))))								
#3			LOS	prospective reimbursement	diagnosis related groups		11	0	0
	((((("LOS"[MeSH Terms]) AND prospective reimbursement[MeSH Terms]))) AND diagnosis related group[MeSH Terms]) NOT ((("LOS"[MeSH Terms]) AND prospective reimbursement[MeSH Terms]) NOT ((Search (((LOS[MeSH Terms]) AND diagnosis related group[MeSH Terms]) NOT ((((((((hospital) AND LOS[MeSH Terms]) AND prospective reimbursement[MeSH Terms]) OR reimbursement mechanism[MeSH Terms]) AND diagnosis related group[MeSH Terms]) AND health care reform[MeSH Terms])))) NOT (((((((((hospital) AND LOS) AND prospective reimbursement) OR reimbursement mechanism) AND diagnosis related group) AND health care reform)))))))								
#4			LOS		diagnosis related groups	reform	2	2	1
	(((diagnosis related group[MeSH Terms]) AND LOS[MeSH Terms]) AND reform)) NOT ((((((((hospital) AND LOS[MeSH Terms]) AND prospective reimbursement[MeSH Terms]) OR reimbursement mechanism[MeSH Terms]) AND diagnosis related group[MeSH Terms]) AND health care reform[MeSH Terms])))) NOT (((((((((((hospital) AND LOS) AND prospective reimbursement) OR reimbursement mechanism) AND diagnosis related group) AND health care reform))))) NOT ((((((((hospital) AND LOS[MeSH Terms]) AND prospective reimbursement[MeSH Terms]) OR reimbursement mechanism[MeSH Terms]) AND diagnosis related group[MeSH Terms]) AND health care reform[MeSH Terms]))))								

Abs.: Screened Abstracts Inc. In Review included Literature after full text screening.

EconLit Search Strings

String	Terms	Hits	Abs.	Inc.
#5a	LOS in hospital AND diagnosis related groups OR PPS OR incentives	155		
#5b	diagnosis related groups AND incentives	12		
#5c	diagnosis related groups AND LOS AND incentives	1		
#5d	DRG AND incentives	12	37	11
#5e	LOS AND PPS	13		
#5f	incentives AND prospective reimbursement AND hospital	27		
#5g	hospital efficiency AND PPS	10		
#6a	incentives AND lump sum OR PPS AND hospital	155		
			25	3
#6b	DRG OR PPS AND principal-agent theory OR inforamtion asymmetry AND hospital	176		

#6c	hospital AND principal-agent theory/theory/relationship AND DRG OR PPS AND physician	67		
#7	google scholar: "hospital" OR "Krankenhaus" AND "principal-agent theory" from 1970 to 2000	151	1	0

Abs.: Screened Abstracts Inc. In Review included Literature after full text screening.

Appendix B: Review of empirical studies

Overview of international studies (1/4)

Author	Country	Setting	Periods under review	Data source	Method	LOS change after PPS introduction	Limitation
Besstremyannaya (2013)	Japan	Replacement of FFS by diagnose based flat-fee payment, partially starting in 2003	2003-2009	Balanced panel of 347 general hospitals based on annual financial and quality data	Difference-in-Differences estimation	- 7.39% after the first year	Very high LOS in J; Exclusion of hosp. with LOS < 6 days
Bogut et al. (2012)	Croatia	FFS scheme replaced by case-based payments for specific indications starting in 2002	2000-2009	Hospitalization claims data for five common procedures from all 68 hospitals from 2000 to 2009; in total 353,000 cases	Interrupted time series approach, aggregating cases to each month for each hospital	cataracts +0.7*; pneumonia -3.2***; bypass -4.0; appendix -1.3***; Hip repl. -1.6	Broadly defined diagnosis groups (118 in 2005)
Busato and Below (2010)	SUI	Introduction of PPS in some cantons before nationwide implementation provides basis of natural experiment	2003-2007	Complete hospital discharge and outpatient claims data, aggregated to 86 hospital services areas	Small area Difference-in-Differences estimation	No reduction	Spatial model does not account for that only 55% of

							hospital cases occur locally
Cheng et al. (2012)	Taiwan	Introduction of DRG payments for parts of hospital services in 2010; cardiovascular cases reimbursed via DRG act as treatment group	2009-2010	Nationwide NHI service claims from 2009-2010 with ~20,000 observations of cardiovascular cases	Difference-in-Differences estimation with propensity score matching (PSM)	- 10 % within 1 year or - 0.82 days	Control and treatment group are quite different, PSM might not be able to heal that
Choné et al. (2013)	France	Shift from global budget to patient-based payment for public and not-for-profit hospitals starting in 2004	2005-2008	Official health statistics; Balanced panel with 182,610 surgical hospital-DRG pairs from 2005 to 2008	Multivariate OLS regression	- 0.13 days - 2.4 %	For-profit-hospitals not included (faced different shift)

Overview of international studies (2/4)

Author	Country	Setting	Periods under review	Data source	Method	LOS change after PPS introduction	Limitation
Farrar et al. (2009)	GB	Introduction of a case mix based payment system replacing cost and block contracts in England's NHS; natural experiment with Scotland as control group	2003-2004	English hospital episode statistics and Scottish morbidity records from 2002-2005 (2.72 million observations)	Linear Panel data regression with fixed effects	- 0.08 days as first year results	Possible spillover effects and opposing trends
Giammanco (1999)	Italy	Cost-based reimbursement and per-diem payments are replaced by a PPS system in 1995	1995	~ 700,000 inpatient episodes in 88 public and 47 private hospitals in Sicily in all four quarters of 1995	Panel regression using MDCs as observation and period dummies as treatment	- 3.0 days	Use of 1st quarter of reform year as pre reform control
Hamada et al. (2012)	Japan	Replacement of FFS by diagnose based flat-fee payment in 2003	2001-2009	Administrative data from 11,159 myocardial infarct patients in 75 hospitals	Multilevel fixed effects regression to account for hospital cluster	- 2.29 days	Only one indication was examined

MOHW (2000, cited in Kwon, 2003)	Korea	3-year DRG pilot project starting in 1997 as alternative to traditional FFS	1997-2000	54/132/798 pilot hospitals in year 1/2/3	Not stated, possibly DID	- 5.7 %	Methods unclear
Louis et al. (1999)	Italy	Move from global budgeting approach to DRG based payment	1993-1996	Discharge data from 32 hospitals in one region on nine common conditions	Descriptive, no regression	- 0.3 days - 3.3 %	Simple method and restricted to one region
Magnussen and Solstad (1994)	Norway	Fixed grants replaced by mixed system including DRG based payment in pilot hospitals in 1991	1991-1992	Discharge data from 4 pilot hospitals	Comparing pilot with regular care	No significant effect	Small sample of hospitals; Mixed reimbursement system

Overview of international studies (3/4)

Author	Country	Setting	Periods under review	Data source	Method	LOS change after PPS introduction	Limitation
Mikkola and Hakkinen (2002)	Finland	Reform in 1993 incentivized using case-based payment instead of traditional cost-based approach	1994-1998	Discharge data from ~50,000 surgical patients	Count data regression	- 0.28 days (Own calculations)	Small number of indications (three)
Moreno-Serra and Wagstaff (2010)	East Europe/ Central Asia	Shifts from historical budgets to patient-based payment in eastern European and central Asian countries since the 1990s	1990-2004	Aggregated panel data on 28 countries from 1990 to 2004 from various sources	Difference-in-Differences regression	- 0.44 days - 3.5 %	Aggregated data; huge differences between health systems
Perelman and Closon (2007)	BEL	Per-diem payment system for non-medical activities was modified by introducing prospectively determined, risk oriented LOS thresholds	1991-1998	Complete hospital discharge data from 1991-1998; aggregated over “DRGs”, age categories and hospitals, leading to 528,500 yearly observations	Linear Panel data regression with fixed effects	- 1.49 %	Medical procedures still payed via FFS; No comprehensive DRG system

Theurl and Winner (2007)	Austria	Change from per-diem-based to per case-based reimbursement in 1997	1989-2003	Official statistics on 20 diagnostic categories aggregated on the nine Austrian states (n=2,700)	Fixed effects regression	- 0.43 days	Data aggregated on state level
Tsai (2005)	Taiwan	Introduction of per-case payment for specific surgical procedures in 1997 (28 in the first year), replacing traditional FFS; natural experiment	1997-1998	Complete Hospital claims data on hemorrhoidectomy patients (23,638 cases) 9 months before and after the reform	Multivariate linear regression	- 0.69 days	Narrow indication; Simple linear regression
Wang et al. (2010)	Japan	Replacement of FFS by diagnose based flat-fee payment	2002-2003	Official expenditure data of 82 hospitals	Descriptive comparison for targeted hospitals	- 1.6 days - 7.8 %	Descriptive; High LOS in Japan

Overview of international studies (4/4)

Author	Country	Setting	Periods under review	Data source	Method	LOS change after PPS introduction	Limitation
Zhang (2010)	Shanghai	FFS replaced by PPS in a two-year experiment	2004-2005	Random sample of 14000 insured patients suffering from 13 target diseases from one large hospital	Difference-in-Differences with 2 test and 2 pre-test periods	- 13 %	Only one hospital Different system with high deductibles

Overview of studies from the U.S. (1/3)

Author	Setting	Periods under review	Data source	Method	LOS change after PPS introduction	Limitation
Frank and Lave (1989)	Variation in hospital reimbursement methods across state Medicaid programs in the early 1980s act as setting for natural experiment	1981-1984	National hospital discharge survey, annual survey of hospital; selective patient group consisting of 5,216 discharges	Semiparametric extreme value hazard models	- 17 % relative to cost based reimbursement	Analysis only included psychiatric patients
Gilman (2000)	Replacing system entirely based on patient characteristics by diagnosis and procedure related DRGs in New York at the beginning of 1994	1992-1995	New York State hospital discharge data on all HIV related DRGs, 48,458 cases	Difference-in-Differences estimation	- 3.91 days	Narrow indication which underwent enormous medical progress
Hadley et al. (1989)	Shift to Medicare's PPS starting in 1983, replacing cost based payments,	1985	AHA's annual survey of hospitals for 1983 to 1985, final sample with 816 hospitals; administrative data on payment rates	OLS multiple regression analysis	- 10.3 % for first and - 2.6 % for second year	Case mix of patients could not be observed directly, severity could counteract the policy effect
Hsiao and Dunn (1987)	Implementation of DRG based PPS gradually replacing per-diem scheme in New Jersey starting in 1980	1971-1984	AHA's annual survey of hospitals; pooled cross section of the bulk of New Jersey's hospitals from 1971 to 1984 (~100)	Time trend model with GLS specification	Trend change of - 2 % starting in 1980	Hospital specific DRG rates including historical costs

Overview of studies from the U.S. (2/3)

Author	Setting	Periods under review	Data source	Method	LOS change after PPS introduction	Limitation
Kahn et al. (1990); Draper et al. (1990)	Shift to Medicare's PPS starting in 1983, replacing cost based payments	1981-1986	Nationally representative sample of 14,012 Medicare patients suffering from one of five diseases, hospitalized before and after the reform in 1983	Representative sampling; pre-post design with multiple time points	- 24 % or - 3.4 days	Post PPS data from periods, where financial base was still changing a lot; though treatment and control are quite similar, other developments might counteract PPS effects
Lave and Frank (1990)	Medicaid hospital payment systems in the U.S. in 1984 differed fundamentally between the states	1984	Medicaid discharge data from 1,670 hospitals on selected DRGs; Annual survey of hospitals and Area Resource File for hospital and area characteristics; ~134,000 cases	OLS	From 2% for medical non-public hospital cases to 17.9% for psych. Cases in public hospitals	Information on patient characteristics very limited; Level of payment not included in analysis
McCue and Thompson (2006)	Implementation of PPS for Medicare population in inpatient rehabilitation facilities, replacing cost-based scheme starting in 1997, general in effect starting in 2002	2002	Health Care Cost Report Information System data on 120 IRFs using PPS before October 2002 and 26 IRFs using cost based payment throughout 2002	Difference-in-Differences	+ 1.03 days (own calculation)	Simple research design, not accounting for confounding effects like changes in patient characteristics, selection or changes in care delivery
Norton et al. (2002)	Per-diem payment in Massachusetts' Medicaid program replaced by fixed per episode payment in October 1991 lasting until October 1992	1991-1993	Massachusetts Medicaid data base, final sample with 8,509 admissions for 4,193 severely mentally ill patients in 50 hospitals	Linear Panel data regression with random and fixed effects specifications	No significant effect of the payment reform on LOS	PPS only in effect for short time period; restricted to mentally ill patients

Overview of studies from the U.S. (3/3)

Author	Setting	Periods under review	Data source	Method	LOS change after PPS introduction	Limitation
Salkever and Steinwachs (1988)	Maryland applied both per case and per service payments starting in 1976, providing a setting for a natural experiment	1976-1981	Medicare cost reports and discharge data of 46 hospitals, 22 of which with experience of per case payments; 9 common DRGs were evaluated	Linear Panel data regression with fixed effects	no significant results overall, possible effect for some subgroups of hospitals	Short period of PPS in hospitals (avg. 2 years); Generalizability problematic as Maryland's FFS scheme was already quite stringent
Sood et al. (2008)	Implementation of PPS for Medicare population in inpatient rehabilitation facilities replacing cost-based scheme starting in 1997, general in effect starting in 2002	2001-2003	Medicare cost reports and claims data for acute care and rehabilitation for 108,692 stroke patients, 92,142 hip fracture, 229,705 joint replacement in 1145 IRFs; 4 quarters in 2001 as pre periods, 10 afterwards as post periods	Linear Panel data regression with fixed effects and instrumental variable approach to disentangle effect of marginal and average reimbursement	- 3.2 % for joint repl. upto 10.5 % for stroke patients	Short post PPS period (1.5 years)

Appendix D: Data Sources

LOS, age, gender, no. of cases per diagnostic group

Krankenhausstatistik - Diagnosedaten der Patienten und Patientinnen in Krankenhäusern. Made available by https://www.gbe-bund.de; 14th October 2014 and 30th March 2016.

Table: Diagnosedaten der Krankenhäuser ab 2000 (Fälle/Sterbefälle, Berechnungs- und Belegungstage, durchschnittliche Verweildauer).

Hospital costs in 1000 € per state

Statistisches Bundesamt. Kostennachweis der Krankenhäuser. Wiesbaden, 2000-2012.

Beds, cases per 100,000, medical personal, non-medical personal, share for profit, hospital beds, bed utilization

Statistisches Bundesamt. Grunddaten der Krankenhäuser. Wiesbaden, 2000-2012.

Outpatient physicians per 100,000 inhabitants

Kassenärztliche Bundesvereinigung – Bundesarztregister. Made available by https://www.gbe-bund.de; 14th October 2014.

Table: An der vertragsärztlichen Versorgung teilnehmende Ärztinnen und Ärzte sowie Psychotherapeutinnen und -therapeuten (Anzahl). Gliederungsmerkmale: Jahre, Region, Geschlecht, Teilnahmestatus, Arzt- und Psychotherapeutengruppe.

Number of long term care beds

Pflegestatistik - Ambulante und stationäre Pflegeeinrichtungen: Grunddaten, Personalbestand, Pflegebedürftige, Empfänger und Empfängerinnen von Pflegegeldleistungen. Made available by https://www.gbe-bund.de; 30th March 2016.

Table: Pflegeheime und verfügbare Plätze in Pflegeheimen. Gliederungsmerkmale: Jahre, Region, Art der Einrichtungen/Plätze, Träger

Quota of Long term care patients

Pflegestatistik - Ambulante und stationäre Pflegeeinrichtungen: Grunddaten, Personalbestand, Pflegebedürftige, Empfänger und Empfängerinnen von Pflegegeldleistungen. Made available by https://www.gbe-bund.de; 30th March 2016.

Table: Pflegebedürftige (Anzahl und Quote). Gliederungsmerkmale: Jahre, Region, Alter, Geschlecht.

Available income of private households per inhabitant

Statistisches Landesamt Sachsen-Anhalt, 2015. Indikator K019 - Verfügbares Einkommen privater Haushalte je Einwohner [€]. Last updated spring 2015. [online] Available at:https://www.statistik.sachsen-anhalt.de/apps/Struktur Kompass/indikator/zeitreihe/119 [Accessed: 30th March 2016].

Appendix E: Abbreviations for disease groups and states

ICD10 - GM	Description
Z00-Z99	Faktoren, die den Gesundheitszustand beeinflussen und zur Inanspruchnahme des Gesundheitswesens führen
A00-B99	Bestimmte infektiöse und parasitäre Krankheiten
C00-D48	Neubildungen
D50-D90	Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems
E00-E90	Endokrine, Ernährungs- und Stoffwechselkrankheiten
F00-F99	Psychische und Verhaltensstörungen
G00-G99	Krankheiten des Nervensystems
H00-H59	Krankheiten des Auges und der Augenanhangsgebilde
H60-H95	Krankheiten des Ohres und des Warzenfortsatzes
I00-I99	Krankheiten des Kreislaufsystems
J00-J99	Krankheiten des Atmungssystems
K00-K93	Krankheiten des Verdauungssystems
L00-L99	Krankheiten der Haut und der Unterhaut
M00-M99	Krankheiten des Muskel-Skelett-Systems und des Bindegewebes
N00-N99	Krankheiten des Urogenitalsystems
O00-O99	Schwangerschaft, Geburt und Wochenbett
P00-P96	Bestimmte Zustände, die ihren Ursprung in der Perinatalperiode haben
Q00-Q99	Angeborene Fehlbildungen, Deformitäten und Chromosomenanomalien
R00-R99	Symptome und abnorme klinische und Laborbefunde, die anderenorts nicht klassifiziert sind
S00-T98	Verletzungen, Vergiftungen und bestimmte andere Folgen äußerer Ursachen

Source: Deutsches Institut für Medizinische Dokumentation und Information (2015).

Abbreviation	State
BW	Baden-Württemberg
BY	Bayern
BE	Berlin
BB	Brandenburg
HB	Bremen
HH	Hamburg
HE	Hessen
MV	Mecklenburg-Vorpommern
NI	Niedersachsen
NW	Nordrhein-Westfalen
RP	Rheinland-Pfalz
SL	Saarland
SN	Sachsen
ST	Sachsen-Anhalt
SH	Schleswig-Holstein
TH	Thüringen

Appendix F: Marginal effects robustness checks

	(1) Year dummies	(2) Dropping city states	(3) Dropping LOS >20
Treatment dummy	-0.693	-0.123	-0.145**
	(0.505)	(0.0804)	(0.0636)
Share 65 to 75 in %	0.0170	0.0201	0.0142
	(0.0146)	(0.0156)	(0.0126)
Share 75 to 85 in %	0.0159	0.0201	0.0420***
	(0.0143)	(0.0153)	(0.0128)
Share over 85 in %	0.0674***	0.0530***	0.0743***
	(0.0188)	(0.0196)	(0.0157)
Share female cases in %	-0.00771	-0.00967	-0.0244**
	(0.0123)	(0.0143)	(0.0108)
Income HH member in €1000	-0.114*	-0.125***	-0.161***
	(0.0620)	(0.0481)	(0.0347)
Costs per case in €1,000	-0.0230	-0.575***	-0.392***
	(0.171)	(0.197)	(0.126)
Beds per 100k P in 100	0.693***	0.732***	0.600***
	(0.171)	(0.146)	(0.119)
Cases per 100k P in 1,000	-0.172***	-0.204***	-0.227***
	(0.0641)	(0.0600)	(0.0432)
Share for profit hospitals in %	0.00106	0.00208	0.00274
	(0.00335)	(0.00333)	(0.00283)
Non med per 100k cases in 100	0.0322*	0.0565***	0.0629***
	(0.0193)	(0.0219)	(0.0133)
Med per 100k cases in 100	-0.0435	-0.0533	-0.182**
	(0.114)	(0.106)	(0.0791)
Bed utilization in %	0.0627***	0.0550***	0.0521***
	(0.0139)	(0.0103)	(0.00859)
Outpatient phys. per 100k P	0.00990	0.0166**	0.0186***
	(0.00643)	(0.00733)	(0.00532)
N	4160	3380	4028
pseudo R^2	0.2614	0.2650	0.1814
AIC	16152.9	13117.2	15468.6
BIC	16311.2	13203.0	15556.8

Average marginal effects; Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	(4) Excl. gender	(5) Excl. age	(6) Excl. age gender
Treatment dummy	-0.166** (0.0712)	-0.126* (0.0694)	-0.126* (0.0697)
Share 65 to 75 in %	0.0107 (0.0127)	-	-
Share 75 to 85 in %	0.0232* (0.0135)	-	-
Share over 85 in %	0.0545*** (0.0178)	-	-
Share female cases in %	-	0.000438 (0.0118)	-
Income HH member in €1000	-0.169*** (0.0411)	-0.149*** (0.0409)	-0.149*** (0.0409)
Costs per case in €1,000	-0.346*** (0.131)	-0.271** (0.125)	-0.271** (0.125)
Beds per 100k P in 100	0.639*** (0.122)	0.661*** (0.122)	0.661*** (0.121)
Cases per 100k P in 1,000	-0.186*** (0.0496)	-0.165*** (0.0500)	-0.165*** (0.0500)
Share for profit hospitals in %	0.00233 (0.00326)	0.00146 (0.00317)	0.00146 (0.00317)
Non med per 100k cases in 100	0.0559*** (0.0155)	0.0543*** (0.0153)	0.0543*** (0.0153)
Med per 100k cases in 100	-0.0598 (0.0927)	-0.0322 (0.0892)	-0.0323 (0.0890)
Bed utilization in %	0.0477*** (0.00902)	0.0500*** (0.00835)	0.0500*** (0.00835)
Outpatient phys. per 100k P	0.0137** (0.00616)	0.0119** (0.00608)	0.0119** (0.00608)
N	4160	4160	4160
pseudo R^2	0.2613	0.2612	0.2612
AIC	16129.9	16128.9	16126.9
BIC	16212.3	16198.6	16190.3

Average marginal effects; Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	(7) Drop aggregates	(8) Drop year 2003	(9) Keep if 90% <65 years
Treatment dummy	-1.158***	-0.196**	-0.467***
	(0.0682)	(0.0972)	(0.126)
Share 65 to 75 in %	-0.00476	0.0181	0.0272***
	(0.0129)	(0.0119)	(0.0885)
Share 75 to 85 in %	-0.00310	0.0293**	-0.0372***
	(0.0115)	(0.0139)	(0.0144)
Share over 85 in %	-0.0695***	0.0395**	0.0191*
	(0.0155)	(0.0178)	(0.0109)
Share female cases in %	0.0175	-0.0000390	-0.0166
	(0.0134)	(0.0123)	(0.0194)
Income HH member in €1,000	-	-0.158***	-0.214**
		(0.0396)	(0.101)
Costs per case in €1,000	-	-0.301**	-0.246
		(0.132)	(0.407)
Beds per 100k P in 100	-	0.579***	-0.0330
		(0.131)	(0.290)
Cases per 100k P in 1,000	-	-0.175***	0.0605
		(0.0495)	(0.116)
Share for profit hospitals in %	-	0.00319	0.00388
		(0.00339)	(0.00934)
Non med per 100k cases in 100	-	0.0635***	0.117***
		(0.0161)	(0.0300)
Med per 100k cases in 100	-	-0.110	-0.450**
		(0.0981)	(0.205)
Bed utilization in %	-	0.0471***	-0.00645
		(0.0112)	(0.0151)
Outpatient phys. per 100k P	-	0.0122**	0.00850
		(0.00608)	(0.0113)
N	4160	3840	724
pseudo R^2	0.2585	0.2613	0.2471
AIC	16176.1	14868.3	2716.2
BIC	16207.8	14955.9	2780.4

Average marginal effects; Standard errors in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Literature

Akerlof, G.A., 1970. The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, 84(3), pp. 488–500.

Arrow, K.J., 1963. Uncertainty and the welfare economics of medical care. *American Economic Review*. 53(5), pp. 941–973.

Ärzteblatt, 2016. BGH: *Landkreise dürfen ihre Krankenhäuser finanziell unterstützen.* Available at: http://www.aerzteblatt.de/nachrichten/66147/BGH-Landkreise-duerfen-ihre-Krankenhaeuser-finanziell-unterstuetzen [Accessed: 31 March 2016].

Ashenfelter, O., 1978. Estimating the Effect of Training Programs on Earnings. *The Review of Economics and Statistics*, 60(1), pp. 47-57.

Augurzky, B. and Schmitz, H., 2013. *Wissenschaftliche Untersuchung zu den Ursachenunterschiedlicher Basisfallwerte der Länder als Grundlage der Krankenhausfinanzierung*: *Forschungsprojekt im Auftrag des Bundesministeriums für Gesundheit.* Essen: Rheinisch-Westfälisches Institut für Wirtschaftsforschung.

Augurzky, B., Gülker, R., Mennicken, R., Felder, S., Meyer, S., Wasem, J., Gülker, H. and Siemssen, N., 2012. *Mengenentwicklung und Mengensteuerung stationärer Leistungen: Endbericht: Forschungsprojekt im Auftrag des GKV-Spitzenverbandes.* Essen: Rheinisch-Westfälisches Institut für Wirtschaftsforschung.

Besstremyannaya, G., 2013. The impact of Japanese hospital financing reform on hospital efficiency. A difference-in-difference approach. *Japanese Economic Review*, 64(3), pp. 337–362.

Blum, K., Löffert, S., Offermanns, M. and Steffen, P., 2010. *Krankenhaus Barometer: Umfrage 2010.* Düsseldorf: Deutsches Krankenhausinstitut.

Böcking, W., Ahrens, U., Kirch, W. and Milakovic, M., 2005. First results of the introduction of DRGs in Germany and overview of experience from other DRG countries. *Journal of Public Health*, 13(3), pp. 128–137.

Bogut, M., Voncina, L. and Yeh, E., 2012. Impact of Hospital Provider Payment Reforms in Croatia. *The World Bank Policy Research Working Paper*, No. 5992.

Bölt, U. and Graf, T., 2012. 20 Jahre Krankenhausstatistik. *Wirtschaft und Statistik*, February Issue, pp. 112–139.

Buchanan, A., 1988. Principal/agent theory and decision making in health care. *Bioethics*, 2(4), pp. 317–333.

Buhr, P. and Klinke, S., 2006. Qualitative Folgen der DRG-Einführung für Arbeitsbedingungen und Versorgung im Krankenhaus unter Bedingungen fortgesetzter Budgetierung: Eine vergleichende Auswertung von vier Fallstudien. *Discussion Papers, Research Group Public Health from Social Science Research Center Berlin (WZB)*, No. SP I 2006-311.

Bulmer, M.G., 1979. *Principles of statistics*. 2nd ed. Dover Publications: New York.

Busato, A. and Below, G.V., 2010. The implementation of DRG-based hospital reimbursement in Switzerland: A population-based perspective. *Health Research Policy and Systems*, 8(1), pp. 1-6.

Cameron, A.C. and Miller, D.L., 2015. A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2), pp. 317–372.

Cameron, A.C. and Trivedi, P.K., 1998. *Regression Analysis of Count Data.* Econometric Society Monograph No.30. Cambridge University Press.

Cameron, A.C. and Trivedi, P.K., 2001. Essentials of count data regression. In B.H. Baltagi, ed. 2001. *A companion to theoretical econometrics*. Oxford: Blackwell Publishing. 331-348.

Cameron, A.C. and Trivedi, P.K., 2005. *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press.

Cameron, A.C. and Trivedi, P.K., 2015. Count Panel Data. In B.H. Baltagi, ed. 2015. *The Oxford handbook of panel data*. Oxford: Oxford University Press, pp. 233–256.

Cameron, A.C., Gelbach, J.B. and Miller, D.L., 2011. Robust Inference with Multiway Clustering. *Journal of Business & Economic Statistics*, 29(2), pp. 238–249.

Chalkley, M. and Malcomson, J.M., 2000. Government purchasing of health services. In A.J. Culyer and J.P. Newhouse, eds. 2000. *Handbook of health economics*. Oxford: Elsevier. pp. 847–890.

Chalkley, M. and Malcomson, J.M., 2002. Cost sharing in health service provision: an empirical assessment of cost savings. *Journal of Public Economics,* Vol. 84(2002), 219–249.

Cheng, S.-H., Chen, C.-C. and Tsai, S.-L., 2012. The impacts of DRG-based payments on health care provider behaviors under a universal coverage system: A population-based study. *Health Policy*, 107(2-3), pp. 202–208.

Choné, P., Evain, F., Wilner, L. and Yilmaz, E., 2013. Introducing activity-based payment in the hospital industry: Evidence from French data. *CESifo Working Paper*, No. 4304.

Clarke, A., 1996. Why are we trying to reduce length of stay? Evaluation of the costs and benefits of reducing time in hospital must start from the objectives that govern change. *Quality and Safety in Health Care*, 5(3), pp. 172–179.

Debatin, J., 2008. Krankenhäuser: Mehr Qualität und Effizienz durch Wettbewerb. In V. Schumpelick and B. Vogel, eds. 2007. *Medizin zwischen Humanität und Wettbewerb: Probleme, Trends und Perspektiven: Beiträge des Symposiums vom 27. bis 30. September 2007 in Cadenabbia*. Freiburg: Herder. pp. 392–399.

Deutscher Bundestag, 2001. *Entwurf eines Gesetzes zur Einführung des diagnose-orientierten Fallpauschalensystems für Krankenhäuser (Fallpauschalengesetz-FPG).* Printed matter 14/6893, as of date 09/11/2001.

Deutscher Bundestag, 2003. *Entwurf eines Gesetzes zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz – GMG).* Printed matter 15/1525, as of date 09/08/2003.

Deutsches Institut für Medizinische Dokumentation und Information, 2015. *Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme 10. Revision German Modification Version 2016 Mit Aktualisierung vom 21.12.2015.* [online] Available at: https://www.dimdi.de/static/de/klassi/icd-10-gm/kodesuche/onlinefassungen/htm lgm2016/ [Accessed 12 April 2016].

Dowling, W.L., 1974. Prospective Reimbursement of Hospitals. *Inquiry*, 11(3), pp. 163–180.

Dranove, D., 1987. Rate-Setting by Diagnosis Related Groups and Hospital Specialization. *The Rand Journal of Economics*, 18(3), pp. 417–427.

Draper, D., Kahn, K.L., Reinisch, E.J., Sherwood, M.J., Carney, M.F., Kosecoff, J., Keeler, E.B., Rogers, W.H., Savitt, H., Allen, H., Wells, K.B., Reboussin, D. and Brook, R.H., 1990. Studying the Effects of the DRG-Based Prospective Payment System on Quality of Care: Design, Sampling, and Fieldwork. *Journal of the American Medical Association*, 264(15), pp. 1956–1961.

Duckett, S.J., Jackson, T., Hatcher, J., Richards, H., Murphy, K., 2013. *The why, the what and the how of activity-based funding in Canada: a resource for health system funders and hospital managers.* Ottawa: Canadian Institute for Health Information.

Ellis, R.P. and McGuire, T.G., 1986. Provider behavior under prospective reimbursement. Cost sharing and supply. *Journal of Health Economics*, 5(2), pp. 129–151.

Ellis, R.P. and McGuire, T.G., 1996. Hospital response to prospective payment: moral hazard, selection, and practice-style effects. *Journal of health economics*, 15(3), pp. 257–277.

Farrar, S., Yi, D., Sutton, M., Chalkley, M., Sussex, J. and Scott, A., 2009. Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. *British Medical Journal*, 339(b3047), pp. 1-8.

Firebaugh, G., 1978. A Rule for Inferring Individual-Level Relationships from Aggregate Data. *American Sociological Review*, 43(4), pp. 557–572.

Fischer, W. (ed.), 2015. G-DRG - German Diagnosis Related Groups (Deutschland). In: W. Fischer, ed. 2000. *Diagnosis Related Groups (DRGs) und verwandte Patientenklassifikationssysteme*. Latest update on October 10th 2015. Zentrum für Informatik und wirtschaftliche Medizin. [online] Available at: http://fischer-zim.ch/text-pcssa/t-ga-H4-System-GDRG-0003.htm [Accessed 27th April 2016].

Fischer, W. (ed.), 2000. *Diagnosis Related Groups (DRGs) und verwandte Patientenklassifikationssysteme.* Survey on behalf of the DKG. Zentrum für Informatik und wirtschaftliche Medizin. [online] Available at: http://fischer-zim.ch/studien/DRG-Systeme-0003-Info.htm [Accessed 27th April 2016].

Fleßa, S., 2013. *Grundzüge der Krankenhausbetriebslehre:* Part 1. 3rd ed. Munich: Oldenbourg Wissenschaftsverlag.

Flintrop, J., 2006. Auswirkungen der DRG-Einführung: Die ökonomische Logik wird zum Maß der Dinge. *Deutsches Ärzteblatt International*, 103(46), pp. A-3082 / B-2683 / C-2574.

Folland, S., Goodman, A.C. and Stano, M., 2013. *The economics of health and health care*. 7th ed. Upper Saddle River, NJ: Pearson Education Inc.

Frank, R.G. and Lave, J.R., 1989. A Comparison of Hospital Responses to Reimbursement Policies for Medicaid Psychiatric Patients. *The Rand Journal of Economics*, 20(4), pp. 588–600.

Friedrich, J., Leber, W.-D. and Wolff, J., 2010. Basisfallwerte - zur Preis und Produktivitätsentwicklung stationärer Leistungen. In: J. Klauber, M. Geraedts and J. Friedrich, eds. 2010. *Krankenhaus-Report 2010: Schwerpunkt: Krankenhausversorgung in der Krise?.* Stuttgart: Schattauer. pp. 127–147.

Fürstenberg, T., Laschat, M., Zich, K., Klein, S., Gierling, P., Nolting, H.-D. and Schmidt, T., 2013. *G-DRG Begleitforschung gemäß §17b Abs. 8 KHG: Endbericht des dritten Forschungszyklus (2008-2010). Untersuchung im Auftrag des deutschen DRG-Instituts (InEK).* Berlin: IGES Institut.

Furubotn, E.G. and Richter, R., 2000. *Institutions and economic theory: The contribution of the new institutional economics*. Ann Arbor: University of Michigan Press.

Geissler, A., Scheller-Kreinsen, D., Quentin, W. and Busse, R., 2011. Germany: Understanding G-DRGs. In: R. Busse, A. Geissler, W. Quentin and M. Wiley, eds. 2011. *Diagnosis-related groups in Europe - moving towards transparency, efficiency and quality in hospitals.* Maidenhead: Open University Press. pp. 243–272.

Gelman, A. and Hill, J., 2007. *Data analysis using regression and multilevel/hierarchical models: Analytical methods for social research*. Cambridge: Cambridge University Press.

Giammanco, M.D., 1999. The short-term response of hospitals to the introduction of the DRG based prospective payment system: some evidence from Italy. *Giornale degli Economisti e Annali di Economia*, 58(1), pp. 27–62.

Gilman, B.H., 2000. Hospital response to DRG refinements: the impact of multiple reimbursement incentives on inpatient length of stay. *Health Economics,* 9(4), pp. 277–294.

Grobe, T., 2005. Stationäre Versorgung – Krankenhausbehandlungen. In E. Swart and P. Ihle, eds. 2005. *Routinedaten im Gesundheitswesen: Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven.* Bern: Verlag Hans Huber. pp. 79–98.

Hadley, J., Zuckerman, S. and Feder, J., 1989. Profits and Fiscal Pressure in the Prospective Payment System: Their Impacts on Hospitals. *Inquiry*, 26(3), pp. 354–365.

Hamada, H., Sekimoto, M. and Imanaka, Y., 2012. Effects of the per-diem prospective payment system with DRG-like grouping system (DPC/PDPS) on resource usage and healthcare quality in Japan. *Health Policy*, 107(2-3), pp. 194–201.

Hauck, K. and Zhao, X., 2011. How dangerous is a day in hospital? A model of adverse events and length of stay for medical inpatients. *Medical Care*, 49(12), pp. 1068–1075.

Hellerstein, D.M., 1991. Using Count Data Models in Travel Cost Analysis with Aggregate Data. *American Journal of Agricultural Economics*, 73(3), p. 860-866.

Hensen, P., Beissert, S., Bruckner-Tuderman, L., Luger, T.A., Roeder, N. and Muller, M.L., 2008. Introduction of diagnosis-related groups in Germany. Evaluation of impact on in-patient care in a dermatological setting. *The European Journal of Public Health*, 18(1), pp. 85–91.

Herwartz, H. and Strumann, C., 2014. Hospital efficiency under prospective reimbursement schemes: an empirical assessment for the case of Germany. *The European Journal of Health Economics*, 15(2), pp. 175–186.

Hilgers, S., 2011. *DRG-Vergütung in deutschen Krankenhäusern: Auswirkungen auf Verweildauer und Behandlungsqualität.* Wiesbaden: Gabler Verlag.

Hodgkin, D. and McGuire, T.G., 1994. Payment levels and hospital response to prospective payment. *Journal of Health Economics*, 13(1), pp. 1–29.

Hsiao, W.C. and Dunn, D.L., 1987. The Impact of DRG Payment on New Jersey Hospitals. *Inquiry*, 24(3), pp. 212–220.

Jensen, M.C. and Meckling, W.H., 1976. Theory of the firm. Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), pp. 305–360.

Kahn, K.L., Keeler, E.B., Sherwood, M.J., Rogers, W.H., Draper, D., Bentow, S.S., Reinisch, E.J., Rubenstein, L.V., Kosecoff, J. and Brook, R.H., 1990. Comparing outcomes of care before and after implementation of the DRG-based prospective payment system. *Journal of the American Medical Association*, 264(15), pp. 1984–1988.

KHG - Krankenhausentgeltgesetz, 2016. Last amended by Art. 1 G v. 12/10/2015 I 2229

Klages, M., 2010. Personalmanagment. In J. Debatin, A. Ekkerkamp and B. Schulte, eds. 2010. *Krankenhausmanagement: Strategien, Konzepte, Methoden.* Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

Kobel, C., Thuilliez, J., Bellanger, M. and Pfeiffer, K.-P., 2011. DRG systems and similar patient classification systems in Europe. In: R. Busse, A. Geissler, W. Quentin and M. Wiley, eds. 2011. *Diagnosis-related groups in Europe - moving towards transparency, efficiency and quality in hospitals.* Maidenhead: Open University Press. pp. 37–58.

Kourie, D.G., 1976. A Length of Stay Index to Monitor Efficiency of Service to General Surgery In-Patients. *Operational Research Quarterly*, 27(4), p. 959-970.

Kunze, B. and Nagel, E., 2010. Die Auswirkungen der DRG-Einführung auf Berufsethos, Arbeitsbedingungen und Versorgungsprozesse in Krankenhäusern – Case Management als Lösungsstrategie? In: F.W. Schwartz and P. Angerer, eds. 2010. *Arbeitsbedingungen und Befinden von Ärztinnen und Ärzten. Befunde und Interventionen.* Cologne: Deutscher Ärzte-Verlag Köln. pp. 263-270.

Kwon, S., 2003. Payment system reform for health care providers in Korea. *Health Policy and Planning*, 18(1), pp. 84–92.

Lave, J. and Frank, R.G., 1990. Effect of the structure of hospital payment on length of stay. *Health Services Research*, 25(2), pp. 327–347.

Levaggi, R., 2005. Hospital health care: pricing and quality control in a spatial model with asymmetry of information. *International Journal of Health Care Finance and Economics*, 5(4), pp. 327–349.

Lindrooth, R.C. and Weisbrod, B.A., 2007. Do religious nonprofit and for-profit organizations respond differently to financial incentives? The hospice industry. *Journal of Health Economics*, 26(2), pp. 342–357.

Louis, D.Z., Yuen, E.J., Braga, M., Cicchetti, A., Rabinowitz, C., Laine, C. and Gonnella, J.S., 1999. Impact of a DRG-based hospital financing system on quality and outcomes of care in Italy. *Health Services Research*, 34(1 part 2), pp. 405–415.

Ma, C.-T.A., 1994. Health Care Payment Systems: Cost and Quality Incentives. *Journal of Economics & Management Strategy*, 3(1), pp. 93–112.

Magnussen, J. and Solstad, K., 1994. Case-based hospital financing: the case of Norway. *Health Policy*, 28(1), pp. 23–36.

McCue, M.J. and Thompson, J.M., 2006. Early Effects of the Prospective Payment System on Inpatient Rehabilitation Hospital Performance. *Archives of Physical Medicine and Rehabilitation*, 87(2), pp. 198–202.

McKillop, I., Pink, G.H. and Johnson, L.M., 2001. *The financial management of acute care in Canada: A review of funding, performance monitoring and reporting practices*. Ottawa: Canadian Institute for Health Information.

Mikkola, H. and Hakkinen, U., 2002. The effects of case-based pricing on length of stay for common surgical procedures. *Journal of Health Services Research & Policy*, 7(2), pp. 90–97.

Milgrom, P.R. and Roberts, J., 1992. *Economics, organization and management*. London: Prentice Hall.

MOHW - Ministry of Health and Welfare, 2000. *Evaluation of the 3rd-year pilot program on DRG-based payment*. Seoul: Ministry of Health and Welfare.

Moreno-Serra, R. and 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), pp. 585–602.

Müller, R., 2007. *Einfluss der DRGs auf den Entlassungszeitpunkt im Krankenhaus: Ergebnisse einer sozialwissenschaftlichen Studie*. Medizin und Gesellschaft.

Norton, E.C., van Houtven, C.H., Lindrooth, R.C., Normand, S.-L.T. and Dickey, B., 2002. Does prospective payment reduce inpatient length of stay?. *Health Economics*, 11(5), pp. 377–387.

Oberender, P. and Zerth, J., 2010. *Wachstumsmarkt Gesundheit.* In: UTB für Wissenschaft, Uni-Taschenbücher. 3rd edition. Stuttgart: UTB GmbH.

OECD - Organization for Economic Cooperation and Development, 2015. *Health at a Glance 2015. OECD Indicators.* Paris: OECD Publishing.

OECD, 2016. Health Care Utilisation: Hospital average length of stay by diagnostic categories. Latest update November 2nd 2015. [online] Available at: http://stats.oecd.org/# [Accessed February 15th 2016].

Perelman, J. and Closon, M.-C., 2007. Hospital response to prospective financing of in-patient days. The Belgian case. *Health Policy*, 84(2-3), pp. 200–209.

Pope, G.C., 1989. Hospital nonprice competition and Medicare reimbursement policy. *Journal of Health Economics*, 8(2), pp. 147–172.

Porter, M.E. and Guth, C., 2012. *Redefining German Health Care. Moving to a Value-Based System.* Springer Verlag Berlin Heidelberg.

Rau, F., Roeder, N. and Hensen, P., 2009. Zum Stand der deutschen DRG-Einführung: Erkenntnisse, Erfahrungen und Meinungen. In: F. Rau, N. Roeder and P. Hensen, eds. 2009. *Auswirkungen der DRG-Einführung in Deutschland: Standortbestimmung und Perspektiven.* Stuttgart: Kohlhammer. pp. 9–22.

Roeder, N., Hensen, P., Franz, D., eds. 2014. *Gesundheitsökonomie, Gesundheitssystem und öffentliche Gesundheitspflege*. 2nd ed. Cologne: Deutscher Ärzte-Verlag Köln.

Roemer, M.I., 1961. Bed supply and hospital utilization: a natural experiment. *Hospitals*, Vol. 35, pp. 36–42.

Rogerson, W.P., 1993. Choice of Treatment Intensities by a Nonprofit Hospital under Prospective Pricing. *Journal of Economics and Management Strategy*, 3(1), pp.7-52.

Rutz, S., 2006. *Die Einführung von Diagnosis Related Groups in Deutschland: Interessen - Anreize - erste Ergebnisse*. Lohmar-Köln: Eul Verlag.

Salkever, D.S. and Steinwachs, D.M., 1988. Utilization and case-mix impacts of per case payment in Maryland. *Health Care Financing Review*, 9(3), pp. 23–32.

Sappington, D.E., 1991. Incentives in Principal-Agent Relationships. *Journal of Economic Perspectives*, 5(2), pp. 45–66.

Schnoor, J., Daumann, F. and Kaisers, U., 2011. Demografie und (Multi-) Morbiditätsentwicklung im Krankenhaus. *Gesundheitsökonomie & Qualitätsmanagement*, 16(5), pp. 310–315.

Schreyögg, J., Tiemann, O. and Busse, R., 2005. The DRG reimbursement system in Germany. *Euro Observer*, 7(4), pp. 4–6.

Schreyögg, J., Bäuml, M., Krämer, J., Dette, T., Busse, R., Geissler, A., 2014. *Forschungsauftrag zur Mengenentwicklung nach §17b Abs. 9 KHG. Endbericht Juli 2014.* Hamburg Center for Health Economics.

Seidman, R.L. and Frank, R.G., 1985. Hospital responses to incentives in alternative reimbursement systems. *Journal of Behavioral Economics*, 14(1), pp. 155–180.

SBG V – Sozialgesetzbuch Fünftes Buch, 2016. Last amended by Art. 2 Abs. 12 G v. 17.02.2016 I 203.

Shleifer, A., 1985. A Theory of Yardstick Competition. *Rand Journal of Economics*, 16(3), pp. 319-327.

Sloan, F.A. and Valvona, J., 1986. Why has hospital length of stay declined? An evaluation of alternative theories. *Social Science & Medicine*, 22(1), pp. 63–73.

Sood, N., Buntin, M.B. and Escarce, J.J., 2008. Does how much and how you pay matter? Evidence from the inpatient rehabilitation care prospective payment system. *Journal of Health Economics*, 27(4), pp. 1046–1059.

Destatis – Statistisches Bundesamt, 2013. *Diagnosedaten der Patienten und Patientinnen in Krankenhäusern (einschl. Sterbe- und Stundenfälle).* Fachserie 12 Reihe 6.2.1 – 2012. Wiesbaden.

Destatis - Statistisches Bundesamt, 2016a. *Krankenhausstatistik.* [online] Available at: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/Krankenhaeuser/Methoden/Krankenhausstatistik.html [Accessed 31 March 2016].

Destatis - Statistisches Bundesamt, 2016b. *Pflegestatistik.* [online] Available at: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/Pflege/Methoden/Pflegestatistik.html [Accessed 1 April 2016].

Destatis - Statistisches Bundesamt, 2016c. Durchschnittliche Verweildauer in deutschen Krankenhäusern in den Jahren 1992 bis 2014 (in Tagen). [online] Available at: http://de.statista.com/statistik/daten/studie/2604/umfrage/durchschnittliche-verweildauer-im-krankenhaus-seit-1992/ [Accessed 15 February 2016].

Theurl, E. and Winner, H., 2007. The impact of hospital financing on the length of stay: evidence from Austria. *Health policy*, 82(3), pp. 375–389.

Thomas, J.W., Guire, K.E. and Horvat, G.G., 1997. Is patient length of stay related to quality of care?. *Hospital & Health Services Administration*, 42(4), pp. 489–507.

Toepfer, R., 2011. Publikationsverhalten in den Wirtschaftswissenschaften - Eine bibliometrische Analyse auf Basis der Fachdatenbank ECONIS. In: J. Warmbrunn, ed. 2011. *Die Kraft der digitalen Unordnung: 32. Arbeits- und Fortbildungstagung der ASpB*, *Sektion 5 im Deutschen Bibliotheksverband, 22. bis 25. September 2009 in der Universität Karlsruhe*. Karlsruhe: KIT Scientific Publishing. pp. 241-253.

Trivedi, P.K., 2010. *Recent Developments in Panel Models for Count Data*: *Prepared for 2010 Mexican Stata Users Group meeting*. [online] Available at: http://fmwww.bc.edu/repec/msug2010/mex10sug_trivedi.pdf [Accessed 22 April 2016].

Tsai, Y.-W., 2005. The effect of changing reimbursement policies on quality of in-patient care, from fee-for-service to prospective payment. *International Journal for Quality in Health Care*, 17(5), pp. 421–426.

vdek – Verband der Ersatzkassen, 2016. *Landesbasisfallwerte (LBFW) 2016 nach dem Krankenhausentgeltgesetz.* [online] Available at: https://www.vdek.com/vertrags partner/Krankenhaeuser/landesbasisfallwerte.html [Accessed 29 February 2016].

Vries, E.N. de, Ramrattan, M.A., Smorenburg, S.M., Gouma, D.J. and Boermeester, M.A., 2008. The incidence and nature of in-hospital adverse events: a systematic review. *Quality & Safety in Health Care*, 17(3), pp. 216–223.

Wang, K., Li, P., Chen, L., Kato, K., Kobayashi, M. and Yamauchi, K., 2010. Impact of the Japanese Diagnosis Procedure Combination-based Payment System in Japan. *Journal of Medical Systems*, 34(1), pp. 95–100.

Winkelmann, R., 2008. *Econometric Analysis of Count Data*. 5th edition. Berlin, Heidelberg: Springer.

Wooldridge, J.M., 2009. *Introductory econometrics: A modern approach*. 4th edition. Mason, OH: South Western Cengage Learning.

Yeh, H.W., Gajewski, B., Mukhopadhyay, P. and Behbod, F., 2012. The Zero-truncated Poisson with Right Censoring: an Application to Translational Breast Cancer Research. *Statistics in Biopharmaceutical Research*, 4(3), pp. 252–263.

Yilmaz, E. and Raynaud, D., 2013. The influence of social deprivation on length of hospitalization. *The European Journal of Health Economics*, 14(2), pp. 243–252.

Zhang, J., 2010. The impact of a diagnosis-related group-based prospective payment experiment: The experience of Shanghai. *Applied Economics Letters*, 17(18), pp. 1797–1803.