HWI HAMBURG
HOCHSCHULÜBERGREIFENDER STUDIENGANG
WIRTSCHAFTSINGENIEURWESEN HAMBURG
Universität Hamburg
Hochschule für Angewandte Wissenschaften Hamburg
Helmut-Schmidt-Universität Hamburg

MASTERARBEIT

von

Jakob Everding

Matrikelnummer: 6376014

How does PEPP affect the German hospital market? First insights from a systematic literature review and an econometric analysis.

1. Gutachter:	Prof. Dr. Jonas Schreyögg
2. Gutachter:	Dr. Matthias Bäuml
Bereich:	Management im Gesundheitswesen
vorgelegt am:	22.11.2016

Content

Tables

Figures

Abstract

Background: A new prospective reimbursement and patient classification system for psychiatric and psychosomatic health care (PEPP) will be introduced in Germany as of 2017. In 2009, PEPP was officially announced and the development of the system began. PEPP has been introduced on a voluntary, budget neutral basis in 2013.

Objective: This thesis examines the development of hospital admissions as well as the intensity of care prior and subsequently to the announcement and optional introduction of PEPP. I investigate the effects of altered financial incentives due to changes in the reimbursement mechanism on the German hospital market.

Methods: A combination of qualitative and quantitative methods has been used. First, a systematic literature review was carried out. Building on the findings of international scholars, econometric models were set up and tested using mainly ordinary least squares regressions. Furthermore, different samples and a fixed effects approach were employed as part of sensitivity analysis to test for the models' robustness. The development of important proxy variables for the intensive and extensive margin of mental health care was analyzed.

Results: The results of this thesis show that the intensity of care increased subsequently to the announcement of PEPP in 2009. In particular, the number of treatments per case measured by OPS codes increased by approximately 260 percent more in psychiatric and psychosomatic wards than in somatic wards subsequently to the announcement of PEPP. Furthermore, graphical data analysis also shows changes in the extensive margin. Hospital admissions increased more in psychosomatic and psychiatric wards than in somatic wards after the announcement of PEPP. However, this effect was no further analyzed empirically because of the scope of this thesis.

Conclusion: It appears likely that the hospital market responds to PEPP along the intensive and extensive margin. However, no causal relationship could be established yet. More data, especially from periods after the binding introduction of PEPP in 2017, is needed. Future research is necessary to study the effects of PEPP on the hospital market in more detail. Policy makers might profit from such research given that particularly in the beginning, the system's grouping algorithm and cost weights need to be continuously adapted, for instance.

1 – Introduction

Health insurance is compulsory in Germany since 2009. Approximately 90 percent of the German population are accordingly covered within the statutory health insurance (SHI) system (Schmitz, 2013). The remaining are civil servants, self-employed, soldiers or individuals whose earnings exceed a set threshold allowing them to buy private health insurance. Basically, the whole German population is fully covered either within the SHI system or private health insurance schemes. (Schmitz, 2013; Wörz and Busse, 2005)

In 2005, a diagnosis-related group system, the G-DRG system, was implemented for the classification of patients and reimbursement of health services in the hospital sector in Germany except for psychiatric and psychosomatic health care (Wörz and Busse, 2005). In 2009, the §17d Krankenhausfinanzierungsgesetz (KHG) was enacted in Germany, determining the development of a new patient classification and payment system specifically designed for psychiatric and psychosomatic institutions in the hospital sector, the "Pauschalierendes Entgeltsystem für psychiatrische und psychosomatische Einrichtungen" (hereafter referred to as PEPP)[1] (Geissler and Quentin, 2010). In the following, the term ward is used with respect to this range of institutions, including inpatient wards just as semi-stationary day clinics, for instance. Within the German health care system, PEPP directly affects all psychiatric and psychosomatic wards offering inpatient and semi-stationary health services (GKV-Spitzenverband, 2009).

Prior to the binding introduction, which was originally scheduled for 2015 (see footnote 1) and will very probably take place as of 2017, PEPP has been introduced on a voluntary, budget neutral basis in 2013 (Busse et al., 2013, p. 282; Geissler and Quentin, 2010). Budget neutrality will be maintained until 2019. During the voluntary trial phase, the PEPP systematics and grouping algorithms are adapted every year to incorporate new insights provided by the data of participating hospitals. Thus, pinpointing the effects of the introduction of PEPP as a whole to a single year is most likely not possible. Consequently, dynamic effects on the hospital market which will be observed over a period of several years rather than a single structural break are expected, presumably.

[1] The design and development of PEPP was announced in the 2009 legislation, followed by the Gesetz zur Einführung eines pauschalierenden Entgeltsystems für psychiatrische und psychosomatische Einrichtungen (PsychEntgG) in 2012. This second act was again changed in 2014 and regulates the introduction of PEPP. Nonetheless, I refer to the 2009 legislation as the starting point of PEPP

The effect of changes in reimbursement and financial incentives in general on health care markets has interested scholars internationally so that a wide range of research is available on different aspects of this topic to date, primarily from the USA (e.g. Clemens and Gottlieb (2014), Dafny (2005), Finkelstein (2007)). Overall, this research points out that the behavior of health care providers, outpatient and inpatient, is affected by the individual financial incentives created by policy reforms. As reflected by the studies' partly diverging and contradictory findings, the health providers' response seems to depend on numerous factors such as the health system's country specific characteristics as well as on the composition and design of incentives. For instance, Dafny (2005), among others, finds that intensity of supplied health care is correlated with average insurance coverage. In consequence, if two countries differ in that respect, generalizing evidence from one to the other seems to be questionable. Also, the effect of market-wide changes of incentives or prices on health systems can differ substantially from the effect of small-scale reforms (Finkelstein, 2007). Although there are also high quality studies with focus on general hospitals available from Europe (e.g. Barbetta et al. (2007) and Melberg et al. (2016)) and Germany in particular (e.g. Herwartz and Strumann (2014)), transferring insights from past reforms to the current case of PEPP is thus not entirely possible due to underlying differences in the studied reforms and health sector characteristics. Another restricting factor regarding generalizability of results is that there exist substantial differences between mental health care and other, i.e. physical or somatic, health care (Frank and McGuire, 2005, p. 895). To date, there is hardly any research on how changes in financial incentives as part of policy reforms affect the supply of mental health care or, more broadly, on drivers of patient health in psychiatric or psychosomatic wards in Germany (Haas et al., 2013). Furthermore, before implementation, such health care reforms may practically only be assessed on a speculative rather than evidence informed basis because of manifold possible combinations and characteristics of incentives (Ellis and McGuire, 1996; Wörz and Busse, 2005).

Therefore, relatively little is known about PEPP's possible consequences on the German hospital market yet. This thesis aims at adding to the discussion by examining the effects of PEPP on the volume of mental health care provided in the timeframe from 2006 to 2014 and thus covering the official announcement of PEPP in 2009 as well as the start of the phase-in period in 2013. However, as the mandatory introduction of PEPP is scheduled for 2017, there is obviously no data subsequently to this available yet. I try to describe how hospitals prepare for the price shock induced by PEPP while the system's new financial incentives are tested in

hospital wards on a voluntary basis. Due to data restrictions and limited opportunities for empirical analysis as explained in the following sections, I will focus on qualitatively discussing and presenting important related literature in detail and thereby provide the opportunity to put my empirical findings and interpretations into perspective with current research.

This thesis is structured as follows. Section 2 gives some background information on the old and new payment system for psychiatric and psychosomatic wards in Germany. A systematic review of relevant literature is presented in section 3. These parts also lay the basis for the assumptions analyzed empirically. Section 4 provides an overview of the data followed by a brief description of the empirical approach in section 5. Results are presented in section 6 and discussed in section 7. Eventually, section 8 concludes.

2 – Background

Prior to PEPP, psychiatric and psychosomatic hospital cases were reimbursed on the basis of the so called Bundespflegesatzverordnung (BPflV), which was also applied for the reimbursement of general inpatient health care prior to the G-DRG system. Reimbursement for psychiatric and psychosomatic wards as regulated by BPflV was principally day-based. (Bundesministerium für Gesundheit, 2016) There was no adjustment of reimbursement height based on e.g. patient characteristics or supplied treatments (Geissler and Quentin, 2010). However, staffing of psychiatric and psychosomatric wards was regulated by the psychiatry personnel directive (Psychiatrie-Personalverordnung, PSYCH-PV) and thus, hospitals did not have any financial incentives to alter the number or ratios of medical staff. Instead, the appropriate number of nurses, psychiatrists and other medical experts was determined using the PSYCH-PV classification system. Under PEPP, the PSYCH-PV classification system was applied and refined in order to assign patients into cost homogenous groups (Busse et al., 2013, pp. 352–355; GKV-Spitzenverband, 2009). Therefore, no response of hospitals to incentives created by PEPP on the number of physicians or nurses is expected.

As the successor of the BPflV reimbursement mechanism, PEPP has been introduced with the intention to increase cost efficiency and quality of mental health care (Bundesministerium für Gesundheit, 2016; GKV-Spitzenverband, 2009). One of the main goals was to develop a reimbursement system which is prospective, comprehensive, and performance oriented, as formulated in §17d KHG. Similar developments concerning health policy reforms took place in numerous other countries over the past years (Haas et al., 2013; Pletscher, 2016). However, in contrast to the development of the G-DRG system, which is based on the Australian DRG system, no existing reimbursement mechanisms from other countries could be identified as appropriate examples to build upon. The legal mandate was hence regarded as particularly challenging. (GKV-Spitzenverband, 2009) Moreover, this underpins the difficulty also for scholars to predict the providers' responses to the new reimbursement's incentives. The PEPP system was developed by the self-governing bodies at federal level, i.e. the German Hospital Federation (Deutsche Krankenhausgesellschaft, DKG), the National Association of Statutory Health Insurance Funds (Spitzenverband Bund der Krankenkassen, GKV-Spitzenverband), and the Association of Private Health Insurances (Verband der Privaten Krankenversicherung, PKV-Verband). By means of the Institute for the Hospital Reimbursement System (Institut für das Entgeltsystem im Krankenhaus, InEK), which was jointly founded by all three federal self-

governing bodies in 2001, grouping algorithms were developed and are refined every year in the sense of a "learning system", similar to the process of the G-DRG system. Different stakeholder groups such as professional associations, e.g. the German Association for Psychiatry, Psychotherapy and Psychosomatics (Deutsche Gesellschaft für Psychiatrie, Psychotherapie, Psychosomatik und Nervenheilkunde, DGPPN), are involved in this process and may propose particular changes of the current system to the InEK (Institut für das Entgeltsystem im Krankenhaus, 2015a). In principal, propositions may be submitted by everyone (Institut für das Entgeltsystem im Krankenhaus, 2015d). For the years 2014 to 2016, the InEK received more than 200 propositions, from single hospitals or hospital associations (Institut für das Entgeltsystem im Krankenhaus, 2016). The high number of participation reflects the public attention PEPP raised among stakeholder groups right from the start. As described by Geissler and Quentin (2010), the federal self-governing bodies are mainly supportive, while professional medical and patients' associations like the DGPPN and the Initiative Patients with Mental Illness (Aktion Psychisch Kranke) are strongly opposed concerning PEPP. However, also the three members of the federal self-governing bodies faced major disagreements during the process of developing PEPP. For instance, they failed to agree on a reimbursement scheme for the year 2013. Due to this, that year's PEPP system had to be put into effect by substitute performance of the Federal Ministry of Health (Bundesministerium für Gesundheit, BMG), based on propositions developed by the InEK (see Bundesministerium für Gesundheit (2012)). I give some broad examples in the following as possible reasons for the debate and dispute.

All different prospective payment systems share the basic idea of reimbursing health care providers with a fixed price which is designed to reflect the average resource consumption of a group of patients (Cutler, 1995; Pletscher, 2016). While hospitals realize profits if the treatment costs for a case are lower than the received reimbursement, they will face losses if costs exceed the case's reimbursement. Like DRG systems, PEPP therefore induces a yardstick competition in the sense of a relative, performance based reimbursement system as described by Shleifer (1985). Under prospective reimbursement, hospitals are incentivized to reduce costs of care below a threshold which is determined by the market's aggregate cost structure and performance (Besstremyannaya, 2013). Hospitals thus face a substantial financial risk and pressure under prospective payment compared to other reimbursement schemes (Herwartz and Strumann, 2014).

PEPP is basically a prospective payment system which has been aligned to the properties in German psychiatric and psychosomatic wards. For every group as defined by the classification system, the InEK calculates per-diem cost weights which are the basis of the final reimbursement. The basic cost weight is then once more adjusted based on patient characteristics and supplied treatments. For instance, supplementary per-diem reimbursement is granted for patients with intensive treatment needs such as individual supervision for adults (Institut für das Entgeltsystem im Krankenhaus, 2015a). Currently, reimbursement amounts are obtained by multiplying the final cost weights with a hospital individual base rate. For treatments which can be pinpointed to a specific application period, supplementary reimbursement may be added to this amount. The hospital individual base rates are scheduled to be replaced with federal state (Bundesland) level base rates, starting as of 2019. (Institut für das Entgeltsystem im Krankenhaus, 2015a) This reimbursement reform is likely to once again mark a significant price shock, differing in magnitude for the single hospitals within each state. As summarized by Institut für das Entgeltsystem im Krankenhaus (2015b), each case is classified by a PEPP-code, which consists of the following four alphanumeric components. First, a letter indicates whether the concerned patient is an inpatient (P) or semi-stationary (T) case. Second, another letter indicates the ward in which the patient was treated. The algorithm allows for differentiation between general psychiatry (letter A), psychosomatic ward (P), child and adolescent psychiatry (K) and codes everything other than that as an error (F). Furthermore, the second letter may also be assigned the number 0 instead of a letter, indicating what is termed a "Prä-PEPP" (pre-group). Again, the presence of errors in the coding algorithm emphasizes the need for further adjustment of the learning system (Institut für das Entgeltsystem im Krankenhaus, 2015a). Third, the first two letters are followed by two numbers which indicate the patient's diagnostic group. Information on diagnoses are retrieved using the German version of the tenth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10-GM). Finally, another letter indicates the patient's resource consumption on a four-point scale ranging from highest (A) to lowest (D) resource consumption. If differentiation in resource consumption is not provided for, the letters A to D are replaced with a Z instead. Altogether, resource consumption considers the patient age as well as applied procedures, which are reported by means of the current version of the Operations- und Prozedurenschlüssel (OPS codes). In accordance with the PEPP algorithm, an inpatient case (P) treated in general psychiatry (A) with psychological disturbances and behavior disturbances by psychotrope substances as the diagnostic category (02) and high treatment intensity (B) will thus be classified into the PEPP-group PA02B eventually, for

example. For a detailed description of the PEPP classification algorithm, please see Institut für das Entgeltsystem im Krankenhaus (2015b).

In its purest form, prospective reimbursement creates incentives for patient selection and case shifting, which might lead to under-treatment of cost intensive patients with severe treatment needs (Sharma, 2009). Analogously, this incentive might lead to over-treatment of patients whose treatments cost less than the granted reimbursement and who are therefore considered as profitable by hospitals. (Melberg et al., 2016) However, PEPP reimbursement takes into account the length of stay and treatments supplied to patients during their stay. Yet, as this marks a change compared to the old reimbursement mechanism, the related incentive might have changed as well. As a result, professional medical and patients' associations fear decreasing quality of mental health care due to PEPP. However, it must be noted that similar concerns arose in the context of the introduction of the G-DRG system (Geissler and Quentin, 2010). While the marginal price effect under pure prospective payment, i.e. the incentive to extend the treatment of admitted patients, is zero (Cutler, 1995), it is expected to be positive under PEPP reimbursement due to this modification. This assumption is also supported by Dafny (2005), who finds that the marginal price effect remains positive if supplied procedures are relevant for reimbursement calculation. Similarly, Gilman (2000) argues that if supply of procedures is relevant for the calculation of reimbursement height, hospitals are expected to increase intensity of care in form of supplied procedures per case compared to reimbursement mechanisms with zero marginal price effect. If, on the other hand, the marginal price effect still decreases from pre- to post-treatment, i.e. the reimbursement changes, intensity of care is expected to decrease, too. Yet, this decline will not be as steep as with the marginal price effect being eliminated completely. It is accordingly difficult to predict the marginal price effect of PEPP on the German hospital market. This ambiguous relationship is also underpinned by Pletscher (2016) who states that the elimination of the marginal price effect does not necessarily lead to a decreased average length of stay for inpatient mental health care. Furthermore, hospitals might also respond to such incentives by upcoding, i.e. illicitly classifying patients and their treatments in a way that provides them a higher reimbursement. This would increase the financial burden of payers as treatment costs truly remain equal for providers (Dafny, 2005; Melberg et al., 2016). Again, adverse outcomes might eventuate from PEPP. Generally, hospitals are found to increase attention to coding procedure after the introduction of prospective payment systems (Serdén et al., 2003). Hence, evaluating how PEPP affects the intensity of care in practice is highly important.

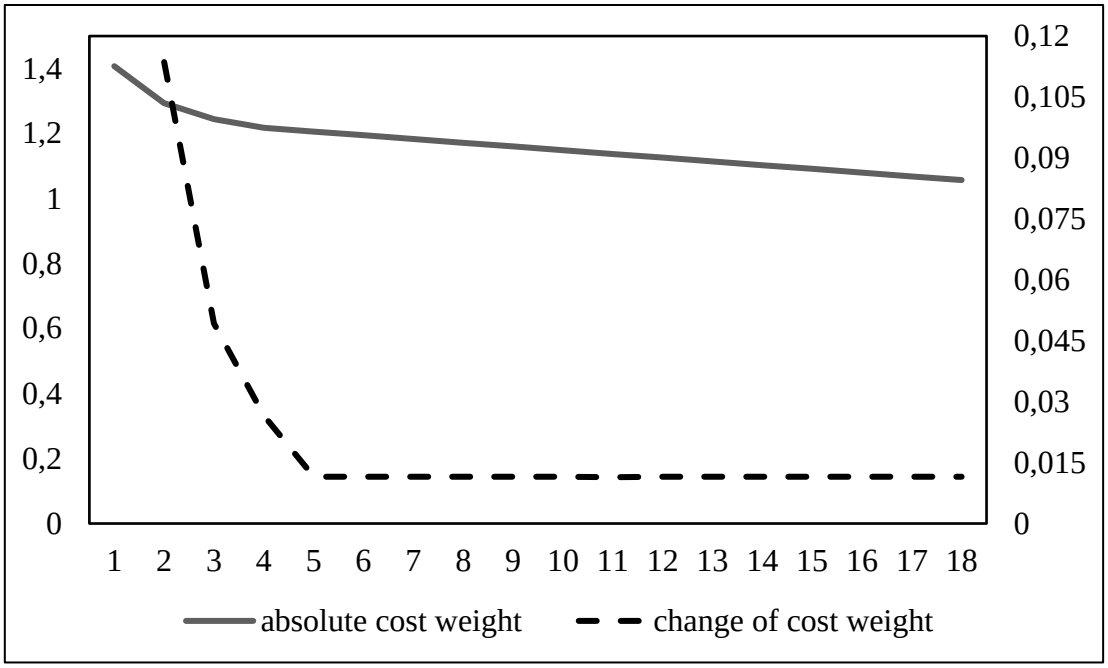

FIGURE 1 – RELATIONSHIP OF COST WEIGHT AND LENGTH OF STAY

Source: Own illustration based on Institut für das Entgeltsystem im Krankenhaus (2015c)
Notes: The figure exemplarily describes the decrease in cost weight over time. For exemplary reasons, the PEPP group PA02B from 2016 is used. The x-axis represents length of stay in days.

However, prospective payment systems do not only alter the marginal price effect, but also the average price effect, i.e. the incentive for hospitals to expand the number of patients admitted in the first place (Hodgkin and McGuire, 1994). Under prospective payment, the average price effect is expected to be positive (Dafny, 2005). This effect is also expected for PEPP. Yet, as length of stay is relevant for the calculation of PEPP reimbursement as a degressive factor, the average price effect will decrease with increased length of stay, too. For a graphical representation of this relationship, see Figure 1. As displayed, the 2016 per-diem cost weight of the PEPP-group PA02B decreases from 1.4073 for a one-day length of stay to 1.1488 for a ten-day length of stay. After the fifth day, the change of cost weight remains constant at -0.0115 per day. Depending on the hospital's individual cost structure and base rate, the marginal reimbursement received for an additional day may fall below the hospital's marginal treatment costs. In this case, hospitals will realize losses for each additional day the patient is not discharged, given that no supplementary reimbursement regulations apply.

In summary, the introduction of PEPP might be associated with decreased public medical spending, if upcoding and related adverse responses can be successfully prevented. While assessing the effect of PEPP on costs of health care is open to question, another interesting and

similarly ambiguous issue is PEPP's implication for quality of mental health care and patients' health outcomes. This aspect in particular is presumably most concerning for the general public, giving a brief explanation for the described debate. Although increased transparency of the reimbursement mechanism is aspired (see e.g. Institut für das Entgeltsystem im Krankenhaus (2015a)), the public perception appears to be rather the opposite (Geissler and Quentin, 2010), further adding to the problem and concerns.

Therefore, the main objectives for empirical analysis in this thesis are (i) to describe the development of the number of cases prior and subsequently to both the announcement (2009) and optional introduction (2013) of PEPP and (ii) to describe the development of the number of OPS codes per case. Albeit all methodological limitations, I try to answer how the hospital market and particularly the volume of health care for psychiatric and psychosomatic wards in Germany would be like today if PEPP was not introduced.

3 – Systematic literature review

This systematic literature review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 checklist (see Moher et al. (2009)). It was performed in September and October 2016. In the following, the structure and process of the review is described in detail.

3.1 – Study selection

As eligibility criteria, characteristics for inclusion and exclusion were set up. Because the rationale of this review was to find studies with an underlying research question similar to the one given in this study, most inclusion and exclusion criteria were derived from the research question.

Correspondingly, publications were included which (i) considered changes in prices or incentives under (ii) prospective payment for (iii) hospitals or physicians, and (iv) analyzed primarily outcomes such as intensity of care or admissions. Furthermore, the articles had to (v) be written in English or German and (vi) utilize mainly econometric or similar economic methods. On the contrary, publications with the following characteristics were excluded from the literature review. (i) Duplicates, (ii) articles without full-text access, and (iii) non peer-reviewed publications.

3.2 – Information sources

The focus of the literature review were economic articles, in alignment with the approach of this thesis. Therefore, EconLit and Business Source Complete in EBSCOhost and the Web of Science Core Collection were selected for systematic online data base searching. The search was complemented by selective use of Google Scholar and scanning of websites of several European and German health authorities. However, as presented in Figure 2, no academic literature was found on these websites. Instead, they mainly contained valuable background information.

3.3 – Search strategy

Starting with partitioning the research question and analyzing its components, the seven groups (i) mental health terms, (ii) setting, (iii) incentive type, (iv) outcomes, (v) methods, (vi) associated terms, and (vii) system classification were identified. Afterwards, similar terms were found using mainly the online services of the Oxford English Dictionary[2] as well as the online services of the Duden[3], a German dictionary. However, German search terms were limited to the group of mental health terms in the final configuration of the search algorithm. These terms were then assigned to the different categories. Finally, combinations of the terms and search categories presented in Table 1 have been used for data base searching.[4]

TABLE 1 – TERMS FOR SEARCH ALGORITHM

Mental	Setting	Incentive Type	Outcomes	Methods	Associated Terms	System Classification
psychiatr*	hospital	change*	*admission	economic*	impact*	drg
psychosom*	health insurance	incentive	cases	econometric	respons*	patient classification
mental	*patient	payment	introduction	difference-in-difference	effect	diagnosis-related groups
		lump-sum	outcome			
		flat rate				
		fixed rate				
		cost				

Source: Own illustration

[2]http://www.oed.com; e.g. psychiatry is defined as "The branch of medicine concerned with the causes, diagnosis, treatment, and prevention of mental illness." (Oxford English Dictionary)

[3] http://www.duden.de

[4] Searches included for instance the following combination of search terms: (hospital OR health insurance OR *patient) AND (incentive OR payment OR lump-sum OR fixed rate) AND (*admission OR introduction OR outcome) AND (economic* OR econometric OR difference-in-difference) AND (impact* OR respons* OR effect); the combination of search groups of this search was hence: setting AND incentive type AND outcomes AND methods AND associated terms

3.4 – Data collection

All publications identified through data base searching were subsequently assessed manually. The process was performed by one researcher and is documented in detail in Figure 2. First, titles, abstracts and key words of the publications were screened using the criteria as defined above. If no clear decision whether to include or exclude a paper from the review was possible on the basis of the title, abstract and key words, the publication's full text including its references was analyzed. After this stage of screening, the full-text articles of all publications preliminarily selected for inclusion were assessed for eligibility. Of the 41 publications identified at this stage, 17 were again excluded after reviewing the full-text. These 17 publications were mainly excluded because their focus was not closely related to the objectives of the thesis.

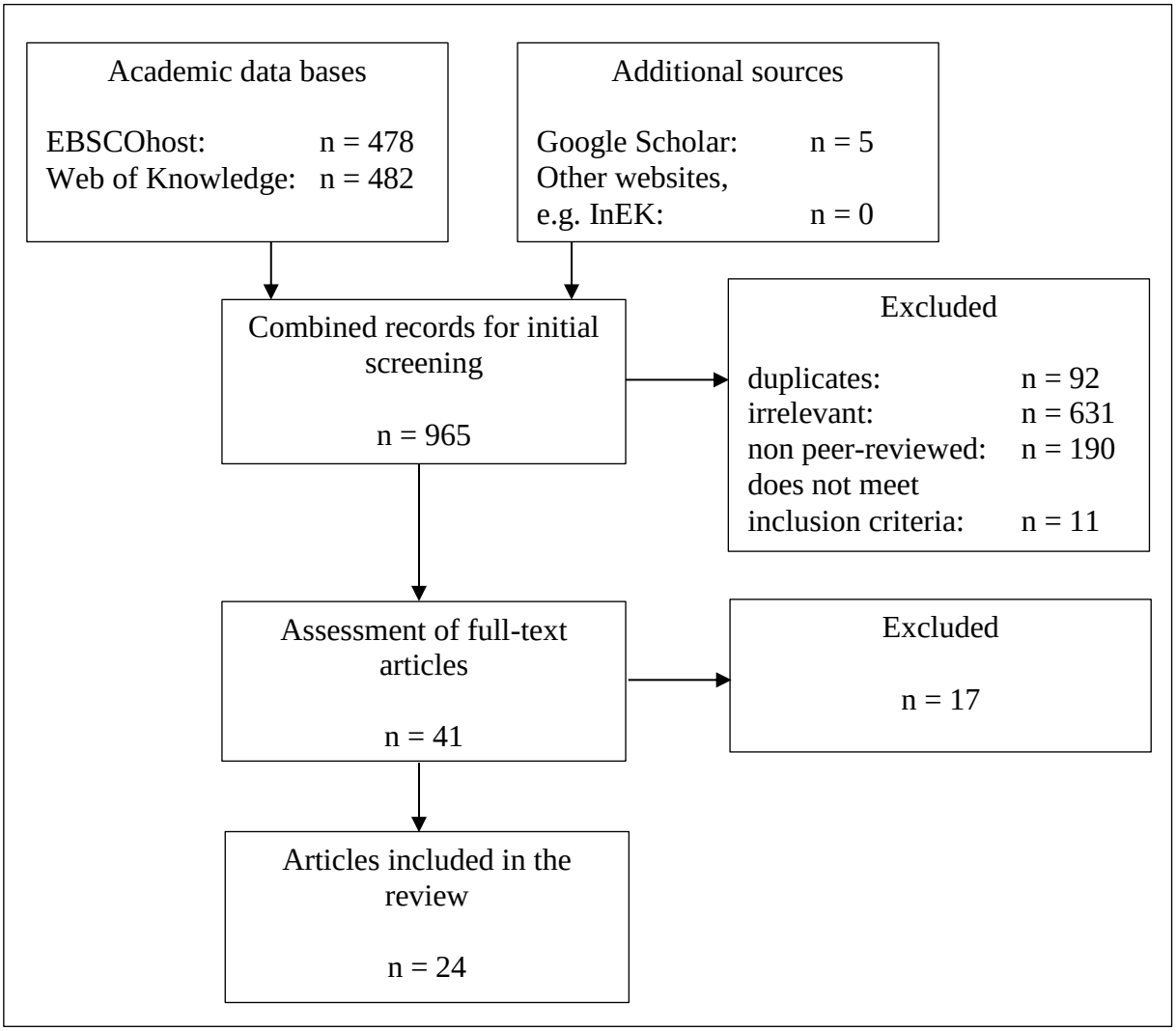

FIGURE 2 – PROCESS DIAGRAM OF LITERATURE REVIEW

Source: Adapted from Moher et al. (2009)

3.5 – Discussion of results

The results of the literature review are presented in Table 3. Here, all 24 included articles are summarized briefly. Key information concerning objectives, methodological approaches, limitations and main findings were elaborated from all of these articles.

All 24 articles included address some kind of health policy reform, which evoked changes of incentives for different players in the health care market. With 13 out of 24 included articles (54 percent), most of the evidence available comes from the USA, followed by Germany with four articles (approximately 17 percent) as presented in Table 2. In total, the literature review's results contain evidence from nine different countries, four of which are European.

TABLE 2 – DISTRIBUTION OF COUNTRIES REPRESENTED IN LITERATURE REVIEW

USA	Germany	Italy	China	Japan	Norway	Australia	Switzerland	Total
13	4	2	1	1	1	1	1	24
0.540	0.167	0.083	0.042	0.042	0.042	0.042	0.042	1

Source: Own illustration

Considering the methodological approach, 22 articles (91.67 percent) used parametric methods such as ordinary least squares (OLS) regressions. Besides, non-parametric methods such as data envelopment analysis, e.g. most used for analyzing changes in efficiency, were used in five articles (20.08 percent). Altogether, difference-in-differences techniques were mentioned in ten of the 24 articles (41.67 percent) as methods used for studying the effect of reimbursement changes. (see Table 3)

Furthermore, the articles assessed the change of patients' health outcomes (N=16, 66.67 percent), i.e. the quality of provided health care, hospital efficiency and performance (N=7, 29.17 percent), adverse outcomes (N=3, 12.50 percent) as well as the development of the number of cases and wards (N=4, 16.67 percent), i.e. the extensive margin. Basically, the included literature has predominantly analyzed effects of health policy reforms on the intensive margin rather than on the extensive margin. The presented numbers add up to more than the number of included articles simply due to the fact that some studies focused on more than one of the above categories.

For the case of outpatient care in the USA, Clemens and Gottlieb (2014) find that the intensity of health care increases in response to increased reimbursement rates. Also assessing outpatient health care, Schmitz (2013) finds similar evidence from Germany, stating that physicians'

response to financial incentives mainly affects the intensive rather than the extensive margin. That is, the number of cases is hardly affected by price changes.

Consistent also for inpatient mental health care, prospective reimbursement did not affect the number of cases shortly after its introduction in the USA as observed by Norton et al. (2002). Yet, they argue that prospective payment does in theory create an incentive for hospitals to alter the number of cases. Furthermore, they do not find evidence for changes in readmission rates as a measure for quality of care in response to prospective payment. Clemens and Gottlieb (2014) do not observe changes in the number of outpatient cases in response to altered financial incentives, too. In contrast to this, Cutler (1995), Finkelstein (2007) and Besstremyannaya (2013) present proof from the USA and Japan for the exact opposite relationship. However, according to Cutler (1995), readmission rates primarily increase because of coding practice and not as a result of truly increased morbidity. Thus, these findings does not necessarily have to be contradictory and might be explained by different model designs. However, Finkelstein (2007) calculates that Medicare prospective payment is associated with a 32 percent increase of hospital admissions in the first five years subsequently to its introduction.

Analyzing short-term mortality, Cutler (1995) additionally finds that the quality of care decreases subsequently to the introduction of prospective reimbursement. This finding is supported by Shen (2003), who studies the effects of prospective reimbursement on mortality specifically for patients with acute myocardial infarction (AMI) in the USA. In contrast to this, for patients with AMI, congestive heart failure (CHF) or chronic obstructive pulmonary disease (COPD) in Italy, Cavalieri et al. (2013) identify an increase of 30-day mortality in regions where reimbursement is predominantly prospective. Given that their research design does not allow to establish an explicit relationship between the reimbursement mechanism and short-term mortality and due to an observation period of only two years, this finding should be interpreted carefully. Yet, introducing prospective reimbursement does apparently not have any effect on long-term mortality (Cutler, 1995; Shen, 2003). Also studying AMI, although not related to prospective payment in particular, Schreyögg and Stargardt (2010) find a negative relationship of treatment costs and mortality rates for the case of Germany.

As outlined by Cotterill (2008), length of stay as another common measure for the quality of general health care is found to be longer in specialized psychiatric institutions than in other settings. He furthermore finds that the development of the number of admitted patients varied by age group. However, this finding cannot be attributed to effects of prospective payment because of restrictions of the chosen research design. Nevertheless, interpreting this finding, he concludes that this can be explained by different growth of these age groups' population.

Moreover, the response of hospitals to price changes is very high among the subgroup of patients with mental illnesses as a primary diagnosis compared to other patient groups (Melberg et al., 2016). Yet, the actual magnitude of the price change determines whether there will be effects e.g. on quality of care in the first place (Finkelstein, 2007; Pletscher, 2016). For instance, Kaestner and Guardado (2008) assess the effects of a 10 percent change in reimbursement level in the USA and find that quality of care still remains unaffected.

Also concerning quality of care in the USA, Gilman (2000) finds that under prospective reimbursement systems which, similar to the PEPP system, take supplied treatments into account, the average intensity of care decreases. This behavior, which Ellis and McGuire (1996) describe as "moral hazard" in the context of mental health care in the USA, in turn leads to decreased length of stay. Supporting this finding, Sharma (2009) provides evidence from general hospitals in Australia. In fact, she finds that prospective reimbursement is associated with decreased length of stay among high outlier cases. Furthermore, she states that this reduction in length of stay is due to improved quality of care. In turn, she observes an increase in the number of inlier cases, especially for diagnoses with high waiting times.

However, Haas et al. (2013) suggest that when assessing quality of particularly mental health care, the traditionally utilized measures such as length of stay and readmission rates are widely inappropriate as they do not reflect the specific characteristics and needs of patients with mental illnesses. Additionally, Dafny (2005) argues that length of stay is an inappropriate measure for quality of care in general as it may be both positively or negatively correlated with quality of care. Further examples for the limitations of length of stay as a measure for quality of care are presented by Pletscher (2016) who finds that it varies significantly within cost groups in inpatient mental health care. In addition, Frank and McGuire (2005, p. 896) argue that mental illnesses are frequently chronic conditions so that length of stay is expected to be biased among these patients. Jian and Guo's (2009) findings support these arguments, too. They show that the introduction of prospective reimbursement did not affect length of stay in inpatient mental health care in China. Yet, Norton et al. (2002) present strong contrasting evidence. Their empirical analysis clearly suggests that length of stay in inpatient mental health care is affected by changes in reimbursement.

Jian and Guo (2009) furthermore state that per diem reimbursement is associated with increased medical spending. This finding is consistent with evidence from the USA by Finkelstein (2007), who calculates that the introduction of the Medicare prospective payment system increased public medical spending by 37 percent. Considering hospital efficiency and performance, Meltzer et al. (2002) observe that expenditure per case decreases with increased competition.

Given that prospective payments aim to simulate a competitive market in the sense of a yardstick competition, their finding might indicate that prospective reimbursement decreases medical spending due to decreased costs per case. Yet, increased medical spending is found not to affect patients' health outcomes (Duggan, 2000). In addition, Haas et al. (2013) show that decreased medical spending decreases health outcomes of mentally ill patients in Germany. However, using a cost-benefit-calculation based on US data, Shen (2003) shows that little reductions in quality of care yield high financial savings. Thus, cost reductions might be reasonable depending on the initial level of quality of care. Moreover, she finds that competition is mainly expressed by quality related outcomes, as prices are fixed. Increases in cost and technical efficiency subsequently to the introduction of prospective reimbursement is observed by Besstremyannaya (2013), while Herwartz and Strumann (2014) and Barbetta et al. (2007) present evidence for declining technical efficiency as a result of prospective reimbursement. Again, this ambiguous relationship due to the seemingly contradictory findings could as well be explained by differences in the studied reimbursement mechanisms. This would once more emphasize the difficulty of generalizing findings concerning reimbursement reforms. In order to counteract low generalizability, the scope of the literature review is rather narrow. For example, there are substantial differences in the behavior of skilled nursing-facilities compared to other providers of health care so that such providers were not included in the review (White and Nguyen, 2011).

Referring to a rather methodological issue, Norton et al. (2002) describe the importance of correct empirical model specification. They argue that measuring the effect of the price change due to the introduction of a prospective reimbursement system by a simple model with only a single dummy variable jointly for the marginal and the average price effect will likely fail to study either effect, especially if only one of the effects is truly statistically significant. Previously, the same issue was already identified by Ellis and McGuire (1996). While some studies do not follow this proposition, there are several findings which address either the marginal or the average price effect in particular. For instance, Dafny (2005) states that prospective reimbursement eliminates the marginal price effect if the provided treatments are not relevant for reimbursement calculation. Moreover, Pletscher (2016) examines that the magnitude of the marginal price effect varies across diagnostic-groups. Cutler (1995) assumed that the magnitude of the marginal price effect is equal across groups, which was in turn criticized as a major shortcoming by Dafny (2005). Assuming such variation in the true marginal effect under prospective reimbursement, an incentive for hospitals to participate in patient-selection, i.e. cream skimming, is created (Pletscher, 2016). The same incentive might

also be caused by increased competition (Herwartz and Strumann, 2012). In consequence, adverse outcomes might occur if hospitals respond to such incentives, eventually leading to under- or over-treatment of patients (Melberg et al., 2016). For instance, Ellis and McGuire (1996) attribute a 3.0 day decrease in length of stay to patient selection caused by the introduction of prospective reimbursement. They call this response "practice-style effects". Additionally, Schmitz (2013) presents evidence for changes in the case-mix of hospitals in response to altered financial incentives. This might be an indicator for the presence of patient selection activities. However, Sloan et al. (1988) find the same for the development of the case-mix for a different reimbursement setting but no evidence for patient selection or case-shifting at the same time. Another adverse outcome that might occur is upcoding, which increases the financial burden of payers as reimbursement increases while costs remain unchanged (Cutler, 1995; Melberg et al., 2016). For example, Cutler (1995) presents evidence suggesting that upcoding and related coding practices account for most of the observed changes in the marginal price effect rather than true changes in the population's morbidity. As argued by Cotterill (2008), upcoding might be even more present in mental than in somatic health care because differentiating between mental illnesses is highly challenging. When in doubt, hospitals might thus choose the more profitable diagnosis, mostly without risking fines. Moreover, Cutler (1995) finds that male patients are more likely to experience adverse outcomes than female patients. Thus, controlling for patient level characteristics might be reasonable. While Shen (2003) argues that this might lead to an exaggeration of the (increased) statistical significance, Ellis and McGuire (1996) also hold the opinion that biased estimates might occur without controlling for patient level characteristics because they might be correlated with other variables, e.g. at hospital level. Strong evidence for upcoding as a common behavior of hospitals is provided by Dafny (2005). In particular, she finds that especially for-profit hospitals respond to the observed introduction of prospective payment in the USA mainly by upcoding, after weighing the thereby increased profits against the risks of fines and even criminal charges. Furthermore, she finds that upcoding is in fact most present in cost groups for which the incentive for such illicit practices is highest.

Studying how ownership and other hospital characteristics affect their behavior towards financial incentives, Barbetta et al. (2007) find that public hospitals respond slower to changed reimbursements than their private, both not-for-profit and for-profit, counterparts. This is basically in line with Dafny's (2005) finding of upcoding being most present in private for-profit hospitals. Additional supporting evidence on this is contributed by Duggan (2000), who observes that patient selection is less present in public hospitals. In alignment with this finding,

Gilman (2000) argues that hospitals which effectively practice cream-skimming are expected to, on average, have more severely ill patients subsequently to the introduction of a prospective payment system if supplied treatments are relevant for reimbursement calculation. This is due to the fact that these formerly unprofitable patients become profitable under such rearranged price effects. Hence, competition for profitable patients depends on the magnitude of price changes. Duggan (2000) furthermore finds that type of ownership does not explain differences in hospital behavior or efficiency. He finds that not-for-profit hospitals are not more altruistic than for-profit hospitals in the USA. Instead, differences in hospital behavior are rather associated with the hospital's payment system. In particular, he finds that public hospitals fall behind private hospitals because of what he calls the "soft budget constraint". That is, government-owned hospitals are often either not or less affected by financial incentives created by policy reforms at state or country level as e.g. changes of reimbursement may be eliminated by opposed policy actions of their owners at county level. For the case of Italy, Barbetta et al. (2007) present evidence in support of Duggan's (2000) findings. Obviously due to the soft budget constraint, they find that hospital efficiency differs by ownership type. In contrast, Herwartz and Strumann (2012) find no such effect for Germany. Again, there might be several reasons for this contradiction. One interesting explanation is provided by Lindrooth et al. (2007) who present evidence similar to the findings of Herwartz and Strumann (2012). They argue that for-profit hospitals are not expected to respond to price changes because they may have already cut intensity of health care to a minimally acceptable level before the reimbursement reform in order to maximize profits. Provided the soft budget constraint and thus no incentives for public hospitals, both hospital types are expected to be similarly non-responsive. Under this assumption, Herwartz and Strumann's (2012) finding might indicate that German for-profit hospitals, in contrast to their Italian equivalents, have already provided only minimally acceptable quality of health care prior to the observed reform. However, there is no empirical evidence for this hypothesis.

TABLE 3 - SUMMARY OF INCLUDED ARTICLES

Author, Year, Observed Country	Setting/ Target Group	Objective/ Research Question	Data Source, Sample, Observed Period	Methods	Limitations	Main Findings
Barbetta et al., 2007, Italy	General hospitals, by ownership	Does hospital performance in response to financial incentives differ between public and private not-for-profit hospitals?	Italian National Ministry of Health, balanced panel with N=531 NHS hospitals, 1995-2000	Parametric (corrected ordinary least squares, stochastic frontier analysis) and non-parametric (data envelopment analysis)	No control for variation of case-mix, systematic practice of discharging patients early, upcoding and patient selection (case shifting)	Performance explained by payment system rather than by ownership, declines after introduction of PPS (DEA, SF $p<0.05$)
Melberg et al., 2016, Norway	General hospitals	Do changes in reimbursement under prospective payment affect hospitals' treatment volume?	Directorate of Health, Norwegian Patient Registry, All Norwegian hospitals with N=3090 DRGs, 2006-2013	Regression analyses	Effects of new technologies not controlled for, only national level data, income elasticity of treatments unobserved	10 percentage increase in DRG-weight associated with increase between 0.76 and 1.10 percent in change in hospital activity ($p<0.05$)
Clemens and Gottlieb, 2014, USA	Outpatient, physicians	Is health care supply and patient health affected by physicians' incentives?	Medicare data, 5 percent random sample, 1993-2005	Ordinary least squares regressions (difference-in-difference analyses)	Patient outcomes restricted to mortality and hospitalizations, quality of life mostly unobserved	Price changes have little effect on extensive margin, 2 percent rise in payment rates leads to 3 percent increase of intensive margin

TABLE 3 CONTINUED - SUMMARY OF INCLUDED ARTICLES

Author, Year, Observed Country	Setting/ Target Group	Objective/ Research Question	Data Source, Sample, Observed Period	Methods	Limitations	Main Findings
Meltzer et al., 2002, USA	General hospitals, by patient cost	Does competition under prospective payment reduce treatment costs differentially for high- and low-cost patients?	California Office of Statewide Health Planning and Development, with all nonfederal Californian hospitals, 1983 and 1993	Ordinary least squares and quantile regressions, difference-in-differences approach	No data prior to introduction of PPS (1983), no data on postdischarge healthcare resource use, only one measure of competition	Increased competition under PPS associated with reduced expenditures, especially for cost-intensive patients
Jian and Guo, 2009, China	Psychiatric hospitals	Do day-based payments increase length of stay in psychiatric hospital wards?	Inpatient record (FPIR) database, N=1964 cases, 2002-2006	Regression analyses (for censored data: COX regression), difference-in-differences approach	Inconsistent findings (depend on chosen control group), lack of data to control for hospital's cost containment	Day-based payments associated with increased daily spending of 12 percent ($p<0.05$), LOS not affected
Norton et al., 2002, USA	Psychiatric hospitals (not-for-profit and public)	Does prospective payment reduce inpatient length of stay?	Massachusetts Medicaid data base and American Hospital Association, N=8509 cases, 1991-1993	Ordinary least squares and quantile regressions	No control group study design, short observation period	10 percent rise in average price leads to increase of length of stay between 0.16 and 0.20 ($p<0.01$), marginal price effect on length of stay insignificant

TABLE 3 CONTINUED - SUMMARY OF INCLUDED ARTICLES

Author, Year, Observed Country	Setting/ Target Group	Objective/ Research Question	Data Source, Sample, Observed Period	Methods	Limitations	Main Findings
Herwartz and Strumann, 2014, Germany	General hospitals	Do prospective payment systems affect hospital efficiency?	Statistical offices of the federal states and GENESIS regional data base, N=1600 hospitals, 1995-2006	Non-parametric (data envelopment analysis) and parametric (stochastic frontier analysis)	No sufficient control for case shifting or postdischarge healthcare resource use	Introduction of prospective payments associated with decline in hospital efficiency
Duggan, 2000, USA	General hospitals, by ownership	Does ownership affect the responsiveness of hospitals to financial incentives?	California Office of Statewide Health Planning and Development, N=397 hospitals, 1987-1995	Ordinary least squares regressions	No robustness test for measure of quality of care	Responsiveness to financial incentives depends on the "soft budget constraint" rather than on ownership ($p<0.01$), health outcomes not affected by increased medical spending
Gilman, 2000, USA	General hospitals	How do procedure-based payment systems affect intensity of inpatient care?	New York State hospital discharge files, 1992-1995	Regressions, difference-in-differences approach	Insufficient control for severity of cases and trends	Procedure-based payment systems associated with decline of average intensity of care ($p<0.05$)

TABLE 3 CONTINUED - SUMMARY OF INCLUDED ARTICLES

Author, Year, Observed Country	Setting/ Target Group	Objective/ Research Question	Data Source, Sample, Observed Period	Methods	Limitations	Main Findings
Dafny, 2005, USA	General hospitals	How do hospitals respond to price changes?	Medicare Provider Analysis and Review files and Federal Register and Medicare Cost Reports and American Hospital Association, 1985-1991	Regressions (ordinary least squares, IVs)	Partly no control for individual hospital-year trends, hence estimated elasticities concerned may be upward-biased	Hospitals, particularly for-profit, responded to price changes mainly by upcoding; case selection or intensity of care not affected
Sharma, 2009, Australia	General hospitals	Do prospective payment effects differ for cases with low, middle, and high cost intensity?	Victorian Admitted Episodes Dataset, N=177 DRGs (balanced panel), 1998/07-2005/06	Stochastic kernel approach (non-parametric)	Analysis limited to suburban and teaching hospitals	LOS for high outliers decreased under PPS due to increased quality of care, number of inliers increased for DRGs with high waiting times
Cotterill, 2008, USA	Psychiatric hospitals	How did psychiatric inpatient claims develop over time before implementation of prospective payment?	Medicare Provider Analysis and Review files and Medicare Denominator files, 1987-2004	Descriptive statistics	No causal analysis, observed period did not include introduction of PPS (no treatment)	Number of admissions developed differently for age groups due to different growth of population groups

TABLE 3 CONTINUED - SUMMARY OF INCLUDED ARTICLES

Author, Year, Observed Country	Setting/ Target Group	Objective/ Research Question	Data Source, Sample, Observed Period	Methods	Limitations	Main Findings
Kaestner and Guardado, 2008, USA	General hospitals	Is hospital staffing and patient health affected by changes in reimbursement level and introduction of prospective payment?	Prospective Payment System Payment Impact files and American Hospital Association Annual Surveys and Nationwide Inpatient Survey, 1994-2001	Regression models (e.g., Poisson)	Only one measure for resource use which might also be biased, small sample sizes, patient age limited to between 40 and 80	Hospital staffing and patient outcomes not affected by introduction of prospective payment and 10 percent change in reimbursement level
Herwartz and Strumann, 2012, Germany	General hospitals	Do prospective payment systems affect competition of local hospitals?	Statistical offices of the federal states and GENESIS regional data base, N=1600 hospitals, 1995-2006	Non-parametric (data envelopment analysis), parametric (spatial regression and stochastic frontier analysis)	SARAR resuts might be biased due to two-step estimation approach	Introduction of prospective payments associated with increase of competition, might also imply patient selection
Schmitz, 2013, Germany	Outpatient, physicians	Do financial incentives affect the outpatient intensive and extensive margins?	German Socio-Economic Panel, N=29,132 individuals, 1995-2002	Two stage (hurdle) model, logit and negative binomial model, difference-in differences approach	No appropriate control group, difference-in-difference measures likely to be biased	Changes in physicians' incentives affected patient mix and intensive, but not extensive margin of doctor visits

TABLE 3 CONTINUED - SUMMARY OF INCLUDED ARTICLES

Author, Year, Observed Country	Setting/ Target Group	Objective/ Research Question	Data Source, Sample, Observed Period	Methods	Limitations	Main Findings
Cavalieri et al., 2013, Italy	General hospitals, AMI, CHF and COPD cases	Does introduction of prospective payment increase quality of care in hospitals?	Italian National Program for Outcome Assessment, 2009-2010	Truncated regression models	Short observation period, no pre-test design, no control for supplementary funding schemes alongside PPS	Use of prospective reimbursement associated with decreased 30-day mortality rate
Finkelstein, 2007, USA	General hospitals	Does introduction of Medicare (prospective payment) increase public medical spending?	Annual surveys of the American Hospital Association, N=6,500 hospitals, 1948-1953 and 1955-1975	Ordinary least squares (weithted, unweighted) and probit regressions, generalized linear model, difference-in-differences approach	Data on cardiac technologies not available before Medicare	Introduction of Medicare associated with 37 percent increase of medical spending ($p<0.05$), increase of admissions by 32 percent ($p<0.01$) over first five years
Shen, 2003, USA	General hospitals, AMI cases	Do prospective payment systems affect inpatient quality of care?	PPS Payment Impact File and Medicare hospital cost reports and American Hospital Association Annual Surveys, N=1800 hospitals, 1985-90, 1990-94	Regressions (long-distance model, two-stage least squares)	Sample limited to urban hospitals, data on changes in reimbursement not available for all years, only one measure for quality of care	1percent decrease in PPS associated with 0.66 (3.21) percent higher 30-day mortality rate in 1985-90, $p<0.1$ (1990-94, $p<0.05$), no effect on longer term outcomes

TABLE 3 CONTINUED - SUMMARY OF INCLUDED ARTICLES

Author, Year, Observed Country	Setting/ Target Group	Objective/ Research Question	Data Source, Sample, Observed Period	Methods	Limitations	Main Findings
Besstremyannaya, 2013, Japan	General hospitals, public ownership	Does introduction of prospective payment increase quality of care in public hospitals?	Ministry of Internal Affairs & Communications and Ministry of Health, Labor and Welfare, N=347 hospitals, 2003-2009	Non-parametric (data envelopment analysis) and parametric (stochastic frontier analysis), DiD approach	No control for case-mix, small number of observed post-treatment periods	Introduction of prospective payment associated with minor (1 to 3 percent) increase in technical and cost efficiency
Cutler, 1995, USA	General hospitals	Does introduction of prospective payment affect health outcomes and coding practice in hospitals?	Medicare and Social Security records, 1981-1988	Logistic regressions, proportional hazard models	Marginal reimbursement effect unlikely to be equal for all DRGs, causing biased estimates	Decreased PPS' average (marginal) price effect assoc. with higher (lower) short-term mortality, increased readmissions due to coding practice
Haas et al., 2013, Germany	Psychiatric hospitals (psychosomatic illnesses)	Do changes in treatment cost affect quality of inpatient mental health care?	Charité Universitaetsmedizin Berlin patient data, N=101 patients, 2006/1-2010/6	Ordinary least squares regression, generalized linear model	Small sample size (especially for sub-samples), data only from one hospital (generalizability questionable)	100 Euro increase in price associated with 0.1 (0.4) increased MCS-8 score for patients with(out) somatic illness ($p<0.05$)

TABLE 3 CONTINUED - SUMMARY OF INCLUDED ARTICLES

Author, Year, Observed Country	Setting/ Target Group	Objective/ Research Question	Data Source, Sample, Observed Period	Methods	Limitations	Main Findings
Pletscher, 2016, Switzerland	Psychiatric hospitals	Does introduction of mixed prospective payment affect length of stay in inpatient mental health care?	PSYREC data base, N=13 hospitals, 2/2008-12/2011	Hazard model, difference-in-differences approach	Only one hospital in intervention group, small sample size, short observation period (only one pre-intervention year)	Introduction of mixed prospective reimbursement may affect length of stay, if changes in (marginal) price effects are large enough
Lindrooth et al., 2007, USA	General hospitals	Do changes in price under prospective payment affect intensity of care?	Healthcare Cost and Utilization Project State Inpatient data base, 1996 and 2000	Quantile regressions, difference-in-differences approach	Only two observation periods, no control for upcoding	Decreased prices under prospective payment associated with decreased intensity of care only for profitable patients
Ellis and McGuire, 1996, USA	Psychiatric hospitals	How does introduction of prospective payments affect the intensive and extensive margin of inpatient mental health care?	Medicaid claims data and New Hampshire Hospital records and state discharge files, 1987-1992	Ordinary least squares regressions, difference-in-differences approach	In general, low R^2 (coefficient of determination) values	Introduction of PPS associated with 1.8 [3.0] day decrease in length of stay due to moral hazard (intensity of care) [practice-style (patient selection)]

Altogether, longer panel datasets would be needed for several studies to observe also lagged effects and enable more precise causal interpretation of empirical results (see e.g. Barbetta et al. (2007), Clemens and Gottlieb (2014)). Moreover, reasons for contradictory findings are rather speculative and mostly not addressed in the single articles. In general, these might be differences in country characteristics and reimbursement reforms. Among others, Finkelstein (2007) gives examples for the limited generalizability of results due to differences in the studied policy reforms. Given limited means of controlling for latent situational factors which are likely to affect outcomes, i.e. the dependent variables in an econometric model (see e.g. Dafny (2005) and Gilman (2000)), results are likely to be differentially biased. In turn, even if the true effect of two reimbursement reforms might be similar, the observed hospital behavior might differ between countries. However, it is unclear to what extent this possible bias has affected the presented results. Although the quality of all articles presented in this review can be expected to be high in general as they are all published in renowned peer-reviewed journals, some of the articles obviously stand out (e.g. Clemens and Gottlieb (2014), Cutler (1995), Dafny (2005), Duggan (2000), Ellis and McGuire (1996), Finkelstein (2007)). While assessing the results presented in Table 3 and discussed in section 3.5, this quality aspect should be kept in mind when confronted with contradictory findings.

4 – Data description

Hospitals in Germany are legally obliged to produce quality reports since 2005 according to §136b Abs. 1 Satz 1 Nr. 3 SGB V (Sozialgesetzbuch Fünftes Buch). These reports are administered by the federal joint committee (Gemeinsamer Bundesausschuss, G-BA) and are available for not-for-profit, e.g. scientific, purposes upon request. Initially, they were published every other year. Since 2012, the reports are published annually. In September 2016, hospital quality reports according to §137 Abs. 3 Satz 1 Nr. 4 SGB V were available for the years 2006, 2008, 2010, and 2012 through 2014 as raw datasets in the xml format from the G-BA. The empirical analysis of this thesis is based on these data[5]. Since the quality reports are secondary hospital data, the analysis' external validity is considered to be high.

The quality reports are structured as follows. First, general information on the hospital is given. For example, the address, ownership status as well as contact details and names of the board of directors are listed, among others. Second, structural data is presented, providing details on the number of cases and beds at hospital level, the range of supplied medical services and the number and qualification of medical staff, for instance. Also, information on cases, diagnoses and treatments, encoded by ICD-10-GM and OPS codes are given at ward level, respectively. Third, detailed analyses of various quality indicators are presented. This section is evidently the main focus of the reports and assesses hospital quality in detail. However, for the purpose of this thesis, mainly information provided in the first two sections are relevant. Hence, data processing marked a key step prior to the actual empirical analyses. The process of retrieving relevant data is briefly described in the following.

Initially, the raw data files have been imported using Microsoft SQL Server 2016. Algorithms for scanning and filtering these files have been programmed. As the data structure has been slightly changed over the years, the SQL code had to be partly adapted for each year. All relevant data points for the years 2006, 2008, 2010, and 2012 through 2014 have been grouped into a single table in SQL. Furthermore, the cumulative number of supplied OPS codes, i.e. treatments, at ward level has been calculated in SQL just like the average number of OPS codes per case. In the dataset created, the aggregate number of OPS codes includes OPS which have been reported as mandatory as well as voluntary. The latter are basically only present in psychiatric and psychosomatic wards.

[5] Only parts of these quality reports are used and shown in this thesis. A complete and unmodified version of the reports is available online at www.g-ba.de

Subsequently, the data was imported in Stata 14.1 for further data processing and analysis. Prior to creating dummy variables, missing values were addressed as described below.
Eventually, data were visualized using tables as well as different graphical plots in Stata.
Over the years, the structure of the reports has faced minor modifications. In fact, reporting quality has improved thereby over time so that less missing or ambiguous values are present in recent reports compared to older ones.
After isolating the relevant data from the raw dataset, plausibility checks have been performed. On the whole, 104 observations have been identified which contained values considered as errors. In specific, the site number by which each hospital can be uniquely identified (Institutionskennzeichen) was misspecified in these cases. The characteristics of the set of variables observed for these 104 hospital wards lead to the conclusion that certain variables must have been specified incorrectly, according to the reporting guidelines issued by the G-BA (Gemeinsamer Bundesausschuss, 2016). 22 of these observations could not be specified correctly by hand and had to be deleted from the dataset in consequence. 41 observations were considered as duplicates due to reporting error and were hence deleted from the dataset. Eventually, 41 observations could be manipulated manually in alignment with the prescribed approach demanded by the G-BA and could remain in the dataset. Additionally, the dataset contained 1,345 observations for which the unique identifier at ward level (Fachabteilungsschlüssel) was missing. There may be two reasonable explanations for this occurrence. First, there are certain wards for which the G-BA's reporting guideline has simply not assigned a specific identifier. In this case, the identification code for a random or, respectively, other ward is assigned. This identification code does not convey more information than a missing value in principle. Due to this, the algorithm in SQL Server has been programmed in a manner that does not include the identifier for these random wards in the first place. Second, these hospital wards actually miss the identifier variable, while the remaining variables are specified correctly. This case is rather unlikely, yet possible. By assigning these wards the identifier for a random ward, information on the specific characteristics of the ward is lost in the second case while the observations concerned can remain in the dataset. In the first case, no information is lost by assigning the identifier for a random ward as their original specification is in fact retained. Therefore, the best way to deal with these missing values is to assign the identification code for a random ward to the observations concerned.
Furthermore, the ownership status of 41 observations could not be determined clearly so that these ambiguous observations had to be removed from the dataset, too. In alignment with

literature (see e.g. Duggan (2000)), this thesis differentiates between public, private not-for-profit and private for-profit ownership.

Summary statistics of the sample as described in the following are presented in Table 4. (Summary statistics of the dataset containing 69,042 observation (hereafter: original dataset) plus selected dummy variables are presented in Appendix A, Table 7.) Apart from the listed variables, the dataset contains a non-numeric variable for the 16 federal states in Germany.

TABLE 4 – SUMMARY STATISTICS

	N	Mean	SD	Min.	Max.
Year	63252	2010.6	2.790749	2006	2014
=1 if year>2009	63252	0.686	.4639352	0	1
Hospital level					
Cases, inpatient	63252	19155.7	16807.71	0	135981
Cases, semi-stationary	43420	704.1	1736.059	0	30689
Cases, outpatient	33339	52062.1	128610.5	0	1443326
Number of beds	63252	489.4	424.3726	0	3213
=1 if teaching hospital	63252	0.610	.4877742	0	1
=1 if ownership not-for-profit	63252	0.382	.4858832	0	1
=1 if ownership private	63252	0.182	.3861618	0	1
Ward level					
Cases, inpatient	63252	1678.7	1423.654	0	22187
Cases, semi-stationary	43420	45.19	262.5952	0	14794
Physician	52433	14.44	24.50702	0	429
Med. specialist	52433	7.582	14.03357	0	249
Sum of OPS	63252	3822.4	6722.975	0	1221220
OPS per case	63252	3.287	14.45221	0	1766
=1 if psychiatric/ psychosom. ward	63252	0.051	.2195196	0	1

Sources: Hospital quality reports 2006-2014, G-BA

The dummy variables for the location of the hospitals at state level are not listed due to purposes of illustration and readability. Altogether, 18 variables plus identification variables at the different levels of observation have been created and imported into Stata.

Salient characteristics are briefly presented and discussed. First, the maximum number of outpatient cases at hospital level is noticeable as it appears to be extremely high, also in comparison with the number of inpatient and semi-stationary cases at hospital level. The fact, however, that this variable contains a lot of missing values seems to be plausible simply because there are wards which do not offer outpatient treatment services at all. Second, the maximum number of beds per hospital appeared to be too high for a specific hospital site. Examining the data unveils that this value is most likely a reporting error as the same hospital site reported to have a capacity of exactly 158 beds for all other years available in the dataset. Therefore, this observation is corrected to 158 beds manually. Also, aggregate treatments supplied at ward level, which is measured as the sum of OPS codes in Table 4, appears to be extremely high. Again, this is due to outliers. They can be attributed to wards such as radiology where patients from other wards are treated, but not admitted. Excluding all observations from non-medical wards, i.e. with the identifier for random ward, from the original dataset solves the problem of obviously extremely high numbers of inpatient cases at ward level. This procedure appears to be reasonable as the behavior of these specific wards is not similar to the treatment group and thus does not specifically qualify for the control group. Moreover, all observations which report zero or only one case admitted at ward level are excluded from the sample. Thereby, only wards are included in which the treatment of patients is provided primarily. Wards offering mainly diagnostic services are expected to be excluded, furthermore aligning the control group to the characteristics of the treatment group. This sample as presented in Table 4 includes 63,252 observations in total. Thus, it is very likely to obtain statistically significant results from statistical tests. Subsequently to excluding the described outliers from the original dataset, the maximum number of inpatient cases at ward level has decreased to 22,187 in the sample and the maximum number of OPS per case has decreased to 1,766.

Table 5 presents the percentage share of somatic versus psychiatric and psychosomatic wards per year of the sample. (Similarly, the percentage share of somatic versus psychiatric and psychosomatic wards per year for the original dataset is presented in Appendix A, Table 8.)

As suspected, the vast majority of hospital wards in the sample are somatic wards. Interestingly, the absolute number of psychiatric and psychosomatic wards in 2010 is more than two times as high as in 2008. Simultaneously, the number of somatic wards decreased slightly, resulting in a shift of percentage shares. From then on, the number of psychiatric and psychosomatic wards

in relation to somatic wards continued to increase until 2014, i.e. the last year observed. Both the absolute number and share of psychiatric and psychosomatic wards are more than four times higher in 2014 than in 2006. These developments are similarly found for the original dataset, presented in Appendix A, Table 8.

TABLE 5 – NUMBER OF WARDS BY YEAR AND TYPE

Year	Somatic (Percent)	Psych. (Percent)	Total (Percent)
2006	9553 (98.26)	169 (1.74)	9722 (100)
2008	9929 (98.21)	181 (1.79)	10110 (100)
2010	9544 (94.67)	537 (5.33)	10081 (100)
2012	10291 (93.44)	723 (6.56)	11014 (100)
2013	10337 (92.83)	798 (7.17)	11135 (100)
2014	10387 (92.82)	803 (7.18)	11190 (100)
Total	60041 (94.92)	3211 (5.08)	63252 (100)

Sources: Hospital quality reports 2006-2014, G-BA

5 – Empirical approach

Given that the data includes the two years for which structural breaks might be expected, i.e. the announcement of PEPP in 2009 and the start of its trial phase in 2013, a classic difference-in-differences approach based on OLS regression as described by Bertrand et al. (2004) is used. One important issue pointed out by Bertrand et al. (2004) is inconsistency of standard errors as a result of misspecification of the econometric model. They show that consistent standard errors of difference-in-differences estimates are produced when the data is grouped into pre- and post-treatment periods. Hence, I simply construct a dummy variable for time, grouping the available observations into years before and after the voluntary introduction of PEPP. As the available data does not allow to controll for a vast number of variables, difference-in-differences estimation appears to be the most promising and suitable approach.

The following dependent variables have been identified. The average number of treatments per case calculated from the data is used as a proxy for the intensive margin. Analogously, the number of cases in each ward serves as a proxy for the extensive margin. A more detailed analysis using a bigger set of different dependent variables was not feasible due to data restrictions and the scope of the thesis. Literature suggests to include for example length of stay or also costs of care as additional dependent variables (see Clemens and Gottlieb (2014), Finkelstein (2007)). Particularly for studying mental health care, including measures of quality of life such as the MCS-8 score might be useful (Haas et al., 2013). Yet, the available dependent variables suffice to broadly analyze the effect of PEPP on the extensive margin as well as outline its effect on the intensive margin as a first step. Nonetheless, a more sophisticated empirical approach based on the insights obtained through this thesis is recommended.
Since the approach of this thesis is rather inductive than deductive, the empirical analysis uses mainly descriptive statistical and graphical methods to identify patterns and effects associated with PEPP. Eventually, counterfactual statements derived from the findings shall be formulated. Furthermore, the raw dataset contains a lot of variables, most of which are not relevant in order to answer the posed research question. Hence, resulting counterfactual statements will need further assessment and testing in future research. Of course, year is included as an independent (time) variable in the model to control for effects over time. Also, each ward had to be assigned a unique identifier, for which a hierarchical combination of variables on hospital, site, and ward level has been utilized. The main independent variable

utilized is a dummy variable for the status of the hospital wards, grouping the different wards either into psychiatric and psychosomatic or into somatic, i.e. not psychiatric or psychosomatic. Thereby, a treatment group, i.e. psychiatric and psychosomatic wards, and a control group, i.e. all other wards, is created with the introduction of PEPP marking the treatment. Because of the scope of the thesis, the only dependent variable used in the regressions is the number of OPS codes per case. Moreover, only the effect of the announcement of PEPP in 2009 is analyzed using the regression methods. OLS assumptions and regression diagnostics are briefly addressed in Appendix C.

In the initial model specification, no further independent variables are used which is feasible due to the explained advantages of the difference-in-differences approach. However, it is then all the more important to make sure that both groups have similar characteristics or, as done in a second step, control for the differences of the treatment and control group. Otherwise, the estimated difference-in-differences effect might be overstated. As suggested by Bertrand et al. (2004), only the aggregated time dummy variable is used in model (1) and (2) (see row 2, Table 4). Based on the above description, the econometric model has been specified as follows.

$$\ln(OPS_{ijt}) = \alpha_{it}\,(TREAT_t \times PSYCH_i) + \beta_t\,TREAT_t + \gamma_i\,PSYCH_i + \epsilon_{ijt} \qquad (1)$$

In model (1), $\ln(OPS_{ijt})$ denotes the log number of average OPS codes per case in ward *i* of hospital *j* in year *t*. The log of the dependent variable is used because it does not seem reasonable that PEPP changes the intensity of care by a constant amount but rather by a (nearly) constant percentage. α_{it} is the difference-in-differences estimate and coefficient of the variable of main interest, i.e. the interaction term $TREAT_t \times PSYCH_i$, which consists of the binary dummy variables $TREAT_t$ and $PSYCH_i$ presented in the second and in the last row of Table 4, respectively. Furthermore, β_t is the coefficient of $TREAT_t$ and γ_i is the coefficient of $PSYCH_i$. ϵ_{ijt} is simply an error term of ward *i* of hospital *j* in year *t*.

Additionally, a different model specification with additional independent variables as controls is utilized in order to test the results' robustness and enhance precision. The extended model uses the following specification.

$$\ln(OPS_{ijt}) = \alpha_{it}\,(TREAT_t \times PSYCH_i) + \beta_t\,TREAT_t + \gamma_i\,PSYCH_i + \theta_{jt}\,OWN_{jt} + \vartheta_{jt}\,TEACH_{jt} + \iota_{it}\,PHYS_{it} + \epsilon_{ijt} \qquad (2)$$

Model (2) is principally specified analogously to model (1). In addition to the variables and coefficients described above, control variables at different levels have been added to increase precision. For instance, dummy variables for the hospital's type of ownership (OWN_{jt}), teaching hospital status ($TEACH_{jt}$), and number of physicians ($PHYS_{it}$) at ward *i* in year *t* are used. Note that the number of physicians working in ward *i* may differ over time *t*. Similarly, hospitals can also change their ownership and teaching hospital status over time.
Eventually, a third model has been specified as follows.

$$\ln(OPS_{ijt}) = \alpha_{it}\,(TREAT_t \times PSYCH_i) + \beta_t\,TREAT_t + \gamma_i\,PSYCH_i + \eta_s\,STATE_s + \theta_{jt}\,OWN_{jt} + \vartheta_{jt}\,TEACH_{jt} + \iota_{it}\,PHYS_{it} + \epsilon_{ijt} \quad (3)$$

Model (3) is basically identical to model (2) despite additionally controlling for the location of the hospital. This is done by introducing dummy variables for the federal states ($STATE_s$).

6 – Empirical results

In Figure 3, the development of different variables at ward level is plotted over the time frame covered by the dataset, i.e. 2006 to 2014. Each row comprises two plots which both illustrate the same variable; first the absolute changes in the plot on the left side, and second that variable's changes relative to 2008 on the right side. In each plot, the vertical line denotes the announcement of PEPP in 2009, i.e. the treatment. The dashed lines describe the variables' development for all somatic wards, i.e. the control group. The solid lines describe the same development for all psychiatric and psychosomatic wards, i.e. the treatment group. In the first row, the plotted variable is the mean number of OPS per case, followed by the mean number of admitted cases in the second row, the mean number of OPS codes in the third row and finally the number of wards in the fourth row.

In all plots, a common trend between the treatment and the control group can be recognized clearly before 2009. Basically, there are hardly any changes in all observed variables from 2006 to 2008. Most prominently, the absolute measures of all variables for the treatment group lie below those of the control group before the announcement of PEPP. This is rather trivial for the number of wards and the number of cases simply because of the sample's composition. Yet, it is interesting concerning the number of supplied treatments and the treatment intensity, i.e. the number of OPS per case. Subsequently to the announcement of PEPP, the absolute number of somatic wards is still significantly higher than the number of psychiatric and psychosomatic wards. A similar development is found for the absolute number of cases. However, the development of the absolute measures of the OPS per case and the absolute number of OPS already indicate a structural break between 2008 and 2010. While the total number of OPS in the control group is approximately constant over all observation periods as can be seen in the changes of OPS relative to 2008, the absolute number of OPS in the treatment group increases by an order of almost 10 from 2008 to 2010 and continues to increase until 2012. Consequently, a similar development is observed for the number of OPS per case. For the control group, there are hardly any changes of supplied treatments per case with a minor exception of the year 2012. However, comparing 2014 with 2008, the intensity of treatment does not seem to have changed significantly in somatic wards.

In contrast to this, OPS per case have increased by an order of more than 10 in the treatment group from 2008 to 2014. Furthermore, the number of cases increased more in the treatment group than in the control group subsequently to the treatment. A similar development is

observed for the number of wards, which increased more in the treatment group than in the control group after the treatment.

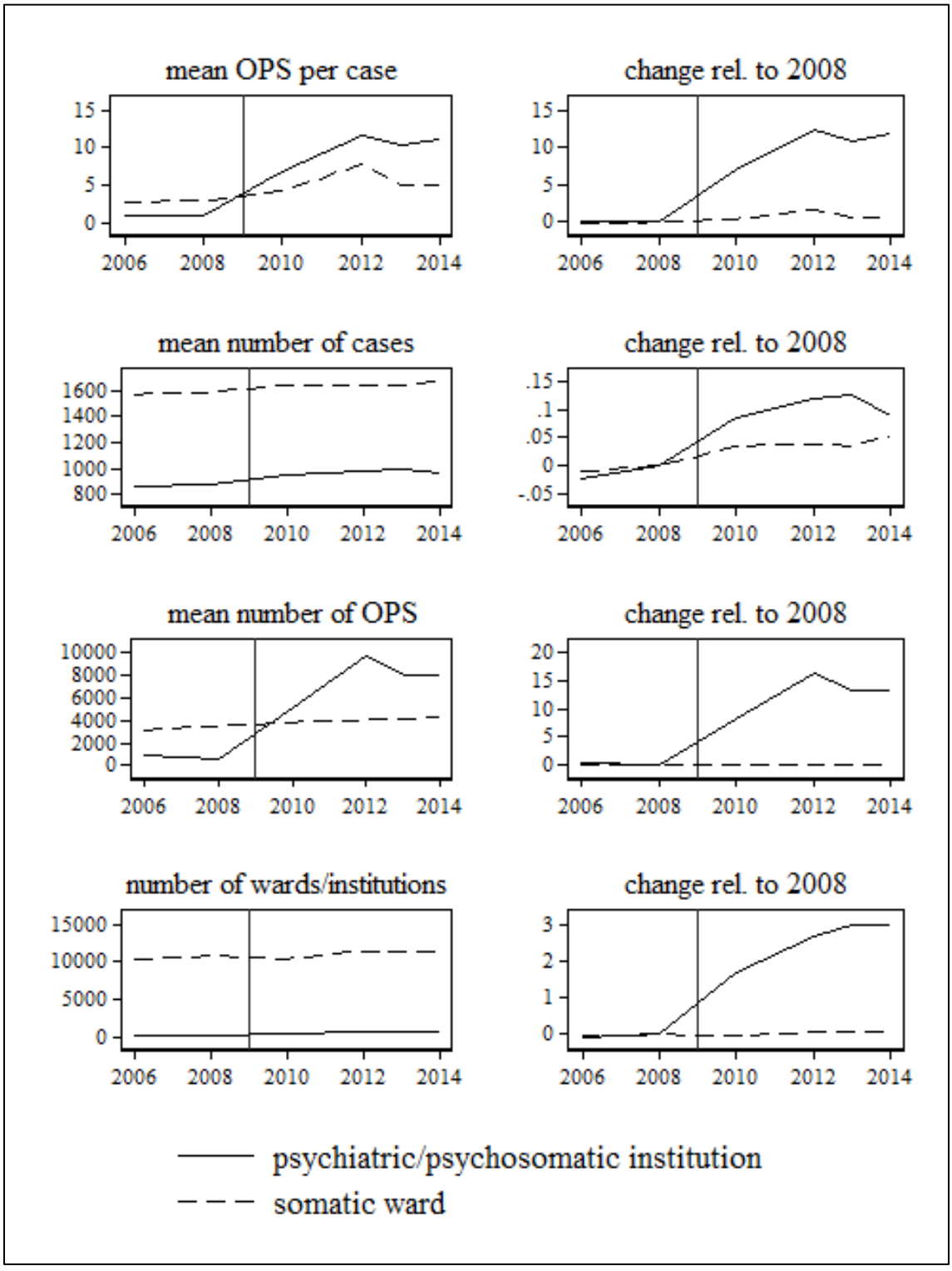

FIGURE 3 – RELATIVE AND ABSOLUTE DEVELOPMENT OF CASES, OPS, WARDS

Sources: Hospital quality reports 2006-2014, G-BA (own illustration)
Notes: The graphical plots are based on the original dataset (n=69,042).

Following the graphical description and analysis of the assumed effects of PEPP on the hospital market, now the regression results will be presented.

Table 6 shows the estimates of the regressions for model 1 in the first column, model 2 in the second column, and model 3 in the third column. At the bottom of the table, the adjusted R-squared, the number of observations and the Bayes information criterion (BIC) are presented for each model. In principal, model 3 has the best fit, which is both indicated by the lowest BIC and the highest adjusted R-squared. Thus, model 3 is able to explain most of the variance in the data in comparison with model 1 and model 2. However, in all three models, the independent variables do not predict more than 15 percent (model 1: 13.3 percent, model 2: 14.6 percent, model 3: 15.0 percent) of the dependent variable's variation. Therefore, relatively much variation is left unexplained, unfortunately. Yet, the finding that model 3 still yields the best results is also consistent with the findings from the sensitivity analysis as described below. In this thesis, I prefer the BIC over the Akaike information criterion (AIC) because the AIC is known to overestimate the statistical significance in large samples.

Due to the fact that the sample size is very large with over 60,000 observations in total, estimates of the independent variables tend to easily become statistically significant. To investigate whether statistical significance is overestimated in this sample, the same models are regressed on a smaller sample as part of sensitivity analysis (see Appendix B, Table 12).

In model 1 as well as model 2, all independent variables are highly statistically significant. Only in model 3, there are results which are not statistically significant at the 0.1 percent level. Namely, there is no statistically significant difference between public hospitals (baseline specification) and private for-profit hospitals. Furthermore, there are no large or statistically significant differences in OPS per case between most federal states. Although not reported in detail in Table 6 due to visibility reasons, the main differences will be described in the following. With the state of Thüringen as the baseline specification, OPS per case were approximately 25 percent lower in Schleswig-Holstein (25.5 percent), Lower Saxony (21.7 percent) and Mecklenburg-Western Pomerania (27.3 percent). Differences for Hamburg, Bremen, Berlin, Saxony, Saarland and Baden-Wuerttemberg are statistically insignificant. The estimates' magnitude of the remaining federal states lies in the described range.

In principle, all models are consistent concerning the direction and magnitude of the estimates. In all three models, the number of OPS per case was approximately 15 percent higher for all wards after the treatment, i.e. after 2009. (Exemplarily for model 1: percentage change of OPS per case is $100 \times 0.144 = 14.4$ percent.) Before 2009, the OPS per case are approximately 112 percent (model 1: 111.6 percent, model 2: 115.9 percent, model 3: 111.5 percent) lower in the

treatment group than in the control group. In model 2 and 3, there are slight differences in the OPS per case between the different ownership types at baseline specification. In private not-for-profit hospitals, the number of OPS per case is approximately 4 percent (model 2: 3.6 percent, model 3: 4.0 percent) lower than in public hospitals. In teaching hospitals, the number of OPS per case was about 19 percent (model 2: 19.6 percent, model 3: 19.0 percent) higher than in non-teaching hospitals. Statistically significant differences are also found for the number of physicians. In particular, model 2 and model 3 both find that the number of OPS per case increases by 5 percent for every 10 additional physicians at ward level in both groups.

TABLE 6 – EFFECT OF PEPP ON LOG OPS PER CASE

	Dependent variable: ln (OPS per case)		
	Model (1)	Model (2)	Model (3)
TREAT	0.144***	0.153***	0.162***
	(0.007)	(0.010)	(0.010)
PSYCH	-1.116***	-1.159***	-1.115***
	(0.080)	(0.047)	(0.083)
TREAT × PSYCH	2.552***	2.624***	2.578***
	(0.082)	(0.050)	(0.085)
OWN = not-for-profit		-0.036***	-0.040***
		(0.007)	(0.008)
OWN = for-profit		0.002	0.016
		(0.009)	(0.010)
TEACH		0.196***	0.190***
		(0.007)	(0.007)
PHYS		0.005***	0.005***
		(0.001)	(0.001)
Constant	0.466***	0.253***	0.344***
	(0.006)	(0.027)	(0.036)
Full set of STATE dummies	No	No	Yes
Adj. R-squared	0.133	0.146	0.150
N	62712	62712	61896
BIC	155720.0	154861.8	152435.8

Sources: Hospital quality reports 2006-2014, G-BA
Notes: All three models in the table are estimated using OLS regression. Standard errors (model 1 and model 3: robust standard errors) in parentheses.
*: Significant at the 5 percent level.
**: Significant at the 1 percent level.
***: Significant at the 0.1 percent level

Yet, the main variable of interest is the interaction term TREAT × PSYCH. After the announcement of PEPP in 2009, the number of OPS per case increased by approximately 260 percent (model 1: 255.2 percent, model 2: 262.4 percent, model 3: 257.8 percent) more in

psychiatric and psychosomatic wards than in somatic wards. This is a finding of both great magnitude as well as statistical significance.

Subsequently to testing the empirical models on the proposed sample, sensitivity analyses have been performed. Particularly, different subsamples have been created and tested using model 1, model 2 and model 3. First, the three models have been tested using the original dataset (n=69,042). As presented in Appendix A, Table 9, similar regression results regarding statistical significance, direction and magnitude of the effects indicate robustness of the empirical models. Second, the sample has been restricted to smaller subsamples. I identified specific wards from the treatment as well as from the control group which I expect to be more similar than the whole treatment and control group. In particular, these are child and adolescent psychiatry as the treatment group and pediatric (somatic) wards as the control group. For instance, I expect the unobserved patient level characteristics to be more similar in this subsample. Therefore, the regression results might be more precise. In contrast to this, as this subsample is substantially smaller than the initial sample, the results might also be less statistically significant. The results of these regressions are presented in Appendix B, Table 12. Again, all three models' estimates are proven to be robust concerning direction, magnitude and statistical significance in general.

In addition, a fixed effects model has been estimated. Because it must be assumed that the proposed models suffer from endogeneity, estimates of the OLS regressions might be inconsistent and biased as further outlined in the limitations addressed in section 8. In general, the fixed effects' estimates are in line with the OLS estimates. Before regressing the model, all (nearly) time invariant control variables such as the location (STATE) and ownership status (OWN) were deleted from the model to avoid misspecification. This is necessary and also reasonable, because the effect of these variables will be expressed by the fixed effects estimates of the alternative model. As described in section 5, it is technically possible that the two variables mentioned, i.e. STATE and OWN, change over time, in fact. However, because changes in these variables are rare in the dataset, it is highly likely that not excluding these variables will lead to collinearity in the fixed effects model. Thus, the main effect left in the fixed effects model was the difference-in-differences estimator, which is similar to the estimates from the OLS regression results (fixed effects estimates not reported). Again, the findings from this sensitivity analysis are generally consistent with the OLS models concerning direction, statistical significance as well as magnitude of the main effect, i.e. the difference-in-differences estimate. Yet, the estimate's magnitude in the fixed effects model is approximately 10 percent lower (approximately 2.30) than in the OLS models (see Table 6).

7 – Discussion

Based on the graphical plots in Table 6, I suspect that PEPP has had a significant effect on the hospital market. In particular, I expect that PEPP is associated with increased intensity of care, indicated by OPS per case, and growth of the mental health care sector in general, indicated by the number of wards. Moreover, PEPP might have an effect on patient selection. The differential development of the number of patients shown in Table 6 may indicate the presence of such an adverse outcome due to PEPP. However, the development of this variable could also be due to true changes in the population's morbidity instead of the altered admission policies of hospitals. Due to data restrictions, the causal relationship of the number of cases and PEPP cannot be studied in this thesis, unfortunately. In general, causal interpretation of the identified relationships is very limited if possible at all in this thesis as described in the following.

Furthermore, the estimates of the difference-in-differences variable are much larger in magnitude than expected. On the one hand, the observed changes in the intensity of mental health care might truly be a result of and attributable to PEPP. On the other hand, there are also several alternative explanations which cannot be ruled out on the basis of the available dataset. One possible alternative explanation is that there have been administrative changes or altered reporting requirements in the hospital quality reports which, somehow, affected psychiatric and psychosomatic wards differently than somatic wards. In this case, the observed increase in intensity of mental health care is most likely at least partly a result of changed reporting practice and not a response to PEPP. Alternatively, hospitals might have simply changed their coding practice rather than their treatment policy in response to PEPP. While this would still be an effect of PEPP, the measure of the effect of PEPP on the true intensity of care would be distorted in this thesis. For instance, if coding of treatments is regarded as rather time-consuming bureaucracy with no direct financial benefit by hospitals, they might simply prefer not to report supplied treatments in detail and thereby save time and money. If, however, hospitals have a financial incentive to extend the reporting of treatment intensity, i.e. the marginal price effect is positive, they are expected to report at least all provided treatments which are relevant for reimbursement calculation. Similar effects are found by Gilman (2000) concerning the intensity of care. This assumption is in line with Dafny's (2005) basic assumptions of the marginal price effect on intensity of care. While I assume that PEPP creates an incentive to increase intensity of mental health care, this side effect of altered reporting policy might likely occur as well. In support of this, Serdén et al. (2003) find that hospitals increase the attention to coding procedure

subsequently to the introduction of prospective reimbursement. However, whether this is necessarily the case is mostly left to speculation. If hospitals have already reduced intensity of health care to a minimally acceptable level prior to the studied policy change as described by Lindrooth et al. (2007) for for-profit hospitals in the USA, they will not be able to expand their reporting of treatments unless they also truly expand the supply of treatments. Hence, the presence of hospital responses differential by type of ownership might indicate such a reporting policy effect of PEPP. In contrast to Barbetta et al. (2007) and Dafny (2005), who also finds that for-profit hospitals respond differently to the introduction of prospective reimbursement than public or not-for-profit hospitals, such differences are not found in this thesis. Rather, the OLS models regressed on the sample (Table 6) find that the supplied OPS per case were statistically significantly lower in not-for-profit hospitals than in the other two types of ownership. These interesting findings can be explained by the soft budget constraint of public hospitals described by Duggan (2000) and Lindrooth et al.'s (2007) hypothesis of minimally acceptable treatment intensity in for-profit hospitals as discussed in section 3.5. However, this result is inconsistent with the results of the pediatric subsample (Table 12) as well as with the results of the fixed effects approach. Furthermore, another simple explanation for the large effect of PEPP might be simply reporting or measurement error. As the data is supplied by the hospitals and processed by the G-BA, hospitals could also have manipulated the reported information. In other words, upcoding might have affected the reported data. Yet, it seems rather unlikely that such adverse outcomes affect psychiatric and psychosomatic wards differently than somatic hospital wards without any treatment such as altered financial incentives. Still, one of the biggest threats and sources of biased estimates when using difference-in-differences techniques are unobserved events which affect the treatment and the control group differentially. Apart from the described examples like altered reporting standards, such an event could also have been a reform or other changes in the G-DRG system which would have most likely affected primarily the control group. Beside the G-DRG system, changes in the reimbursement of outpatient health care, i.e. changes in physicians' financial incentives, could have had differential effects on mental and somatic health care, too. Moreover, increased awareness, reduced stigma towards mental illnesses just as changes in the prevalence of mental illnesses in the German population might have biased the difference-in-differences estimates of the presented models. Despite the robustness of the estimated effects, future research based on a broader data foundation is needed to rule out such alternative explanations. All in all, the observed changes in intensity of care might or might not be entirely attributable

to PEPP directly. Thus, the intensity of mental health care might not have changed as dramatically as the regression estimates and the graphs in Figure 3 indicate.

The empirical results are also surprising considering that by the time of the announcement of PEPP in 2009, quantifying the changes of the average and marginal price effect of PEPP was not possible yet. However, there is broad evidence that the magnitude of the changes in the marginal price effect determines how hospitals will respond to reimbursement changes and whether there will be a response at all (Finkelstein, 2007; Kaestner and Guardado, 2008; Pletscher, 2016). Future research should also incorporate the actual reimbursement changes, possibly even at the level of diagnoses or cost groups. This is also proposed by Dafny (2005) and Pletscher (2016), among others. However, this was not possible in this thesis due to time and data restrictions. The results presented in this thesis must therefore be regarded as sort of an initial and broad estimation of the effects of PEPP on the German hospital market. Following up on this, major shortcomings of this thesis are presented and discussed now.

Concerning the literature review, no grey literature was included subsequently to data base searching. Such articles might have shed light on interesting and particularly recent results, maybe even concerning PEPP specifically. Yet, the data base searching resulted in a vast number of international publications. The included articles have partly limited significance because of the chosen research design. As summarized in Table 3, some of the articles only cover short observation periods or do not have a control group design, for instance. As only published articles are included in the review, so-called publication bias may be an issue. Moreover, only articles published in English and German language were considered. Also, the selection of publications was conducted by only one researcher. In consequence, exclusion and inclusion decisions could not be double-checked.

Concerning the empirical part of the thesis, the reported number of treatments measured by the number of OPS codes is partly censored by the hospitals due to privacy reasons. This affects the treatment and the control group similarly. Therefore, the empirical models are expected to underestimate the true effect of PEPP on the intensive margin.

As the available hospital data has been analyzed on an aggregate level, patient level and individual effects such as upcoding of treatments to specific OPS codes in order to achieve higher reimbursement rates for certain patient groups as described by Dafny (2005) cannot or can only be approximately controlled for. Omitted-variable bias may occur as a result, which limits the model's internal validity. There is vast literature available explaining why including patient level variables is important (see e.g. Ellis and McGuire (1996), Cutler (1995), Norton

et al. (2002) and Cotterill (2008)). In line with literature and in particular with Ellis and McGuire (1996), I suspect that patient level variables may be correlated with both the dependent and independent variables, e.g. at hospital level. Therefore, the proposed models might be endogenous. To formally test whether they are specified correctly, Ramsey's regression specification error test (RESET) has been used additionally. The results suggest that model 2 suffers from omitted variable bias. However, including more of the available variables, e.g. controlling for the hospital size using the number of beds did not improve the test results. Interestingly, the RESET test finds no evidence for omitted variables bias in model 2 when using the sensitivity analysis' pediatric subsample ($p=0.805$). Applying the RESET test to model 1 is technically not possible. Thus, more data and in particular more variables would be needed to obtain less biased regression models.

Apart from omitted variable bias, reverse causality might be also be a source of bias in the proposed models. If the population's average morbidity and thereby patients' treatment needs increase, then the intensity of care is expected to increase, too. However, increasing intensity of care simultaneously decreases the morbidity of the population. Morbidity could be measured by a proxy such as the number of wards, physicians, or cases. Using OPS per case as the dependent variable in the regressions (see e.g. Table 6), developments of both the underlying population's morbidity and the intensity of care are simultaneously, but not necessarily separately represented in the regression models. Both components of the dependent variable may affect each other. These effects will, however, remain unobserved. Moreover, increases in OPS per case might affect physicians at ward level, which can be interpreted as a proxy measure for the size of a ward. In that case, the true effect of physicians on OPS per case will be underestimated. Apart from independent variables, using different sets of dependent variables is therefore recommended for future research.

In general, it is likely that the specified models are endogenous in several ways. There are unfortunately not many options to fix these potential issues as no appropriate instrumental variables are available in the dataset, for example. Again, including supplementary data from different sources is recommended for future research. One possible solution to overcome endogeneity problems due to omitted variable bias is the use of a fixed effects model as described in the sensitivity analysis above. This approach allows to control for unobserved effects at – in this case – ward level. In consequence, problems such as omitted variable bias are minimized. Using the fixed effects approach did not alter the difference-in-differences estimate substantially. Therefore, although results estimated by OLS might be inconsistent, I argue that they are still suitable in terms of indicating the direction as well as the approximate

magnitude of the relationship. Yet, it is important to emphasize once more that unobserved events which took place between 2008 and 2010 just as the announcement of PEPP might have only affected mental health care, for instance. This might lead to significantly biased results which also the fixed effects model cannot correct in the described specification.

Furthermore, empirically studying the hospitals' responses specifically to changes in the marginal and the average price effect was not feasible in this thesis. Yet, Ellis and McGuire (1996) and Norton et al. (2002), among others, present evidence that not measuring the marginal and the average price effect separately will likely cause bias or result in not observing a truly existent effect at all. Yet, referring to Figure 3, measures for the development and changes in both price effects are presented at least graphically. Provided OPS per case can be interpreted as a measure for the hospitals' response to changes in the marginal price effect. Similarly, the number of admitted patients can be interpreted as a measure for the response to changes in the average price effect. In line with Dafny (2005), I expected the average price effect to be positive subsequently to the announcement of PEPP. Based on the plot for the relative change of the number of cases in Figure 3, which increases after 2009, I still expect the average price effect of PEPP to be positive. This increase might also be due to adverse outcomes (Cutler, 1995). However, this effect is no further studied in this thesis.

As stated earlier, the available data covers only periods before the announcement of PEPP and periods in which PEPP is introduced as a voluntary payment mechanism. This might lead to self-selection bias. Furthermore, it is unknown which of the psychiatric and psychosomatic wards have chosen to apply the new reimbursement system and which wards have not. Isolating the effect of the payment component of PEPP on the hospital market is consequently impossible due to the data restrictions, unfortunately. However, all hospitals knew beforehand whether reimbursement and patient classification as regulated by PEPP will affect them, eventually. I argue that it is highly unlikely that these hospitals refuse to prepare for the price shock and requirements concerning patient classification created by PEPP as of 2017. To ensure a fluent introduction of PEPP, these hospitals must timely train staff and adapt their patient classification and reporting standards, for instance. Moreover, a comprehensive policy reform such as PEPP bears the opportunity for hospitals to build a knowledge-edge. Given that under prospective payment, hospitals might face increased competition as observed by Herwartz and Strumann (2012), such an advantage due to experience might be of even higher value. Consequently, the announcement of PEPP marked a perfect opportunity for hospitals to already become familiar with the new reimbursement mechanism at relatively low cost. Although the

available data does not supply evidence on this, I therefore speculate that most hospitals which will be affected by PEPP in 2017 were similarly affected by it in 2009 already.

8 – Conclusion

This thesis aimed to shed light on the development of hospital admissions as well as the intensity of care prior and subsequently to the announcement and optional introduction of PEPP. Using a combination of qualitative and quantitative methods, the development of important proxy variables for the intensive and extensive margin of mental health care was analyzed. First, a systematic literature review has been carried out. Second, econometric models were tested using mainly OLS regressions. The results of this thesis show that the intensity of care increased subsequently to the announcement of PEPP in 2009. In particular, the number of treatments per case measured by OPS codes increased by approximately 260 percent more in psychiatric and psychosomatic wards than in somatic wards subsequently to the announcement of PEPP. Detailed regression results are shown in Table 6. Additionally, sensitivity analyses using a fixed effects approach and, as shown in Table 9 and Table 12, different samples underpin the robustness of the OLS regressions' findings. Furthermore, graphical data analysis also shows changes in the extensive margin. Figure 3 illustrates for instance that hospital admissions increased more in psychosomatic and psychiatric wards than in somatic wards after the announcement of PEPP. However, this effect was no further analyzed empirically because of the scope of this thesis.

The main contribution of this thesis is, apart from the empirical analysis, the detailed overview of the effects of similar policy reforms in different countries as part of the systematic literature review. By means of the literature review, many examples were identified in support of the changes observed in the empirical component of this thesis (see e.g. Lindrooth et al. (2007), Duggan (2000), and Finkelstein (2007)). Yet, it must be noted that comparing the effects of reimbursement reforms across countries is challenging due to differences in the specific policy reforms and country characteristics (Finkelstein, 2007; Frank and McGuire, 2005, pp. 897–898). Building on the literature summarized in Table 3, it appears likely that the described developments of the intensive and the extensive margin are associated with the announcement of PEPP in 2009. In this case, these developments could be interpreted as the hospital market's response to PEPP. However, establishing a causal relationship based on this thesis is not

intended and recommended due to various limitations concerning utilized data and methods. Instead, the empirical results should rather be considered as preliminary findings of possible effects of PEPP on the hospital market which must be complemented by more comprehensive research in the future. While the research questions could not be answered definitively due to the lack of a causal relationship, analyzing the parallel developments of the policy reform and the hospital market was possible. To date, there is hardly any research on how changes in financial incentives as part of policy reforms affect the supply of mental health care or, more broadly, on drivers of patient health in psychiatric or psychosomatic wards in Germany (Haas et al., 2013). Therefore, this thesis still adds important insights to the discussion of the effects of prospective reimbursement systems on mental health care, in specific. Also, most effects of PEPP are expected to unfold in the future, i.e. subsequently to the binding introduction in 2017, and are therefore not observable yet. Thus, it will still take time until the true effects of PEPP on the German hospital market can be determined and studied empirically in a more comprehensive way. Before the mandatory introduction of PEPP, its effects may only be analyzed on a speculative rather than evidence informed basis (Ellis and McGuire, 1996; Wörz and Busse, 2005). For instance, the expected effects of PEPP on hospitals might be compensated by unanticipated effects as similarly described by Duggan (2000). Designing successful reimbursement systems is particularly challenging for mental health care and PEPP in particular (Douven et al., 2015; GKV-Spitzenverband, 2009). In this respect, this thesis may also be considered as an initial evaluation and feedback for policy makers. Given that PEPP is designed as a learning system, such feedback is presumably desired and needed by decision makers, which again emphasizes the importance of the presented findings.

Yet, in the course of this thesis, interesting new questions were raised which might inspire future research to assess the identified relationships and effects in more detail. For instance, it will be enlightening to analyze whether the binding introduction of PEPP in 2017 will mark a structural break similar to the one identified between 2008 and 2010 (see e.g. Figure 3). Also, using other data sources to confirm or disconfirm the findings of this thesis concerning the development of the hospital market before and after 2009 might be interesting. Furthermore, another important focus of future research might be to precisely determine the magnitude of the average and marginal price effect of PEPP on the hospital market.

All in all, this thesis is considered to be a good starting point for future research, especially for analyzing causal relationships and the different assumed effects of PEPP on the German hospital market in more detail.

References

Barbetta, G. P., Turati, G. and Zago, A. M. (2007) 'Behavioral Differences between Public and Private Not-for-Profit Hospitals in the Italian National Health Service', *Health Economics*, vol. 16, no. 1, pp. 75–96.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004) 'How Much Should We Trust Differences-in-Differences Estimates?', *Quarterly Journal of Economics*, vol. 119, no. 1, pp. 249–275.

Besstremyannaya, G. (2013) 'The impact of Japanese hospital financing reform on hospital efficiency: A difference-in-difference approach', *Japanese Economic Review*, vol. 64, no. 3, pp. 337–362.

Bundesministerium für Gesundheit (2012) *Verordnung pauschalierende Entgelte Psychiatrie und Psychosomatik 2013: Amtliche Begründung zum Referentenentwurf* [Online]. Available at http://www.g-drg.de/cms/Aktuelles/Amtliche_Begruendung_zur_PEPPV_2013 (Accessed 1 November 2016).

Bundesministerium für Gesundheit (2016) *Entgelte für voll- und teilstationäre Krankenhausleistungen* [Online]. Available at http://www.bmg.bund.de/themen/krankenversicherung/stationaere-versorgung/krankenhausfinanzierung.html (Accessed 31 October 2016).

Busse, R., Schreyögg, J. and Stargardt, T. (2013) *Management im Gesundheitswesen: Das Lehrbuch für Studium und Praxis*, Springer-Verlag.

Cavalieri, M., Gitto, L. and Guccio, C. (2013) 'Reimbursement systems and quality of hospital care: An empirical analysis for Italy', *Health Policy*, vol. 111, no. 3, pp. 273–289.

Clemens, J. and Gottlieb, J. D. (2014) 'Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?', *American Economic Review*, vol. 104, no. 4, pp. 1320–1349.

Cotterill, P. G. (2008) 'Medicare Psychiatric Admissions, 1987-2004: Does The Past Offer Insights For The Future?', *Health Affairs*, vol. 27, no. 4, pp. 1132–1139.

Cutler, D. M. (1995) 'The Incidence of Adverse Medical Outcomes under Prospective Payment', *Econometrica*, vol. 63, no. 1, pp. 29–50.

Dafny, L. S. (2005) 'How Do Hospitals Respond to Price Changes?', *American Economic Review*, vol. 95, no. 5, pp. 1525–1547.

Douven, R., Remmerswaal, M. and Mosca, I. (2015) 'Unintended effects of reimbursement schedules in mental health care', *Journal of Health Economics*, vol. 42, pp. 139–150.

Duggan, M. G. (2000) 'Hospital Ownership and Public Medical Spending', *Quarterly Journal of Economics*, vol. 115, no. 4, pp. 1343–1373.

Ellis, R. P. and McGuire, T. G. (1996) 'Hospital response to prospective payment: Moral hazard, selection, and practice-style effects', *Journal of Health Economics*, vol. 15, no. 3, pp. 257–277.

Finkelstein, A. (2007) 'The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare', *Quarterly Journal of Economics*, vol. 122, no. 1, pp. 1–37.

Frank, R. G. and McGuire, T. G. (2005) 'Chapter 16 Economics and mental health', in Culyer, A. J. and Newhouse, J. P. (eds) *Handbook of Health Economics*, Amsterdam, Elsevier, pp. 893–954.

Geissler, A. and Quentin, W. (2010) 'Prospective payment for inpatient psychiatric care', *Health Policy Monitor* [Online]. Available at http://www.hpm.org/de/a15/3.pdf (Accessed 30 October 2016).

Gemeinsamer Bundesausschuss (2016) *Regelungen zum Qualitätsbericht der Krankenhäuser, Qb-R* [Online]. Available at https://www.g-ba.de/downloads/62-492-1235/Qb-R_2016-07-21_iK_2016-08-06.pdf (Accessed 29 October 2016).

Gilman, B. H. (2000) 'Hospital Response to DRG Refinements: The Impact of Multiple Reimbursement Incentives on Inpatient Length of Stay', *Health Economics*, vol. 9, no. 4, pp. 277–294.

GKV-Spitzenverband (2009) *Vereinbarung über die Einführung eines pauschalierenden Entgeltsystems für psychiatrische und psychosomatische Einrichtungen gemäß §17 d KHG* [Online]. Available at https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/krankenhaeuser/psychiatrie/KH_2009_11_23_Grundlagenvereinbarung_Psych-Entgeltsystem.pdf (Accessed 1 November 2016).

Haas, L., Stargardt, T., Schreyögg, J., Schlosser, R., Klapp, B. F. and Danzer, G. (2013) 'The Trade-Off between Costs and Quality of Care in the Treatment of Psychosomatic Patients with Somatoform Pain Disorder', *Applied Health Economics and Health Policy*, vol. 11, no. 4, pp. 359–368.

Herwartz, H. and Strumann, C. (2012) 'On the effect of prospective payment on local hospital competition in Germany', *Health Care Management Science*, vol. 15, no. 1, pp. 48–62.

Herwartz, H. and Strumann, C. (2014) 'Hospital Efficiency under Prospective Reimbursement Schemes: An Empirical Assessment for the Case of Germany', *European Journal of Health Economics*, vol. 15, no. 2, pp. 175–186.

Hodgkin, D. and McGuire, T. G. (1994) 'Payment levels and hospital response to prospective payment', *Journal of Health Economics*, vol. 13, no. 1, pp. 1–29.

Institut für das Entgeltsystem im Krankenhaus (2015a) *Abschlussbericht zur Weiterentwicklung des PEPP-Systems für 2016* [Online]. Available at http://www.g-drg.de/cms/inek_site_de/layout/set/einspaltig/PEPP-Entgeltsystem_2016/Abschlussbericht_zur_Entwicklung_des_PEPP-Systems_und_PEPP-Browser/Abschlussbericht_zur_Weiterentwicklung_des_PEPP-Systems_fuer_2016 (Accessed 1 November 2016).

Institut für das Entgeltsystem im Krankenhaus (2015b) *Definitionshandbuch PEPP-Version 2016* [Online]. Available at http://www.g-drg.de/cms/inek_site_de/content/view/full/5652?campaign=Pepp16&kwd=DefHb16 (Accessed 1 November 2016).

Institut für das Entgeltsystem im Krankenhaus (2015c) *Ergänzende Fallbeispiele der Selbstverwaltung zur PEPPV 2016* [Online]. Available at http://www.g-drg.de/cms/inek_site_de/layout/set/einspaltig/PEPP-Entgeltsystem_2016/Abrechnungsbestimmungen (Accessed 1 November 2016).

Institut für das Entgeltsystem im Krankenhaus (2015d) *Verfahrensbeschreibung PEPP-Vorschlagsverfahren* [Online]. Available at http://www.g-drg.de/cms/PEPP-Vorschlagsverfahren/Verfahrensbeschreibung_PEPP-Vorschlagsverfahren (Accessed 1 November 2016).

Institut für das Entgeltsystem im Krankenhaus (2016) *Veröffentlichung der PEPP Vorschläge* [Online]. Available at http://www.g-drg.de/cms/PEPP-Vorschlagsverfahren/Veroeffentlichung_PEPP_Vorschlaege (Accessed 1 November 2016).

Jian, W. and Guo, Y. (2009) 'Does Per-Diem Reimbursement Necessarily Increase Length of Stay? The Case of a Public Psychiatric Hospital', *Health Economics*, vol. 18, S2, p. S97-106.

Kaestner, R. and Guardado, J. (2008) 'Medicare reimbursement, nurse staffing, and patient outcomes', *Journal of Health Economics*, vol. 27, no. 2, pp. 339–361.

Lindrooth, R. C., Bazzoli, G. J. and Clement, J. (2007) 'The Effect of Reimbursement on the Intensity of Hospital Services', *Southern Economic Journal*, vol. 73, no. 3, pp. 575–587.

Lumley, T., Diehr, P., Emerson, S. and Chen, L. (2002) 'The importance of the normality assumption in large public health data sets', *Annual review of public health*, vol. 23, no. 1, pp. 151–169.

Melberg, H. O., Beck Olsen, C. and Pedersen, K. (2016) 'Did hospitals respond to changes in weights of Diagnosis Related Groups in Norway between 2006 and 2013?', *Health Policy*, vol. 120, no. 9, pp. 992–1000.

Meltzer, D., Chung, J. and Basu, A. (2002) 'Does competition under Medicare Prospective Payment selectively reduce expenditures on high-cost patients?', *RAND Journal of Economics*, vol. 33, no. 3, pp. 447–468.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G. (2009) 'Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement', *PLoS medicine*, vol. 6, no. 7, pp. e1000097.

Norton, E. C., van Houtven, C. H., Lindrooth, R. C., Normand, S. T. and Dickey, B. (2002) 'Does Prospective Payment Reduce Inpatient Length of Stay?', *Health Economics*, vol. 11, no. 5, pp. 377–387.

Oxford English Dictionary *"psychiatry, n."* [Online]. Available at http://www.oed.com/view/Entry/153856?redirectedFrom=psychiatry (Accessed 28 October 2016).

Pletscher, M. (2016) 'Marginal Revenue and Length of Stay in Inpatient Psychiatry', *European Journal of Health Economics*, vol. 17, no. 7, pp. 897–910.

Schmitz, H. (2013) 'Practice budgets and the patient mix of physicians – The effect of a remuneration system reform on health care utilisation', *Journal of Health Economics*, vol. 32, no. 6, pp. 1240–1249.

Schreyögg, J. and Stargardt, T. (2010) 'The Trade-Off between costs and outcomes: The case of acute myocardial infarction', *Health services research*, vol. 45, 6p1, pp. 1585–1601.

Serdén, L., Lindqvist, R. and Rosén, M. (2003) 'Have DRG-based prospective payment systems influenced the number of secondary diagnoses in health care administrative data?', *Health Policy*, vol. 65, no. 2, pp. 101–107.

Sharma, A. (2009) 'Inter-DRG resource dynamics in a prospective payment system: a stochastic kernel approach', *Health Care Management Science*, vol. 12, no. 1, pp. 38–55.

Shen, Y.-C. (2003) 'The effect of financial pressure on the quality of care in hospitals', *Journal of Health Economics*, vol. 22, no. 2, pp. 243–269.

Shleifer, A. (1985) 'A theory of yardstick competition', *RAND Journal of Economics*, vol. 16, no. 3, pp. 319–327.

Sloan, F. A., Morrisey, M. A. and Valvona, J. (1988) 'Case Shifting and the Medicare Prospective Payment Systems', *American Journal of Public Health*, vol. 78, no. 5, pp. 553–556.

White, C. and Nguyen, N. (2011) 'How Does the Volume of Post-acute Care Respond to Changes in the Payment Rate?', *Medicare & Medicaid Research Review*, vol. 1, no. 3, pp. E1-E21.

Wörz, M. and Busse, R. (2005) 'Analysing the Impact of Health-Care System Change in the EU Member States--Germany', *Health Economics*, vol. 14, no. 51, p. S133-49.

Appendices

Appendix A

In the following, the summary statistics (Table 7), the number of wards by year and type (Table 8), and the regression results (Table 9) for the original dataset are presented.

TABLE 7 – SUMMARY STATISTICS, ORIGINAL DATASET

	N	Mean	SD	Min.	Max.
Year	69042	2010.7	2.786562	2006	2014
=1 if year>2009	69042	0.688	.4632126	0	1
Hospital level					
Cases, inpatient	69042	19199.2	16748.34	0	135981
Cases, semi-stationary	47518	711.8	1748.408	0	30689
Cases, outpatient	36585	52192.6	127493.7	0	1443326
Number of beds	69042	492.1	425.3952	0	6885
=1 if teaching hospital	69042	0.617	.4862051	0	1
=1 if ownership not-for-profit	69042	0.384	.4863955	0	1
=1 if ownership private	69042	0.180	.3843884	0	1
Ward level					
Cases, inpatient	69042	1568.7	1462.47	0	33361
Cases, semi-stationary	47518	43.36	259.9914	0	14794
Physician	55188	14.22	24.36703	0	429
Med. specialist	55188	7.493	13.89986	0	249
Sum of OPS	69042	4000.6	7057.445	0	1221220
OPS per case	69042	4.899	180.734	0	38123
=1 if psychiatric/ psychosom. ward	69042	0.047	.2126921	0	1

Sources: Hospital quality reports 2006-2014, G-BA

TABLE 8 – NUMBER OF WARDS BY YEAR AND TYPE, ORIGINAL DATASET

Year	Somatic (Percent)	Psych. (Percent)	Total (Percent)
2006	10322 (98.23)	186 (1.77)	10508 (100)
2008	10815 (98.18)	201 (1.82)	11016 (100)
2010	10392 (95.05)	541 (4.95)	10933 (100)
2012	11415 (93.93)	738 (6.07)	12153 (100)
2013	11372 (93.38)	806 (6.62)	12178 (100)
2014	11447 (93.41)	807 (6.59)	12254 (100)
Total	65763 (92.25)	3279 (4.75)	69042 (100)

Sources: Hospital quality reports 2006-2014, G-BA

TABLE 9 – EFFECT OF PEPP ON LOG OPS PER CASE, ORIGINAL DATASET

	Dependent variable: ln (OPS per case)		
	Model (1)	Model (2)	Model (3)
TREAT	0.142***	0.151***	0.161***
	(0.007)	(0.010)	(0.010)
PSYCH	-1.128***	-1.171***	-1.127***
	(0.080)	(0.048)	(0.050)
TREAT × PSYCH	2.555***	2.628***	2.581***
	(0.082)	(0.052)	(0.053)
OWN = not-for-profit		-0.036***	-0.040***
		(0.007)	(0.008)
OWN = for-profit		0.010	0.021*
		(0.009)	(0.010)
TEACH		0.193***	0.187***
		(0.007)	(0.007)
PHYS		0.005***	0.005***
		(0.001)	(0.001)
Constant	0.477***	0.263***	0.350***
	(0.006)	(0.028)	(0.033)
Full set of STATE dummies	No	No	Yes
Adj. R-squared	0.123	0.135	0.138
N	65096	65096	64245
BIC	164817.6	163998.5	161502.6

Sources: Hospital quality reports 2006-2014, G-BA
Notes: All three models in the table are estimated using OLS regression. Standard errors in parentheses (model 1: robust standard errors; Breusch-Pagan test for heteroskedasticity: model 1: p=0.000; model 2: p=0.690; model 3: p=0.513).
*: Significant at the 5 percent level.
**: Significant at the 1 percent level.
***: Significant at the 0.1 percent level

Appendix B

In the following, the summary statistics (Table 10), the number of wards by year and type (Table 11), and the regression results (Table 12) for the subsample (pediatric cases) are presented.

TABLE 10 – SUMMARY STATISTICS, PEDIATRIC SUBSAMPLE

	N	Mean	SD	Min.	Max.
Year	2480	2010.7	2.744194	2006	2014
=1 if year>2009	2480	0.702	.4572863	0	1
Hospital level					
Cases, inpatient	2480	23568.0	18688.13	0	135981
Cases, semi-stationary	1742	1069.8	2179.908	0	30689
Cases, outpatient	1332	73733.8	166490.4	0	1443326
Number of beds	2480	598.3	463.8807	0	3213
=1 if teaching hospital	2480	0.721	.448415	0	1
=1 if ownership not-for-profit	2480	0.313	.4639314	0	1
=1 if ownership private	2480	0.181	.3851359	0	1
Ward level					
Cases, inpatient	2480	1964.5	1393.643	0	6832
Cases, semi-stationary	1742	122.2	362.6508	0	4147
Physician	2392	17.11	27.90747	1	429
Med. specialist	2392	8.438	16.09524	0	249
Sum of OPS	2480	2671.1	2452.095	0	20703
OPS per case	2480	2.407	3.962188	0	40.84953
=1 if psychiatric/ psychosom. ward	2480	0.183	.3864665	0	1

Sources: Hospital quality reports 2006-2014, G-BA

TABLE 11 – NUMBER OF WARDS BY YEAR AND TYPE, PEDIATRIC SUBSAMPLE

Year	Somatic (Percent)	Psych. (Percent)	Total (Percent)
2006	325	24	349
	(93.12)	(6.88)	(100)
2008	360	29	389
	(92.54)	(7.46)	(100)
2010	333	77	410
	(81.22)	(18.78)	(100)
2012	345	100	445
	(77.53)	(22.47)	(100)
2013	335	109	444
	(75.45)	(24.55)	(100)
2014	329	114	443
	(74.27)	(25.73)	(100)
Total	2027	453	2480
	(81.73)	(18.27)	(100)

Sources: Hospital quality reports 2006-2014, G-BA

TABLE 12 – EFFECT OF PEPP ON LOG OPS PER CASE, PEDIATRIC SUBSAMPLE

	Dependent variable: ln (OPS per case)		
	Model (1)	Model (2)	Model (3)
TREAT	0.015	0.015	0.018
	(0.036)	(0.048)	(0.048)
PSYCH	-0.178	-0.187	-0.136
	(0.125)	(0.129)	(0.131)
TREAT × PSYCH	2.453***	2.467***	2.422***
	(0.133)	(0.139)	(0.142)
OWN = not-for-profit		-0.117**	-0.124**
		(0.036)	(0.038)
OWN = for-profit		-0.142**	-0.101*
		(0.043)	(0.045)
TEACH		0.034	0.020
		(0.035)	(0.036)
PHYS		0.000	-0.000
		(0.004)	(0.005)
Constant	-0.230***	-0.198	-0.213
	(0.030)	(0.126)	(0.143)
Full set of STATE dummies	No	No	Yes
Adj. R-squared	0.544	0.547	0.558
N	2445	2445	2410
BIC	5695.2	5708.0	5678.5

Sources: Hospital quality reports 2006-2014, G-BA

Notes: All three models in the table are estimated using OLS regression. Standard errors in parentheses. The Breusch-Pagan test (model 1: p=0.346; model 2: p=0.577; model 3: p=0.615) clearly shows that there is no heteroskedasticity for the models in this subsample, so that robust standard errors are not used here.

*: Significant at the 5 percent level.

**: Significant at the 1 percent level.

***: Significant at the 0.1 percent level

Appendix C

In the following, OLS regression assumptions will be discussed briefly in order to address regression diagnostics. First, linearity in parameters is assumed. However, further testing of this assumption is not possible.

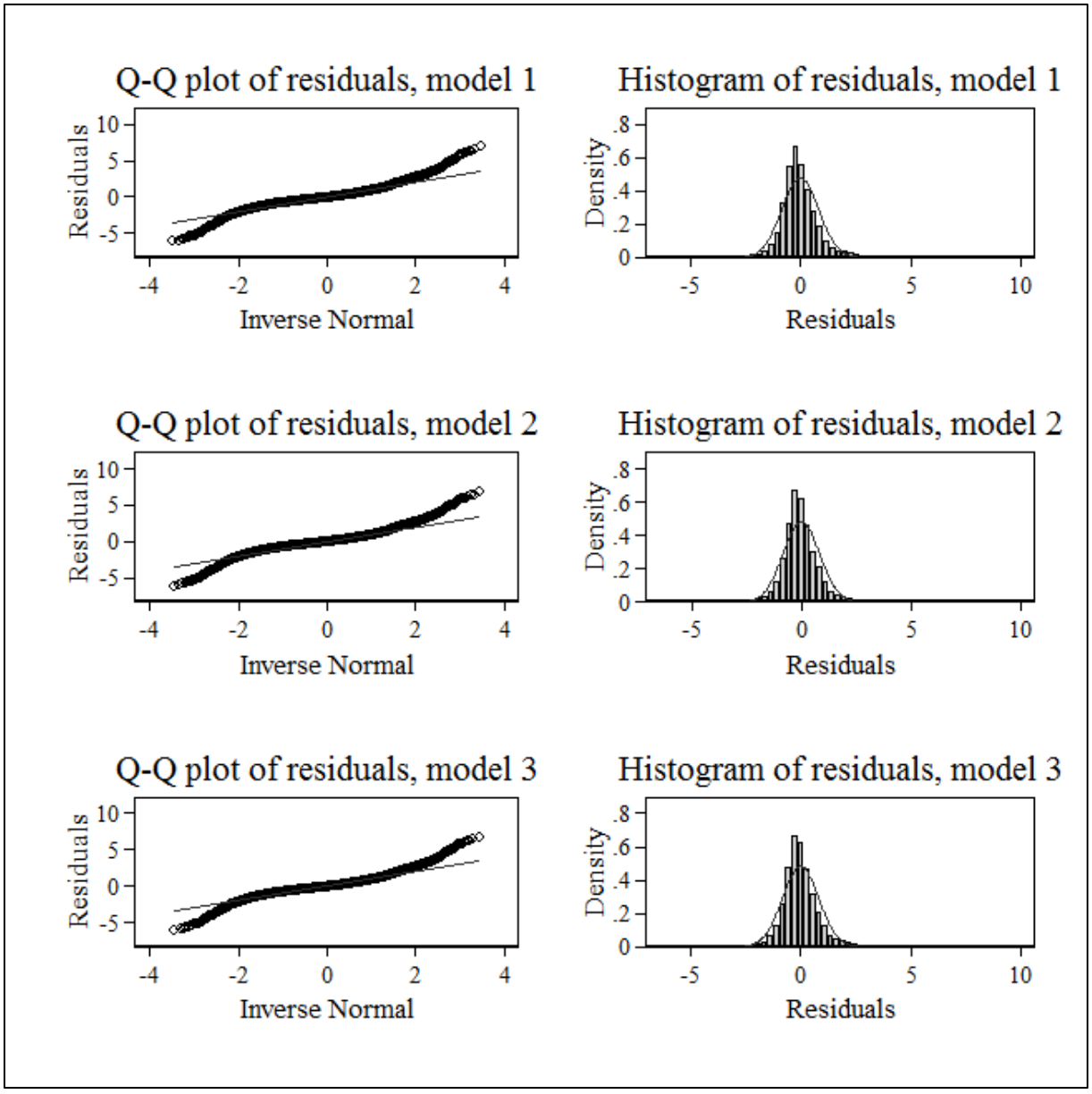

FIGURE 4 – DIAGNOSTIC PLOTS FOR NORMALITY

Sources: Hospital quality reports 2006-2014, G-BA (own illustration)
Notes: The solid line in the three histogram denotes the normal distribution

Concerning the assumption of zero conditional mean, the mean of the predicted residuals of model (1) of 4.31e-9 (model 2: 8.07e-10; model 3: 1.13e-11) indicates that this assumption holds. Although a non-normal distribution of the error terms must be assumed based on

assessment of diagnostic plots for the models' residuals, as displayed in Figure 4[6], I argue that linear regression is still applicable due to the very large sample size. This is also in line with the central limit theorem, which is no further tested, however. In addition, Lumley et al. (2002) present evidence that violation of this assumption is rather neglectable in large datasets. The Breusch-Pagan test has been used to test for heteroskedasticity. As it clearly rejects the hypothesis of homoscedasticity, i.e. equal variance of the error terms, for model 1 and model 3 (model 1: p=0.000; model 2: p=0.152; model 3: p=0.014), robust standard errors are used in the following for these models. Issues concerning specification of the empirical model are further discussed in the limitations.

In addition to the OLS assumptions, the difference-in-differences approach requires another characteristic assumption, called common trend or parallel paths assumption. Following visual inspection of Figure 3, I assume that this assumption holds.

[6] The use of most specific normality tests, e.g. the Shapiro-Wilk or the Shapiro-Francia test, is very restrictive concerning sample sizes. In fact, it is not recommended to use them in samples as large as the one in this thesis.

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit ohne fremde Hilfe selbständig verfasst habe. Ich habe keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellverzeichnis nicht benannten Internet-Quellen – benutzt. Ich habe die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht. Die schriftliche Fassung entspricht der auf dem elektronischen Speichermedium.

Hamburg, 22.11.2016

_____

Jakob Everding