Incentives in Health Care Provision

Dissertation zur Erlangung des Doktorgrades der Wirtschafts-wissenschaften (Dr. oec.)

Fakultät Wirtschafts- und Sozialwissenschaften

Universität Hohenheim

Institut für Health Care & Public Management

vorgelegt von

Johanna Katharina Weinert

aus *Hamburg*

2022

Datum der mündlichen Prüfung: 20.12.2022

Dekan der Fakultät Wirtschafts- und Sozialwissenschaften: Prof. Dr. Jörg Schiller

Erstgutachter: Prof. Dr. Jörg Schiller

Zweitgutachter: Prof. Dr. Christian Ernst

Prüfungsvorsitz: Prof. Dr. Katja Schimmelpfeng

DANKSAGUNG

Ich möchte mich ganz besonders bei meinem Doktorvater Herrn Professor Jörg Schiller bedanken, der mir die Möglichkeit gegeben hat, diese Dissertation zu schreiben. Ich bin sehr dankbar dafür, dass ich mich während meiner gesamten Promotionszeit mit fachlichen Fragen und Problemen an ihn wenden konnte und immer konstruktive Unterstützung erhalten habe. Des Weiteren möchte ich mich auch herzlich für die Unterstützung meines Zweitgutachters Herrn Professor Christian Ernst bedanken. Mein Dank geht auch an Frau Professorin Katja Schimmelpfeng für die Übernahme des Prüfungsvorsitzes meiner mündlichen Prüfung.

Außerdem möchte ich mich bei allen Mitarbeiterinnen und Mitarbeitern des Instituts bedanken, die diesen Ort während meiner Dissertationszeit zu so einem schönen Arbeitsplatz gemacht haben. Insbesondere möchte ich mich bei Tanja, Kylie, Martin, Christopher, Lena und Vanessa für die immer tollen und unterhaltsamen Mittagessen und Kaffeepausen bedanken. Bedanken möchte ich mich weiterhin auch bei Herrn Dr. Simon Dally von der AOK-BW, der mir stets mit Rat und Tat zur Seite stand.

Besonders bedanken möchte ich mich auch bei meinen Eltern Claudia und Ulrich, die immer an mich geglaubt haben und ohne deren Unterstützung diese Arbeit niemals möglich gewesen wäre.

Und natürlich möchte ich mich von ganzem Herzen bei meinem Ehemann Markus bedanken, der mich während der Promotionszeit einige Jahre als Kollege begleitet hat und dessen Unterstützung mich auch durch teils schwierige Phasen während der Fertigstellung der Arbeit getragen hat. Ich möchte mich für seine Liebe, seinen Glauben an mich und seine unendliche Geduld bedanken.

Contents

List of Abbreviations

ATE	Average treatment effect
ATET	Average treatment effect on the treated
DDD	Difference-in-difference-in-difference
FCODS	First case of the day start
FFS	Fee-for-service
OOP-FFS	Fee-for-service paid out-of-pocket
OR	Operating room
SCPC	Selective contract for pediatric care

List of Tables

List of Figures

1 Introduction

The importance of finding ways to ensure high quality health care provision in a cost-effective and -efficient way becomes a more and more pressing issue given the challenges many economies currently face such as ageing populations (Marešová et al., 2015), rising costs caused by advancements in medical technology (Sorenson et al. 2013) and an increasing shortage in qualified personnel (Marć et al., 2019). To address this challenge, economic research analyzed a myriad of different incentive schemes aimed at increasing cost-effectiveness and cost-efficiency including prominent examples such as pay-for-performance, capitated reimbursement or public reporting of health outcomes. Many of these schemes have found their way to real-world implementation and although they oftentimes proved valuable in achieving the desired objectives, they frequently gave rise to unintended changes in health care providers' as well as patients' behavior. Hence, gaining a better understanding of how incentive schemes affect providers' as well as patients' behavior is crucial. This thesis aims at contributing to the existing knowledge by analyzing the effects and implications of different incentive schemes on health care providers' behavior. In the following three chapters, non-monetary as well as monetary incentive schemes targeting in- and outpatient health care service provision are empirically and theoretically examined.

Chapter 2 analyzes empirically how the introduction of a surgical suite governance document affects punctuality in first case of the day starts. Delays in first cases are an indicator for inefficiencies in operating room utilization as their occurrence is directly associated with lower OR utilization rates caused by delayed starts of subsequent surgeries (Does et al., 2009; Van Veen-Berkx et al., 2014; Dexter & Epstein, 2009; Szczesny & Ernst, 2016). Because operating rooms constitute a major driver in hospitals' operating costs (Cardoen et al., 2010), clinic management has a strong interest in incentivizing an efficient use of this resource. This analysis focuses on the implementation of a surgical suite governance document, which explicitly specifies the starting time of the first case of the day and formulates scheduling rules. First case punctuality is an easily observable and measurable performance indicator, which is associated with only minor tracking efforts and consequently low costs. The analysis uses a quasi-experimental setting, which arose from the lagged implementation of an identical governance document in two different hospital sites belonging to the same hospital group. To

assess the question whether the introduction of a governance document was associated with significant reductions in first case delays, a difference-in-difference estimation approach has been implemented. Results indicate that the introduction of a surgical suite governance document is associated with significant reductions in first case delays. The analysis shows that the document's introduction lead to an estimated decrease by roughly one third in first case delays. In conclusion, a surgical suite governance document seems to offer a promising tool to incentivize health care workers' to use costly resources like surgery capacities more efficiently.

The focus of Chapter 3 is the analysis of the effects of a reimbursement change - from fee-for-service paid out-of-pocket (OOP-FFS) to a capitation fee per patient - on health service provision. This change was part of a selective contract in outpatient pediatric care introduced by a large German sickness fund in 2014. Over the past few years, several German sickness funds have implemented selective contracts in outpatient care with the objective to increase quality and cost-effectiveness in health care service provision. Thus, the present analysis aims at deriving further insights on how reimbursement affects service provision and at offering guidelines for future designs of selective contracts. To reflect the special features of the analyzed selective contract, namely that incentives change for both the pediatricians and the patients simultaneously, a theoretical model is set up to derive a testable hypothesis. The model predicts that given pediatricians are not only monetarily incentivized (but also sufficiently concerned about patients' well-being) and that costs associated with screening provision are relatively small, reimbursement change from OOP-FFS to capitation will induce an increase in service provision. In the scheme at hand, participation in the selective contract scheme is neither mandatory for physicians nor for patients, which offers a natural experiment setting as we have observations for both participating and non-participating pediatricians. Using a generalized difference-in-difference approach, the theoretically derived hypothesis is tested empirically. Results indicate that the change from fee-for-service paid out-of-pocket to a capitation fee per patient did lead to a significant increase in provided screenings as the number of diagnoses more than doubles for pediatricians enrolled in the selective contract scheme. These findings indicate that physicians are not solely driven by monetary incentives and that capitation per patient offers a valuable tool to ensure cost control yet simultaneously ensure effective health care provision.

Finally, Chapter 4 examines theoretically to which extent policymakers are able to incentivize hospitals to increase quality provision by actively fostering the link between performance indicator reporting and hospitals' reputation. Empirical findings indicate that potentials for further improvements in health care quality provision still exist (Howell & Zeitlin, 2017) and that quality provision between different hospitals varies (Tsai et al., 2013; Merchant et al., 2012). This makes a better understand of how policymakers can affect quality incentives crucial. By fostering the link between outcome-based performance indicators and reputation, policymakers are able to affect hospital's market share and thereby ultimately hospitals' incentives for quality provision. Ways to strengthen the link are manifold, e.g. by raising awareness about the existence and importance of hospital performance reports or by improving populations' health literacy to ensure that patients are able to decode the information contained in performance indicators properly. Choosing a dynamic framework allows to reflect the fact, that hospitals' reputation resembles a stock variable evolving over time depending on hospital managers' quality decisions in preceding periods. The main finding is that a strengthened link between performance indicators' realization and hospitals' reputations does not necessarily result in stronger incentives for quality provision. In the case where the degree of competition is sufficiently low and the costs associated with quality provision are sufficiently high, an intensified link between performance indicators and reputation induces a decrease in quality provision. If the opposite is true, strengthening the link between performance indicators' realization and hospitals' reputations always results in increases incentives for quality provision.

The thesis concludes with a short summary of key results in chapter 5.

2 Introducing a surgical suite governance document: Effects on first case punctuality in an orthopedic department[1]

Abstract

In this paper, we analyze how the implementation of a surgical suite governance document affects first case punctuality of elective surgeries in an orthopedic department. We exploit the quasi-experimental setting given by the lagged introduction of an identical surgical suite governance document in two different hospital sites belonging to the same hospital group. For our difference-in-difference estimation approach we used clinical data covering the period from April 2012 to March 2015 encompassing information with respect to patients' characteristics (i.e. age, gender and co-morbidities), diagnoses (ICD-10 codes), procedures (OPS codes) and self-reported process times (e.g. incision start times). We restricted our analysis to cases, for which patient admission took place before the day of surgery. In doing so, we prevented the erroneous inclusion of emergency cases and we aimed at overcoming the problem that delays in same-day surgeries oftentimes are attributable to unpunctual arrival of patients, thereby being beyond the control of operating room (OR) staff. Applying a difference-in-difference estimation approach, we find that the introduction of a surgical suite governance document led to significant decreases in morning delays of around eight minutes on average. Furthermore, our findings indicate that OR staff only partially adhere to surgical suite governance document's guidelines. Nevertheless, OR staff was able to improve punctuality also for cases not directly targeted by the OR guidelines, indicating that awareness about deficiencies exists among OR staff and that the latter is able to autonomously develop strategies to overcome those deficiencies.

[1] This chapter is based on joint work at the University of Hohenheim with Martin Richter and Tanja Wollensak, which is yet unpublished.

2.1 Introduction

Operating rooms represent a major cost factor in hospitals (Cardoen et al., 2010) due to high operating and personnel costs.[2] Combined with rising cost pressure in the provision of inpatient health care services induced by the introduction of prospective reimbursement schemes in Germany in 2004 (Szczesny & Ernst., 2016), hospitals have a strong incentive to design OR processes efficiently (Geldner et al., 2002). To improve efficiency, hospitals and OR management need to agree on strategies, which define rules and performance indicators suitable to manage multidisciplinary teams and functional units. Contrary to other industries, where it is common to assess efficiency via financial performance indicators, researchers and hospital practitioners tend to rely on process performance indicators to evaluate OR efficiency (Schuster at al., 2007).

In this context, punctuality of the first case of the day is considered as a suitable key performance indicator for process efficiency (Wong et al., 2010; Van Veen-Berkx et al., 2014) and interdisciplinary team performance (Overdyk et al., 1998; Schuster et al., 2007). Delays in the first case of the day impact OR utilization directly by causing delays in subsequent surgery starts (Does et al., 2009; Van Veen-Berkx et al., 2014; Dexter & Epstein, 2009; Szczesny & Ernst, 2016). Empirical evidence indicates that tardiness is directly associated with real costs induced by OR over-utilization, where increasing labor costs arising due to compensation payments for overtime working hours constitute a major driver for rising costs (Dexter & Epstein, 2009). Furthermore, delays in the first case of the day can cause cancellation of surgeries scheduled later in the day (McIntosh et al., 2006, Dexter and Epstein, 2009). Another important aspect is that punctuality appears to have a signal effect on OR teams' working discipline (Overdyk et al., 1998; Schuster et al., 2007) thereby also affecting OR efficiency indirectly.

Both process and infrastructural factors can cause morning delays. Late arriving anesthesiologists and surgeons (Wright et al., 2010; Wong et al., 2010; Darwish et al., 2016); delays in preparation of operating room set-ups (Wong et al., 2010); and anesthesia induction complexity and insufficient communication between OR and ward personnel (Schuster et al., 2013; Wong et al., 2010; Unger et al., 2009; Geldner et al., 2002) constitute processual factors.

[2] For patients undergoing surgical intervention, about 60% of the treatment costs are incurred in the OR (Berry et al. 2008). International literature estimates costs of (U.S.) $15-20 per minute of staffed OR time for OECD countries (Macario, 2010; Park & Dickerson, 2009).

Examples for infrastructural factors are long transportation distances between wards and ORs, the existence of bottlenecks (e.g. elevators) and the lack of pre-operating waiting area capacity (Szczesny & Ernst, 2016). While overcoming infrastructural factors typically requires substantial investments in construction (such as new buildings or faster and more spacious elevators), processual factors offer a less cost-intensive starting point for improvement measures. To raise process efficiency, e.g. by reducing delays, hospital managers can implement educational programs and incentive schemes fostering team-orientated behavior and adherence to schedules. Several studies have analyzed the effect of educational interventions on processual performance indicators. Scientifically accompanied field trials investigated the impact of staff education and training programs (Truong et al., 1996; Forse et al., 2011), staff education combined with the introduction of target times (Overdyk et al., 1998), and performance reporting and financial rewards (Vitez & Macario, 1998; St. Jacques et al., 2004; Scalea et al., 2014). Key findings were that educational programs seem to have a positive impact on first case of the day starts (FCODS) (Tuong et al. 1996; Overdyk et al. 1998) but that sustainability of improvements requires constant repetition and refreshment of training content. Besides, evidence for cost-effectiveness of financial rewards are ambiguous as difficulties exist with regard to the calculation of real cost savings (Scalea et al., 2014).

An alternative approach to improve process efficiency is the introduction of a surgical suite governance document (to which we refer as “OR charter” in the remainder of this paper), which explicitly specifies process guidelines by setting target times and formulating scheduling rules (Ernst et al., 2012; Bohnenkamp & Braun, 2017). Major advantages of OR charters are minor implementation costs and relatively low tracking requirements (Ernst et al., 2012). Evaluating staffs’ performance relies on easy observable and verifiable performance indicators, such as deviations from targeted starting times, and does not require tracking of individual behavior. Although OR charters are a widespread management tool in German hospitals, surprisingly few studies have analyzed the effects of OR charter implementation on FCODS so far. Ernst el al. (2012) conducted a survey among chief anesthesiologists to assess OR charter’s effectiveness in reducing tardiness of FCODS. Their results indicate that the adoption of an OR charter seems to be associated with a reduction in delays. The authors state that their analysis faces some limitations as it relies on self-reported data on delays, which implies the risk of possible inaccuracy or bias. Additionally, the study builds on a cross-

sectional data set making it impossible to control for unobservable but relevant hospital characteristics.

The present paper contributes to the literature by analyzing the effect of OR charter adoption on FCODS using data from two orthopedic surgery departments belonging to the same German public non-profit clinic group. This represents an area where little research has been conducted so far, with most of OR management literature focusing on tertiary centers. In this context, our contribution is twofold. First, we have a panel data structure as we use repeated processual data for the same two hospital sites covering an observation period of several years. Second, the fact that both departments introduced an identical OR charter at different points of time offers a quasi-experimental design. We analyze whether OR charter adoption affected average morning delays and test the hypothesis that introduction improved punctuality. Additionally, we assess whether hospitals followed the guidelines specified in the OR charter. To the best of our knowledge, the effects of OR charter adoption has never been studied before using archived data in a quasi-experiment setting. We are able to confirm the central findings of Ernst et al. (2012), namely that the implementation of an OR charter constitutes an effective tool in morning delay reduction.

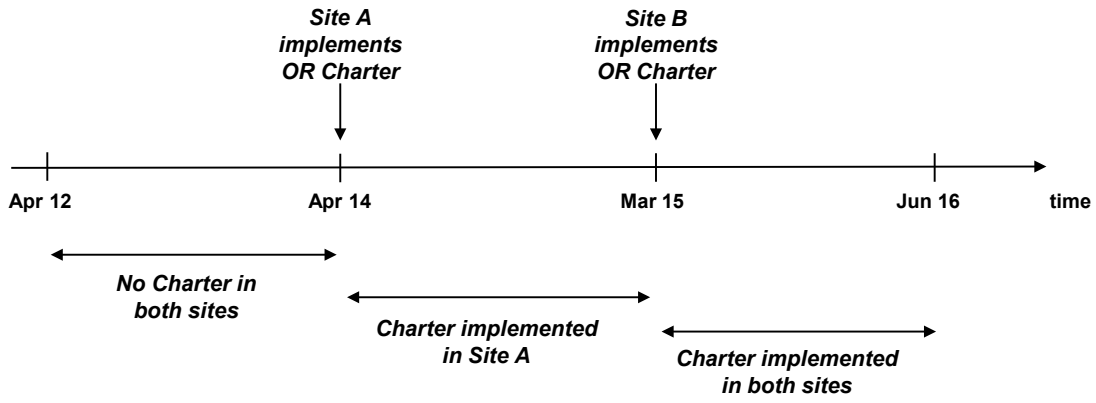
2.2 Background and hypothesis

The provision of surgical interventions requires multiple stakeholders from different professions to work together, which makes the improvement of coordination and cooperation between team members a crucial objective for efficiency improvements (Gfrörer et al. 2005). An OR charter constitutes a written policy statement intended to foster interdisciplinary teamwork by introducing a binding set of rules (Bohnenkamp & Braun, 2017) in order to ensure a more efficient use of OR capacity (Hensel et al., 2005). Generally, an OR charter encompasses guidelines with respect to planning and scheduling rules (Geldner et al., 2002), rules for perioperative processes (Geldner et al., 2002) and target times for FCODS (Ernst et al., 2012). This OR management tool is associated with several advantages. Firstly, the OR charter requires minimum personal tracking efforts, because the measured performance indicators base on stakeholders' jointly performance. Secondly, neither costly monetary incentives nor educational measures are required and finally, this tool offers a high degree of flexibility, as guidelines are easily adaptable to the needs and characteristics of individual

hospitals (Hensel et al., 2005). The latter aspect stresses the importance of an appropriate guideline design in order to unfold potential advantages making a diligent and self-critical analysis of the as-is state a prerequisite for the success of this management tool (Geldner et al., 2002). Additionally, it is important to ensure practicality of guidelines, as employees may otherwise be reluctant to adhere (Bohnenkamp & Braun, 2017). The integration of all affected interest groups and stakeholders is necessary to guarantee acceptance among the target group (Pfannstiel, 2014). To ensure legitimacy and signal the binding nature of OR guidelines to all stakeholders, the OR charter should be directly enacted by hospital's management (Pfannstiel, 2014). Finally, guidelines have to be designed cautiously, as only little understanding exists so far on how an OR charter actually affects individual behavior and working processes. Ernst et al. (2012) hypothesize that positive effects are possibly attributable to peer-pressure on individuals.

In this paper we study the impact of an OR charter introduction on first case tardiness in a German public hospital. Our analysis uses data from two hospitals belonging to a public-owned hospital group operating several hospital sites in south-west Germany. The observation period comprises more than four years starting from April 2012 and ending in June 2016. Both hospital sites introduced an identical OR charter document at two different points of time, where site A adopted the charter on 1 April 2014 and site B on 1 March 2015 (see Figure 2.1). The OR charter's key element is the specification of a target time for incision start at 08:15 AM. Furthermore, the document stipulates that the surgical intervention scheduled as first case of the day should not require peripheral pain therapy and that ward rounds have to be planned and executed in such a way that they do not interfere with punctuality of FCODS. With regard to possible sanctioning mechanisms, the OR charter states that if repeated violations against charter guidelines occur, the responsible head physician has to report to the clinic management and has to justify the reasons for those violations.

Figure 2.1: Scheduling of intervention



In the beginning of the observation period, neither hospital site A nor site B used an OR charter. Then, hospital management introduced an OR charter in Site A in April 2014 and implemented the exact same document eleven month later in March 2015 in site B. The resulting lag in OR charter implementation offers a quasi-experimental setting, where site A experiences treatment and site B acts as the control. We estimate a difference-in-difference model to analyze whether OR charter introduction in site A was associated with a decrease in first case delays.

2.3 Data and Methods

2.3.1 Data and Variables

The data set covers an observation period from April 2012 to June 2016 and encompasses information with respect to patients' characteristics (i.e. age, gender and co-morbidities), procedures (OPS codes), diagnoses (ICD-10 codes), diagnosis-related groups (DRGs), information with respect to hospital stay (e.g. day of admission, day of surgery) and self-reported OR process times (e.g. incision start times).

As our approach seeks to exploit the quasi-experimental character of the data set, we only consider observations until site B introduces the OR charter in March 2015 (and no longer serves as control). Thus, we restrict our data set to the period between April 2012 to February 2015.

A shortcoming of the data set is the lack of differentiation between elective and emergency surgeries. Because our interest lies in the impact of OR charter introduction on plannable OR processes, we solely focus on elective procedures. Therefore, we had to apply a valid strategy

to identify emergency cases. We excluded surgeries taking place outside the regular operating hours, namely surgeries on weekends and public holidays. We also omitted all surgeries, for which admission took place on the day of surgery (we refer to this type of interventions as "same day surgeries" in the remainder of this paper) to ensure that the data set does not contain any emergency surgeries. Additionally, limiting our analysis to non-same day surgeries is sensible in this context, as OR staff has only limited impact on punctuality of same day surgery starts. For this type of intervention, timeliness in FCODS depends crucially on patients' punctual arrival in the hospital. A study by Wong et al. (2010) aggravates this concern as their findings show that delays in same day surgery starts are in many cases attributable to late arriving patients. This implies that OR staff has only limited influence on timeliness of same day surgeries starts compared to elective non-same day surgeries and that the latter offer more starting points for improvements, like e.g. enhancements in inter-departmental communication and coordination of patient transports. Finally, in order to ensure that observed deviations are not attributable to external factors out of hospital's staff control (e.g. technical malfunctions etc.), we choose an approach analogous to Schuster et al. (2013) by formulating a critical threshold for cases included in the data set. As the scheduled starting time was 8:15 AM for both hospital sites during the complete observation period, we set thresholds such that only cases starting after 7:50 AM and before 9:00 AM were included in the data set.

In the following analysis, our outcome variable $deviation$ measures deviations from scheduled starting time. To measure $deviation$ we take the difference between the actual start of the first surgery of the day and the targeted start time at 8:15 AM (where we consider both delayed and early starts). Table 2.1 shows the average deviations for FCODS separately for site A and site B before and after OR charter implementation in April 2014. Before charter introduction, the average delay was substantially higher for site A than for site B (21.53 vs. 13.20 minutes). The difference between sites diminishes in the post-treatment observation period as average delay decreases to 15.33 minutes for Site A and slightly increases to 14.09 minutes for site B.

Table 2.1: Outcome variable measuring deviation from target time in first case of the day before and after introduction of OR charter – Grouped by hospital site

Before OR charter introduction (Apr 2012 – Mar 2014)

Outcome	Site A					Site B				
	Obs.	Mean	SD	Min	Max	Obs.	Mean	SD	Min	Max
Deviation (in minutes)	326	21.531	9.75	-3	44	365	13.202	8.89	-5	44

After OR charter introduction (Apr 2014 – Feb 2015)

Outcome	Site A					Site B				
	Obs.	Mean	SD	Min	Max	Obs.	Mean	SD	Min	Max
Deviation (in minutes)	256	15.328	9.25	-5	44	370	14.092	9.59	-10	43

Figure 2.2: Distribution of deviations from targeted incision time – Grouped by hospital site

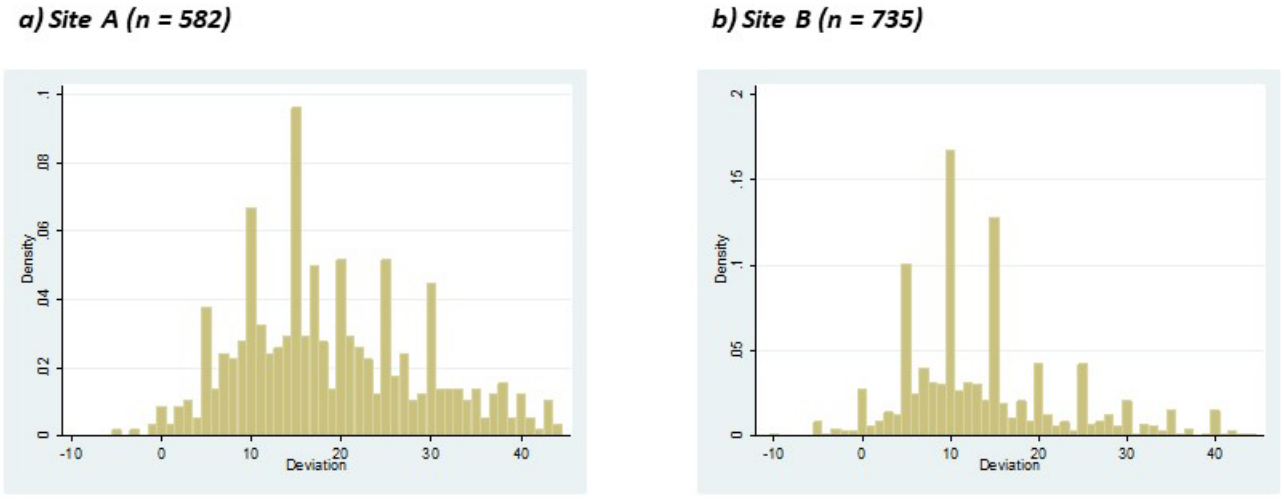


Figure 2.2 shows the distribution of the outcome variable $deviation$ for site A and site B separately. Deviations from targeted incision time seem normally distributed for both sites. The distributions exhibit a pattern of accumulated observations in five-minute intervals, which indicates the presence of a measurement error possibly provoked by rounding errors. These observation clusters are much more prominent for site B and smooth out in both sites over time (Appendix I). A possible explanation for these observation clusters might be OR personnel's preference for rounding starting times instead of accurately documenting them.

The increase in reporting accuracy over time (Appendix I) is likely attributable to rising OR personnels' awareness for documentation quality. Measurement error in the outcome variable poses no threat towards the validity of OLS coefficient estimates as long as no correlation between covariates and rounding errors exists (Wooldridge, 2013). Here, we have no indication to believe that systematic under- or over-reporting of delays took place as rounding was much more prevalent in site B. In this site, no sanctions for delays had been installed during the complete observation period and therefore no systematic incentives for rounding were present.

Table 2.2: Summary statistics for non-same-day surgeries – Grouped by hospital site

	Site A				Site B			
Variable	Obs.	Mean (SD)	Min	Max	Obs.	Mean	Min	Max
Deviation (in minutes)	582	18.802 (10.01)	-5	44	735	13.650 (9.25)	-10	44
Age65	582	0.613 (0.49)	0	1	735	0.722 (0.45)	0	1
Female	582	0.529 (0.50)	0	1	735	0.604 (0.49)	0	1
Obesity	582	0.199 (0.40)	0	1	735	0.114 (0.32)	0	1
Diabetes	582	0.119 (0.32)	0	1	735	0.158 (0.36)	0	1
Surgtime (in minutes)	582	80.519 (36.28)	6	275	735	82.219 (43.32)	2	406
pain	582	0.271 (0.45)	0	1	735	0.224 (0.42)	0	1
Ward	582	0.174 (0.38)	0	1	735	0.193 (0.40)	0	1

Table 2.2 gives an overview of variables characterizing hospitals' patient structure separated by hospital sites. We included binary variables indicating patient's gender ($Female$) and age ($Age65$), where the latter variable becomes zero for all patients younger than 65 years and one otherwise. Furthermore, we introduced variables indicating whether a patient suffers from diabetes ($Diabetes$) or obesity ($Obesity$) as empirical evidence suggests that both

diabetes[3] and obesity[4] are associated with prolonged surgery preparation times. We used ICD-10GM-codes (International Classification of Diseases 10th revision, German Modification) to identify patients suffering from diabetes and/ or obesity. Additionally, we included surgical time in minutes[5] (*Surgtime*) as a proxy for surgery complexity. The reasoning is that more complex surgeries are associated with longer preparation times, implying that complex surgeries are more prone to delays. [6] Head physician's morning ward rounds (*ward*) might pose another threat to timely FCODS, as participating surgeons might experience time conflicts due to prolonged ward rounds resulting in delayed arrivals in the operating room (Schuster at al. 2007). Furthermore, we use information on whether a patient receives a complex peripheral anesthesia induction as additional pain therapy (*pain*). This is of interest as research suggests that complex anesthesia procedures are also associated with delays in morning starts (Unger et al. 2009).

We grouped summary statistics by hospital sites to get a better understanding of differences between the two sites. Table 2.2 shows that patients are on average older in site B compared to site A, as 72 per cent of individuals are older than 65 years in site B compared to 61 per cent in site A. Furthermore, site A has an almost balanced patient sex ratio (53 per cent of patients are female), whereas site B exhibits a higher share of female patients (60 per cent). The proportion of obese patients is almost twice as high for site A (0.20) compared to site B (0.11). In contrast, the share of patients suffering from diabetes is slightly higher for site B than for site A (0.16 vs. 0.12). As occurrence of diabetes correlates with older age (Hassing et al., 2004) this is in line with the observed age structure. With respect to average surgery time, we see only a minor difference between both sites (80.5 minutes in site A and 82.2 minutes in site B). Furthermore, we observe a slightly higher share of interventions requiring peripheral pain therapy in site A compared to site B (27 per cent vs. 22 per cent). For both sites, almost one fifth of observed first cases take place on days with ward rounds (0.17 in site A and 0.19 in site B).

[3] Higher complexity of intraoperative management for patients with diabetes can lead to delays in surgery start (Schiff & Emanuele, 1995).

[4] Obesity is associated with prolonged surgery preparation times due to additional requirements for positioning and anesthesia induction (Raphael et al., 2013).

[5] Surgery time measures the time span between incision and suture.

[6] Current research indicates that applying the scheduling rule 'shortest processing time first' (SCF) constitutes a suitable tool to reduce morning delays (Marcon & Dexter, 2006).

2.3.2 Estimation Strategy

We use a difference-in-difference estimation approach to assess the impact of OR charter introduction on morning delays. The setting at hand resembles a quasi-experiment, where the introduction of the OR charter in site A constitutes the treatment. A key assumption for the validity of the difference-in-difference approach is that in the absence of OR charter adoption, deviations from scheduled starting time would have developed identically over time for both the treatment and the control group. As this is a hypothetical situation, we cannot directly test whether this requirement is satisfied. Thus, we explore whether pre-treatment development of deviations are similar for both sites and assume that the existence of parallel pre-trends is a valid indicator for a parallel development of deviations in the absence of treatment.

Figure 2.3: Time trends for deviations from targeted incision time from April 2012 to February 2015 measured in half-year time intervals (where the sixth (last) interval encompasses only five months due to the data structure) (n=1,317)

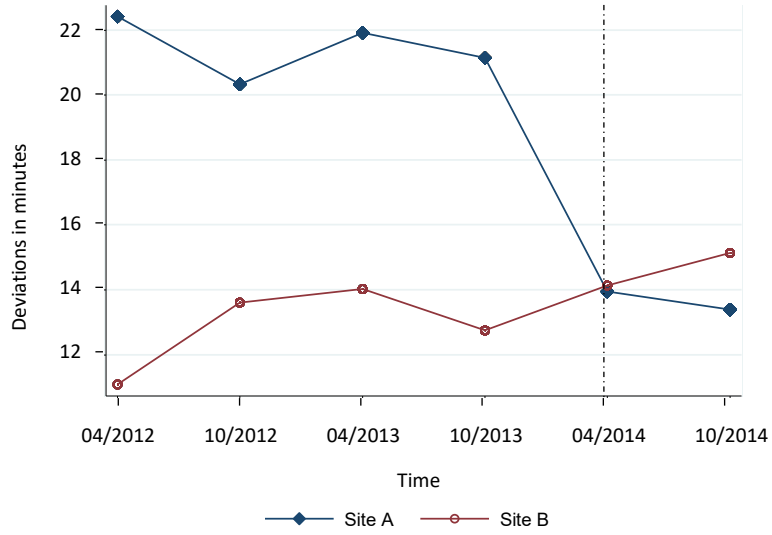


Figure 2.3 depicts how average delay of non-same day surgeries develops over time. Here, the blue line represents average delays for site A and the red line for site B. [7] From visual inspection we can see, that time trends develop almost parallel for both sites in the pre-treatment period, although the time trends differ with respect to levels with a substantially lower level for site B. Furthermore, Figure 2.3 shows that, between the first and the second

[7] See Appendix II for graphical illustration of deviations for the complete data set as well as for surgeries taking place on the day of admission (same day surgeries).

half-year interval, average deviation develops in opposite directions, with a decrease in delays for site A and an increase in delays for site B. However, from October 2012 (the second half-year of observations) onwards, time trends develop parallel until treatment takes place in April 2014. To assess, whether the parallel trend assumption necessary for the validity of the difference-in-difference estimation strategy is satisfied, we conduct further tests on pre-treatment trends (see Appendix III & IV). We introduced placebo treatments (Appendix III) and linear time trends reflecting general time trends in morning delays (Appendix IV) to explore the validity of the parallel trend assumption. Our estimation results show that we cannot reject the null hypothesis of parallel pre-trends. Consequently, we argue that the parallel trend assumption is likely to hold which indicates the internal validity of the difference-in-difference estimation approach. Therefore, we run the following difference-in-difference estimation model:

$$y_{itn} = \beta_0 + \beta_1 post_{itn} * siteA_{itn} + \beta_2 siteA_{itn} + \beta_3 ward_{itn} + \beta_4 pain_{itn} + X_{itn}\gamma + \delta_t + u_{itn} \quad (1)$$

Here, y_{itn} measures the deviation from scheduled starting time (8.15 AM) in hospital site i (with $i = A, B$) in period t for patient n. Dummy variable $siteA_{itn}$ indicates whether patient i underwent surgery in period t in hospital site A. To indicate whether a surgery took place before or after OR charter introduction, we introduce the dummy variable $post_{itn}$. The coefficient of interest is β_1, which measures the difference in average delays between site A and site B after charter introduction took place. The binary variable $ward_{itn}$ indicates whether a head physician ward round took place or not and $pain_{itn}$ whether patient i required peripheral pain therapy. Additionally, δ_t reflects time fixed effects (half-year dummies) and the $1 \times k$ vector X_{itn} contains patient-specific information (age, gender, co-morbidities and surgery duration).

Additionally, we want to explore whether hospitals improve in aspects which have been explicitly addressed in the OR charter. We are especially interested in whether a reduction of tardiness for days where head physicians' ward rounds took place is detectable.

In Eq. (2), we employ a difference-in-difference-in-difference (DDD) strategy in order to assess whether site A improved in reducing delays on days with ward rounds compared to site B and estimated the following model:

$$y_{itn} = \beta_0 + \beta_1 post_{itn} * siteA_{itn} + \beta_2 siteA_{itn} + \beta_3 pain_{itn} + \beta_4 ward_{itn} + \beta_5 ward_{itn} * post_{itn} + \beta_6 ward_{itn} * SiteA + \beta_7 ward_{itn} * siteA_{itn} * post_{itn} + X_{itn}\gamma + \delta_t + u_{itn} \quad (2)$$

2.4 Results

Table 2.3 shows the results of the generalized difference-in-difference estimation approach (Eq. (1)) we employed to estimate the effect of OR charter introduction on FCODS. The coefficient of interest is β_1, which measures the average difference in deviations between treated site A and untreated site B before and after charter introduction. For the base model without covariates (1a), the estimated effect of OR introduction is small in magnitude (-0.93), which translates in an estimated decrease in average delays of one minute. Furthermore, the estimated effect of OR introduction on delays is not significant. As soon as we control for site A (1b), the estimated effect becomes highly significant and increases to a magnitude of -9.53, implying that the introduction of the OR charter is associated with a drop in average delays by nine and half minutes. Introducing patient characteristics and surgery time (1c) leads to a slight decrease in the magnitude of the estimated reduction (to 8.02 minutes). Here, the estimated effect of surgery duration on morning delays is highly significant and positively correlated with delays. Another minute of surgery time is associated with an average delay of 0.07 minutes in morning starts. Furthermore, both obesity and diabetes are associated with significant delays, where obesity leads to an estimated delay of 2.06 minutes and diabetes is associated with a delay of 1.71 minutes. Specification (1d) includes a dummy variable for head physicians' ward rounds. The estimated coefficient for head physicians' ward rounds is positive and highly significant which implies that on days with ward rounds, the first case of the day starts on average with delays of three and a half minutes. All other coefficients remain similar in magnitude in comparison to specification (1c). Finally, we introduce peripheral pain therapy as another covariate and have the fully specified model in (1e). Our estimates suggest that peripheral pain therapy is associated with a delay of 5.13 minutes and that this effect is again highly significant. The estimated treatment effect in the fully specified model (1e) equals -8.23 minutes, which means that average delays decrease by roughly eight minutes for site A

after charter introduction has taken place. For all specifications (1b)-(1e), coefficient estimates for β_2 are roughly the same in magnitude and highly significant. Regression results show that before OR charter has been introduced, FCODS started on average eight minutes later than in site B. The OR charter implementation was associated with a significant decline in delays and resulted in almost complete alignment in average delays between site A and site B.

Table 2.3: Effect of OR charter adoption on deviations from scheduled incision time – Difference-in-difference estimation results

	Dependent variable				
	Deviation from scheduled incision time				
	(1a)	(1b)	(1c)	(1d)	(1e)
SiteA × Post	-0.963	-9.532***	-8.024***	-7.973***	-8.234***
	(0.80)	(1.04)	(0.99)	(0.99)	(0.96)
SiteA		8.569***	8.081***	8.139***	8.000***
		(0.66)	(0.64)	(0.63)	(0.60)
Age65			0.315	0.270	0.020
			(0.54)	(0.54)	(0.52)
Female			1.169*	1.028*	0.739
			(0.49)	(0.49)	(0.47)
Obesity			2.062**	2.020**	1.630*
			(0.71)	(0.69)	(0.66)
Diabetes			1.708*	1.680*	1.891**
			(0.74)	(0.72)	(0.71)
Surgtime			0.071***	0.072***	0.068***
			(0.01)	(0.01)	(0.01)
Ward				3.482***	3.217***
				(0.66)	(0.65)
Pain					5.132***
					(0.56)
Constant	16.200***	12.333***	5.417***	4.749***	4.555***
	(0.73)	(0.67)	(0.97)	(0.96)	(0.93)
Half-year dummies	Yes	Yes	Yes	Yes	Yes
r2	0.021	0.133	0.234	0.252	0.300
N	1317	1317	1317	1317	1317

Robust standard errors are in parentheses. * p<0.05, ** p<0.01, *** p<0.001

To get a further understanding of how the OR charter actually has affected OR staffs' behavior, we analyze, whether observed reductions in delays are attributable to improvements in explicitly addressed aspects. The charter at hand stipulates that cases requiring peripheral pain therapies should not be scheduled as first case of the day and that head physicians ward rounds should not compromise punctual FCODS.

First, we assess the question whether Site A has reduced the number of surgeries requiring pain therapies as first case of the day after OR charter introduction. We conduct a t-test to compare average delays before and after the OR charter has been implemented in site A. The tested null hypothesis H_0 is that the mean share of first cases of the day, which require peripheral pain therapies, does not differ between the pre- and post-treatment period. We test the hypothesis for both hospital sites separately. For site B, we do not expect to reject the null hypothesis as this site is not affected by the OR charter introduction in site A. In contrast, we expect to reject the null hypothesis for site A as it is explicitly stated in the OR charter.

From Table 2.4 we see that no significant changes are detectable in the average share of peripheral pain therapies for both site A and site B. This indicates that site A does not follow the explicitly stated guideline of not scheduling cases requiring peripheral pain therapy as first case of the day.

To analyze whether OR charter introduction had an effect on delays for days with ward rounds, we implement a difference-in-difference-in-difference (DDD) approach. This requires to add the following three additional interaction terms $ward_{itn} * post_{itn}$, $ward_{itn} * SiteA_{itn}$ and $ward_{itn} * SiteA_{itn} * post_{itn}$ to our original estimation equation Eq. (1). Consequently, Eq. (2) describes our DDD estimation approach, where the coefficient β_7 associated with the triple interaction term $ward_{itn} * SiteA_{itn} * post_{itn}$ is of major interest. The estimated coefficient $\hat{\beta}_6$ measures the difference between differences in changes in average morning delays between site A and site B for days with ward rounds and days without ward rounds[8].

[8] The DDD-estimator is equal to $\hat{\beta}_7 = \left[\left(\bar{Y}_{Post,SiteA,Ward} - \bar{Y}_{Pre,SiteA,Ward}\right) - \left(\left(\bar{Y}_{Post,SiteB,Ward} - \bar{Y}_{Pre,SiteB,Ward}\right)\right)\right] - \left[\left[\left(\bar{Y}_{Post,SiteA,No\ ward} - \bar{Y}_{Pre,SiteA,No\ ward}\right) - \left(\left(\bar{Y}_{Post,SiteB,No\ ward} - \bar{Y}_{Pre,SiteB,No\ ward}\right)\right)\right]\right]$

Table 2.4: T-test for changes in the average share of first cases of the day requiring peripheral pain therapies (for site A and site B respectively)

Site A					
	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
Pre-treatment	384	0.271	0.023	0.445	0.226 0.315
Post-treatment	198	0.273	0.032	0.4446	0.210 0.335
Combined	582	0.271	0.018	0.445	0.235 0.308
Diff		-0.002	0.039		-0.078 0.075

diff=mean(0)-mean(1) t=-0.0486

H_0: diff=0 degrees of freedom=580

H_a: diff<0 H_a: diff≠ 0 H_a: diff>0

Pr(T<t)=0.4806 Pr(|T|>|t|)=0.9613 Pr(T>t)=0.5194

Site B					
	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
Pre-treatment	418	0.242	0.021	0.429	0.200 0.283
Post-treatment	317	0.202	0.023	0.402	0.157 0.246
Combined	735	0.224	0.015	0.418	0.194 0.255
Diff		0.040	0.031		-0.021 0.101

Diff=mean(0)-mean(1) t=1.2783

H_0:diff=0 degrees of freedom=733

H_a: diff<0 H_a: diff≠ 0 H_a: diff>0

Pr(T<t)=0.8992 Pr(|T|>|t|)=0.2015 Pr(T>t)=0.1008

As shown in Table 2.5 the difference-in-difference-in-difference model confirms our key results derived from the difference-in-difference specification in Eq. (1). The estimated coefficient $\hat{\beta}_7$ for the triple interaction term *Ward × SiteA × Post* equals -6.91, which means that after OR charter introduction site A exhibits an additional reduction in average delays of almost seven minutes on days with head physician ward rounds compared to days without head physician ward rounds. These results suggest that OR personnel in site A undertook efforts to reduce existing conflicts between ward rounds and first case punctuality and thus act in compliance with OR charter guidelines.

Thus, we have two major findings with respect to the impact of OR charter implementation on OR staffs' behavior. First, OR staff seems to adhere only partially to OR charter guidelines, as we found no evidence for a reduction in the number of first cases requiring peripheral pain therapy, whereas we found substantial reductions in delays for days with head physician ward rounds. Second, our results suggest that OR staff implements alternative measures - not specified in the charter - in order to reduce first case tardiness. As Table 2.5 shows, improvements are not only attributable to reduced delays on days with ward rounds. Instead, site A is able to achieve a reduction in average delays of around seven minutes on all working days (reflected in the estimate coefficient for the interaction term *SiteA × Post*). This implies that OR staff uses strategies other than those explicitly stated in the OR charter to fight delays. These findings highlight the flexibility of this management tool as it seems to foster the usage of knowledge already existing among OR personnel.

Table 2.5: Effect of OR charter adoption on deviations from scheduled incision time on days with head physicians' ward rounds

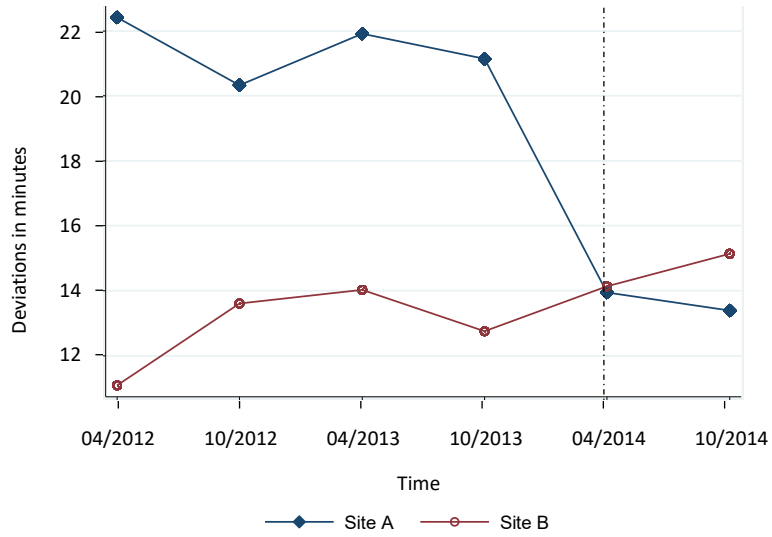
	Dependent variable
	Deviation from scheduled incision time
SiteA × Post	-7.130***
	(1.03)
SiteA	7.233***
	(0.65)
Age65	0.051
	(0.52)
Female	0.756
	(0.47)
Obesity	1.626*
	(0.65)
Diabetes	1.752*
	(0.71)
Surgtime	0.067***
	(0.01)
Pain	5.152***
	(0.56)
Ward	2.504*
	(1.06)
Ward × Post	-0.462
	(1.85)
Ward × SiteA	4.046*
	(1.57)
Ward × SiteA × Post	-6.907**
	(2.64)
Constant	4.866***
	(0.94)
Half-year dummies	Yes
r2	0.309
N	1317

Robust standard errors are in parentheses. * p<0.05, ** p<0.01, *** p<0.001.

2.5 Robustness

In this section, we assess the robustness of our results. First, we restrict our analysis on a subset of homogenous procedures. We do so because orthopedic procedures vary substantially with respect to preoperative processes (e.g. anesthesia induction, patient positioning). We cannot observe these preoperative processes in our data set and are therefore not able to control for them, although they are likely to affect FCODS. Focusing only on one particular subset of procedures allows to ensure homogeneity in preoperative processes thereby helping to circumvent potential confounders. We concentrate on hip related interventions[9], because these procedures account for a large portion of surgical interventions in both sites, thereby representing the largest overlapping set. Visual inspection of pre-treatment deviations for hip surgeries reveals a parallel development in both sites (see Figure 2.4). Again, we only include non-same day surgeries and the resulting subsample encompasses 476 observations.

Figure 2.4: Time trends for average deviations from targeted incision time from April 2012 to February 2015 measured in half-year time intervals (where the sixth (last) interval encompasses only five months due to the data structure) (n=476)



[9] We included patients coded as DRG I 47B. Furthermore, we focused on one procedure (implantation of an endo prosthesis (OPS code: 5-820)) to ensure comparability in perioperative processes.

We estimate the following estimation equation:[10]

$$y_{itn} = \beta_0 + \beta_1 post_{itn} * siteA_{itn} + \beta_2 siteA_{itn} + \beta_3 ward_{itn} + X_{itn}\gamma + \delta_t + u_{itn} \quad (3)$$

Table 2.6 shows the results for estimating Eq. (3) with the hip surgery subsample. The estimated effect of OR charter adoption is highly significant and with a reduction of roughly eight minutes similar in magnitude to the results derived in the main analysis encompassing all orthopedic non-same day surgeries. This result underlines the robustness of our results, as the hip surgery subsample does not suffer from unobserved confounders provoked by systematic differences in perioperative processes. Thus, we can conclude that unobserved surgery-type and process-related confounders do not seem to bias our analysis systematically.

[10] We exclude the covariate *pain* as our data set records no hip surgeries where a peripheral pain therapy has been conducted.

Table 2.6: Effect of OR charter adoption on deviations from scheduled incision time for hip surgeries subsample

	Dependent variable			
	Deviation from scheduled incision time			
	(1a)	(1b)	(1c)	(1d)
SiteA × Post	-8.787***	-8.492***	-8.573***	-7.735***
	(-1.58)	(-1.57)	(-1.55)	(-1.49)
SiteA	9.683***	9.505***	8.372***	8.288***
	(-0.99)	(-1.01)	(-1.15)	(-1.10)
Age65		0.408	0.407	0.239
		(-0.75)	(-0.75)	(-0.73)
Female		-0.222	-0.069	-0.127
		(-0.71)	(-0.71)	(-0.69)
Obes		0.822	0.321	0.41
		(-1.03)	(-1.00)	(-0.97)
Diab		2.055	1.975	1.868
		(-1.07)	(-1.04)	(-0.99)
Surgtime			0.051*	0.049*
			(-0.02)	(-0.02)
Ward				4.378***
				(-0.96)
Constant	10.576***	9.912***	6.471***	5.738**
	(-0.8)	(-1.02)	(-1.87)	(-1.85)
Half year dummies	Yes	Yes	Yes	Yes
r2	0.211	0.222	0.234	0.277
N	476	476	476	476

Robust standard errors are in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Additionally, we want to assess whether our findings are robust with respect to the choice of the time window specification for first case starts. In order to minimize the impact of deviations provoked by factors beyond the control of OR staff (technical problems etc.), we only considered cases in our main analysis, which started within the time window between 7:50 AM and 9:00 AM. To explore the robustness of our results with respect to the choice of the time window, we re-run the difference-in-difference estimation equation Eq. (1) with four more conservative (shorter) time windows. The results in Table 2.7 indicate that the estimated effect of OR charter introduction is robust with respect to alternative specifications as the estimated treatment effect remains highly significant for all considered specifications. However, the magnitude of the estimated coefficient for the interaction term *SiteA × Post* varies. The estimated effect is the largest for the broadest time window (≥-25 & ≤45), for which we have an estimated reduction in delays of around eight minutes. All other time window specifications are narrower and allow for smaller maximum delays. The shorter the considered maximum delay the smaller the estimated treatment effect. If we only consider cases with a maximum delay of 30 minutes, the estimated treatment effect decreases to around six minutes. This seems plausible as the smaller the specified threshold for considered delays becomes, the cases with large delays from targeted incision time are removed from the sample. Additionally, this finding indicates that observed improvements in delay reductions were not only driven by the prevention of extremely large delays but instead by improving overall punctuality.

Table 2.7: Alternative specifications for considered time windows (the grey column represents the time window specification used in the main analysis)

	Dependent variable				
	Deviation from scheduled incision time				
	$\geq -5\& \leq 30$	$\geq -10\& \leq 40$	$\geq -25\& \leq 45$	$\geq -15\& \leq 35$	$\geq -15\& \leq 30$
SiteA × Post	-6.335***	-8.079***	-8.234***	-7.171***	-6.271***
	(0.8)	(0.94)	(0.96)	(0.87)	(0.80)
SiteA	6.447***	7.766***	8.000***	7.112***	6.442***
	(0.49)	(0.58)	(0.60)	(0.54)	(0.49)
age65	0.116	0.348	0.020	0.108	0.087
	(0.43)	(0.50)	(0.52)	(0.46)	(0.43)
Female	0.618	0.469	0.739	0.731+	0.660+
	(0.39)	(0.46)	(0.47)	(0.42)	(0.39)
Obesity	1.238*	1.536*	1.630*	1.387*	1.247*
	(0.55)	(0.63)	(0.66)	(0.60)	(0.55)
Diabetes	1.565**	1.149+	1.891**	1.536*	1.452*
	(0.58)	(0.66)	(0.71)	(0.64)	(0.58)
Surgtime	0.051***	0.064***	0.068***	0.057***	0.052***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ward	2.162***	2.923***	3.217***	2.398***	2.178***
	(0.57)	(0.63)	(0.65)	(0.61)	(0.57)
Pain	4.413***	5.024***	5.132***	4.784***	4.419***
	(0.47)	(0.54)	(0.56)	(0.51)	(0.47)
Constant	5.399***	4.842***	4.555***	5.221***	5.340***
	(0.77)	(0.91)	(0.93)	(0.85)	(0.78)
Half-year dummies	Yes	Yes	Yes	Yes	Yes
r2	0.277	0.289	0.300	0.279	0.278
N	1,193	1,300	1,317	1,255	1,194

Robust standard errors are in parentheses. * p<0.05, ** p<0.01, *** p<0.001

2.6 Conclusion

Our analysis of the OR charter introduction in an orthopedic department suggests that this tool is effective in reducing delays in elective non-same day surgeries. Estimation results indicate that OR charter introduction led to a decrease in average delays by around eight minutes on average. We exploited a quasi-experimental setting, where two hospitals sites belonging to the same hospital group introduced an identical OR charter with a time lag of eleven months. The validity of the estimated effect depends on the suitability of the control

group. Parallel pre-trends in average deviations from scheduled incision time for both sites indicate the suitability of our control group. Furthermore, both sites belong to the same hospital group implying that both hospital sites are subject to identical management decisions.

Our estimation results show that ward rounds and peripheral pain therapies were associated with substantial delays. Interestingly, OR management has explicitly addressed these two aspects in the OR charter guidelines, which indicates that OR personnel and management were already aware of weaknesses in perioperative processes when setting up the OR charter document. Regarding peripheral pain therapies, our analysis indicates that OR personnel refrained from adopting the guideline as the number of first case surgeries requiring peripheral pain therapy did not significantly decline after OR charter introduction. The picture is different with respect to delays on days with head physician ward rounds, for which we observe substantial reductions in morning delays. Furthermore, we also detect significant reductions in FCODS delays on days where no ward rounds took place. This indicates that OR personnel found alternative ways – which have not been explicitly stated in the OR charter - to reduce delays in first cases. This supports the evidence for the effectiveness of using general guidelines rather instead of formulating strict specifications. Our results indicate that knowledge with respect to process deficiencies already exists among the affected employees and that OR staff is able to develop effective strategies to fight delayed surgery starts autonomously.

These results underline the importance of using process knowledge existing among targeted employees when designing or revising guidelines. Furthermore, in a highly complex working environment with a limited level of standardization in working processes, it is important to formulate guidelines in a non-restrictive way in order to ensure that OR staff can flexibly react to unforeseen occurences. This also implies that OR staff needs to be integrated into the OR charter's development process to ensure that valuable knowledge about deficiencies and working processes are taken into account and to foster acceptance among employees.

Additionally, it would be fruitful to analyze whether combining an OR charter with other management tools such as quality reporting or financial incentives is a sensible option to realize possible synergies.

Finally, our analysis faces some limitations. First, we solely consider orthopedic interventions and therefore we cannot make any predictions regarding the effects of OR charter introduction in other specialties. As delays in FCODS might be different among specialties (e.g. due to differences in perioperative processes), the effectiveness and success of an OR charter document possibly varies. Using a representative sample of hospitals with various disciplines to gain further insights might be a promising approach. Finally, more evidence is needed on the complex relationship between reduced delays in FCODS and the resulting effects on costs in the OR and subsequent areas (Szczesny & Ernst, 2016), as determining real cost savings through improved FCODS is challenging (Scalea et al., 2014).

3 Impact of reimbursement change on physicians' screening behavior[11]

Abstract

In this paper, we analyze the effect of a reimbursement change from fee-for-service paid out-of-pocket to capitation per patient on health service provision behavior for a specific preventive service (i.e. amblyopia screening). Our approach exploits a natural experiment generated by the introduction of a selective contract of pediatric care in 2014. Our panel data from a large public German sickness fund covers routine and billing data from 2011-2017. We set up a theoretical model to derive testable predictions regarding physicians' screening behavior. We employ a difference-in-difference approach to analyze amblyopia-screening behavior for both participating and non-participating pediatricians. Empirical results are in line with our theoretical prediction as we find a significant increase in the screening rate for participating pediatricians compared to non-participating ones. This indicates that capitation does not necessarily lead to a decrease in service provision.

[11] This chapter is based on joint work at the University of Hohenheim with Tanja Wollensak, which is yet unpublished.

3.1 Introduction

Health economics literature has long had a strong interest in studying how over, under, and incorrect provision of care can be addressed by designing more intelligent reimbursement systems for physicians. Different reimbursement schemes for health service provision imply differing incentives for service provision. Prospective reimbursement such as capitation shifts financial risk towards the provider and thereby affects incentives for service provision and quality (Ellis & McGuire, 1993; Ma, 1994). In contrast, retrospective reimbursement schemes like fee-for-service (FFS) (paid for by health insurers) shift financial risk towards the health insurer (Ellis & McGuire, 1986). Overall, this literature suggests that FFS incentivizes providers to increase the number of services, whereas prospective payment schemes such as capitation, flat rates and capitated fees could potentially prevent overtreatment (Ellis & McGuire, 1986; Ellis & McGuire, 1993; Pauly, 1995; Henning-Schmidt et al., 2011). Generally, physicians have economic incentives to reduce the amount of services as reimbursement becomes increasingly more independent from service provision (Hennig-Schmidt & Wiesen, 2014). Unfortunately, this may cause the pendulum to swing in the other direction and result in an underprovision of health services and treatment quality (Henning-Schmidt et al., 2011).

In Germany, FFS or FFS-based models traditionally constituted the predominant way of reimbursement for physicians in the outpatient sector[12]. With the introduction of gate-keeping models in 2008[13], a shift from FFS to more prospective reimbursement models has taken place for several medical services. The main objectives of gate-keeping models were cost containment in service provision, improvements in service quality and the strengthening of general practitioners' coordination function (Klora et al., 2017). During the last two decades, several German sickness funds expanded gate-keeping models to other medical specialists and introduced selective contract schemes. A large German sickness fund operating in South-West Germany launched a non-mandatory selective contract targeting pediatricians exclusively in 2014. The declared objective of the selective contract for pediatric care (SCPC)

[12] Compensation for outpatient services is actually quite complex. Basically, physicians do not receive payments for services provided directly from the statutory health insurance, but from the Associations of Statutory Health Insurance Physicians (Kassenärztliche Vereinigungen (KV)). The health insurance fund pays a "total compensation" to the KV covering all necessary medical treatments of the insured. The KV then allocates the funds to the physicians mainly according to FFS.

[13] § 73b SGB V (German Social Security Code V).

was the introduction of a more needs-oriented health care provision for children and adolescents. In contrast to standard care, SCPC encompasses several prospective reimbursement elements with a fixed baseline fee per participating patient and additional contact fees (e.g. for chronically ill children). The reimbursement change was designed to streamline administrative requirements, e.g. by reducing documentation standards for capitated services. For the service providers this type of capitation per patient ensures a baseline income thereby reducing financial risks. Additionally, by decoupling reimbursement from the scope of service provision, capitation may limit incentives for service over-provision.

Another major objective of SCPC was to incentivize the provision of effective preventive measures. Negotiations of contractual content and terms took place bilaterally between representatives of the sickness fund and the representatives of resident pediatricians. Both parties agreed on the relevance of ophthalmological screening measures due to the importance of proper eyesight on early childhood development. In this context, amblyopia is one of the most common eye-disorders among young children (Lagrèze, 2010). Prevalence of amblyopia for children varies between 1% and 5% (Ganekal et al., 2013; Oscar et al., 2014; Fu et al., 2014; Faghihi et al., 2017; Aldebasi, 2015) with higher prevalence for adults (Faghihi et al., 2017). Prevalence depends on the population studied (e.g. age group, ethnicity) as well as the study design (e.g. definition of amblyopia[14]). Untreated amblyopia provokes visual impairments in adolescents and adults and increases the risk for visual impairment (Van Leeuwen et al., 2007). Research indicates that the success of amblyopia treatment crucially depends on early detection of risk factors because amblyopia treatment has higher success probabilities for younger children (Flynn et al., 1999; Flynn et al., 1998; Fronius et al. 2014; Williams et al., 2003; Williams et al., 2002; Elflein et al., 2015). Since no standardized system for ophthalmological preventive examinations for children currently exists in Germany, practitioners and pediatricians play a major role in detecting eye disorders in young patients (Elflein & Pitz, 2015).

Under standard care, i.e. for patients not participating in SCPC, the sickness fund does not cover the cost for this type of screening. Hence, patients' parents have to pay out-of-pocket

[14] There is no uniform definition for amblyopia (Lagrèze, 2010). This is one reason for differences of amblyopia prevalence across the world (Hashemi et al., 2018).

to receive the preventive service for their children. To increase the number of provided screenings, SCPC incentive design targets both pediatricians and patients simultaneously. With respect to pediatricians, reimbursement for screenings changes from FFS paid out-of-pocket (abbreviated with OOP-FFS in the remainder of our paper) to capitation. More specifically, participating pediatricians receive a capitation fee per quarter for each visiting patient who is eligible for a for-free screening, regardless of whether the screening actually takes place or not. For participating patients, the service is no longer associated with an out-of-pocket payment but instead the sickness fund fully covers the costs. Note that due to the simultaneous change in incentives, the SCPC deviates from the settings regularly studied in literature, in which changes affect either the health care service provider side or the patient side. To the best of our knowledge, we are the first to analyze such a framework in pediatric care. The introduction of SCPC constitutes a natural experiment. We develop a basic theoretical model to analyze the interplay between modified incentives for physicians and patients and derive predictions about physicians screening provision. In order to analyze the effects of the reimbursement change from OOP-FFS to capitation empirically, we compare amblyopia-screening provision behavior for pediatricians participating in SCPC and pediatricians who remain in the ordinary system of care. We do so by using a difference-in-difference approach. Our theoretical model predicts an increase in amblyopia screening rates for the SCPC compared to the ordinary scheme. Empirical results are in line with our theoretical prediction as we find a significant increase in the screening rate for participating pediatricians compared to non-participating ones.

3.2 Background

3.2.1 The Program

As participation in SCPC is voluntary for pediatricians and patients, SCPC and standard care do co-exist in routine pediatric patient care. Note that only if both the physician and the visiting patient participate in SCPC, treatment and reimbursement occurs under SCPC terms. If a participating pediatrician treats a non-enrollee, she is reimbursed via the standard care regime. Figure 3.1 gives a schematic overview of the reimbursement structure.

Figure 3.1: Classification selective scheme care and standard care

		Physician	
		Selective Scheme	Ordinary Scheme
Patient	Selective Scheme	**Selective Care**	Standard Care
	Ordinary Scheme	Standard Care	Standard Care

In the German statutory health care, most sickness funds do not cover the cost for amblyopia screenings. If a pediatrician undertakes a screening under standard care, parents have to pay around 20 € out of pocket for this service. Whether a child actually receives a screening depends therefore on parents' willingness-to-pay and ability-to-pay. As information asymmetries between provider and parents are obviously present, this willingness-to-pay heavily depends on physician's persuasiveness regarding the beneficial health effects of the amblyopia screening. Contrary to the previous regime, the SCPC framework now specifies that enrolled children between their second and third year of life are eligible for a free screening, meaning that parents face no costs when they opt for a screening. Thus, under SCPC parents' willingness-to-pay and ability-to-pay should no longer have an impact on screening demand.

For pediatricians, reimbursement for screening provision changes from OOP-FFS paid directly by patients to capitation per enrolled patient. Enrolled pediatricians, who verifiably own an appropriate screening device, receive a lump-sum payment paid by the sickness fund if an enrolled child in the eligible age group visits the pediatrician's practice. This lump-sum payment is billable once per quarter and does not depend on the actual provision of the screening. Thus, the pediatrician can receive the lump-sum-capitation fee four times a year per patient at maximum (given that this patient visits the pediatrician's practice in each quarter).

Since only children aged 2-3 years are eligible for the screening, the maximum time span of eligibility equals two years. It is important to note that the pediatricians' receipt of the capitation fee does not depend on actual service provision[15].

To avoid confusing, we want to make clear that our analysis only remotely relates to the traditional FFS vs. capitation literature. Most of the former literature (often tacitly) assumes that both reimbursement regimes are compared under existing third party payment by health insurers. If a third party covers (almost) 100% of the cost, it is obvious that physicians will maximize the number of healthcare services provided under FFS (e.g. Pauly, 1995). If the third party payer then switches to prospective payment or capitation, service provision will usually decrease.[16] Conversely, we study a setting where an OOP-FFS is replaced by a capitation system. The latter was included in the SCPC because the contracting parties agreed that the level of service provision under OOP-FFS was insufficient because of ability-to-pay and willingness-to-pay concerns that led to equity of access problems.

3.2.2 Theoretical Considerations

In order to make testable predictions on how pediatricians' provision behavior changes under SCPC, we first develop a theoretical model in order to reflect the peculiarities of our studied set-up. Our model relies on three central assumptions. First, we assume that pediatricians do not only care about monetary incentives when choosing their utility-maximizing level of service provision, but instead also about their patients' health benefit and well-being. The importance of altruistic motives in physicians' treatment decisions is widely accepted within the field of health economics (Arrow, 1963; Newhouse, 1970; Godager & Wiesen, 2013; Hennig-Schmidt et al., 2011). Second, we assume that health services, which patients have to purchase as out-of-pocket payments, require pediatricians to perform a costly persuasion

[15] The reimbursement scheme works as follows: Any participating pediatrician owning the testing device bills the fee if an eligible patient (i.e. patients is between 2-3 year old and enrolled in SCPC) visits pediatrician's practice. Amblyopia fee is billable once per quarter and it does not matter whether the pediatrician actually conduct the test or not. The latter aspect is what defines the capitation property because the physician could potentially pocket the fee without testing the child. This implies that a physician can invoice the fee eight times per patient at maximum (which is the case when an enrolled patient visits the pediatrician's practice in each quarter during second and third year of life).

[16] A well-known example is the Managed Care system in the U.S. in the 90s of the previous century. HMOs and PPOs limited the available provider pool to patients and switched to capitation. The result was an undersupply of services (at least in patient perception) that led to the so-called managed care backlash (Blendon et al., 1998; Duijmelinck & van de Ven W., 2016).

effort (Wiesemann et al., 2004; Turner et al., 2009)[17]. We do so, because physicians have to explain in detail the advantages of the screening and possible threats of undetected amblyopia on child's future health in order to convince parents to purchase the screening that is not covered by the statutory health care system. We interpret costs in terms of physician's opportunity costs i.e. as forgone revenue generated by conducting other services or treating more patients. In the remainder of this paper, we refer to these costs as persuasion costs. Third, we assume that the physician faces costs when conducting the screening. These costs are associated with the actual service provided (e.g. interpretation of test results, diagnostics, writing referrals to ophthalmologists etc.).

As a first step, we derive the number of screenings pediatricians conduct in the case that only the standard scheme is in place, i.e. in the case that screenings are only available as out-of-pocket services with price p. We assume that patients' parents[18] differ with respect to their willingness-to-pay to conduct an amblyopia screening. patients' parents type is denoted by parameter θ[19]. We assume that θ is continuously and uniformly distributed over the interval $[0,1]$ with density function $g(\theta) = 1$ and cumulative distribution function $G(\theta) = \theta$. According to our assumptions, a pediatrician faces persuasion cost $e(\theta)$ associated with the effort necessary to persuade a parent with type θ to purchase a screening as an out-of-pocket service. Persuasion cost arise due to effort to raise awareness about amblyopia and explain possible consequences if amblyopia remains untreated. We assume that $e(0) = 0$, $e'(\theta) > 0, e''(\theta) > 0$ and $e'''(\theta) < 0$ over the support $(0,1]$, which implies that persuasion cost are strictly increasing and convex in parent type θ and that a parent characterized by type $\theta = 0$ requires no persuation cost at all to purchase the screening.

[17] Wiesemann et al. (2004) as well as Turner et al. (2009) find that willingness-to-pay does not only depend on net income but instead also on patients' health literacy. To build up patients' health literacy, the physician has to educate patients about health related issues. As counseling requires time, the physician faces opportunity costs and a decision problem on how to allocate time.

[18] As the service provision in question targets infants, patient's parents decide whether their child receives a screening. Thus, we will always refer to patients parents' type when analyzing the decision whether to purchase a screening under standard care.

[19] In the theoretical framework at hand, a patient with type $\theta = 0$ exhibits the maximum willingness-to-pay, i.e. the pediatrician faces no consultation costs at all in order to convince the patient to purchase the screening. An increasing θ represents a decreasing willingness-to-pay. Furthermore, we abstract from ability-to-pay because we assume in the present framework that all parents have sufficient financial resources to afford the screening.

Furthermore, we assume that not only the monetary incentive p (the amount of money the pediatrician receives per out-of-pocket screening) increases pediatricians' utility, but also the knowledge that conducting the screening contributes to the child's health development and welfare. To reflect the latter, we introduce $a > 0$ which represents the utility gain provoked by childrens' additional health benefit due to the screening. Additionally, the pediatrician faces costs of service provision, which we denote with c. We assume constant costs per screening that are strictly positive (i.e. $c > 0$). Furthermore, we assume that pediatricians are not able to observe the patients' parents' type θ but instead only have knowledge about the distribution of type $g(\theta)$. Thus, the pediatrician sets ex-ante a utility maximizing consultation effort $e^{opt}(\hat{\theta})$ such that the marginal benefit from another unit of persuasion effort equals its costs. For all patient types' $\theta < \hat{\theta}$ the pediatrician faces marginal costs of $e'(\theta) < e^{opt}(\hat{\theta})$. This implies that the physician conducts the exact amount of persuasion effort necessary to convince the respective parent. In contrast, for all parent types $\theta \geq \hat{\theta}$ the pediatrician faces marginal costs of $e(\hat{\theta})$.

Thus, the pediatrician's overall persuasion costs e^{opt} connected with optimal consultation effort level $\hat{\theta}$ will be equal to:

$$e^{opt}(\hat{\theta}) = \int_0^{\hat{\theta}} e'(\theta)d\theta + \int_{\hat{\theta}}^1 e'(\hat{\theta})d\theta = e(\hat{\theta}) + (1 - \hat{\theta}) * e'(\hat{\theta}) \quad (1)$$

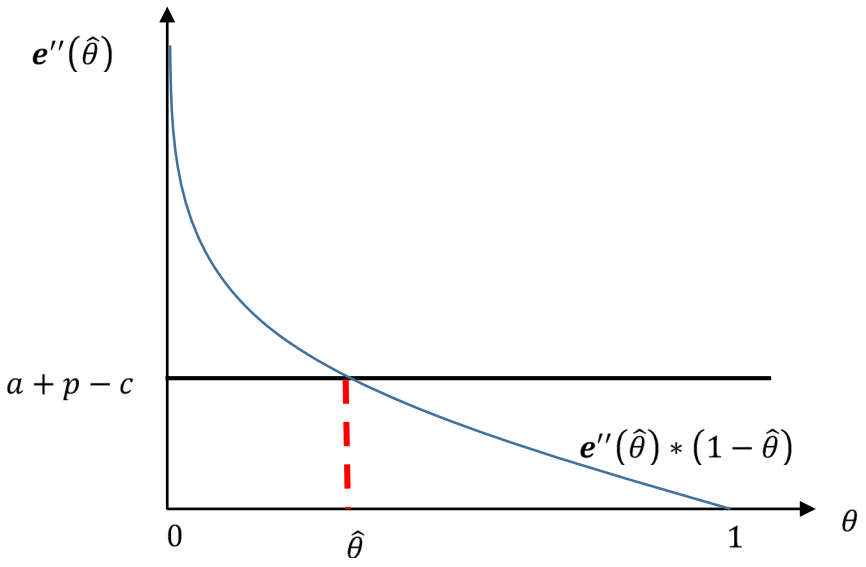
The utility-maximizing physician will choose $e^{opt}(\hat{\theta})$ in such a way that marginal utility from exercising another unit of persuasion effort equals its marginal cost. According to our assumptions, net marginal gain (without the consideration of convincing costs) is a constant and equals $(a + p - c > 0)$. Thus, the sum of benefit derived from patient well-being and out-of-pocket payment is strictly larger than the cost of conducting a screening. We consider this an appropriate assumption for the setting at hand, because amblyopia screenings are associated with minor costs to the pediatrician (Matta et al., 2008; Singman et al., 2013).

Taking the first derivative of (1) we get marginal convincing costs equal to:

$$e^{opt\prime}(\hat{\theta}) = e'(\hat{\theta}) + (-1) * e'(\hat{\theta}) + (1 - \hat{\theta}) * e''(\hat{\theta}) = (1 - \hat{\theta}) * e''(\hat{\theta}) \quad (2)$$

The utility maximizing physician will choose $e^{opt}(\hat{\theta})$ such that marginal utility and marginal cost from conducting another screening becomes equal (see Figure 3.3):

Figure 3.2: Utility maximizing effort



Assuming convex effort costs $e(\theta) = \theta^{\gamma}$ with $1 < \gamma < 2$, $e''(\theta)$ is strictly decreasing over the domain $(0,1]$ implying there is a unique interior solution satisfying $(a+p-c) = (1-\hat{\theta}) * e''(\hat{\theta})$. Thus, the number of conducted screenings under standard care will be equal to $G(\hat{\theta}) = \hat{\theta}$ with $0 < \hat{\theta} < 1$ and $a+p-c > 0$.

Now, we turn to the case when SCPC is introduced and the lump-sum scheme applies for all physician-patient contacts where both, physician and patient, participate in SCPC. We assume that under the selective scheme $e(\theta) = 0$ for all $\theta \in [0,1]$, i.e. as soon as screenings are covered by the statutory health insurance, physicians face no persuasion costs at all. We do so, because the screening is not associated with any discomfort for the screened individual, very quick and no negative side effects whatsoever have been reported to date (Moghaddam et al., 2012; Yan et al., 2015). Thus, we have two possible outcomes for the physicians' optimization problem, where the outcome depends on the relation between the degree of physicians' utility derived from patients' health benefit a and cost per screening c.[20] Physicians will conduct screenings as long as the marginal benefit from performing a screening

[20] If both the pediatrician and the patient are enrolled in SCPC, the price per screening p is no longer relevant for pediatrician's decision on optimal screening provision, as the pediatrician does not receive a reimbursement per screening but instead receives a contact-dependent fee regardless of whether screening takes place or not.

is larger or equal than its marginal cost, i.e. as long as $a \geq c$. In this case, all patients who opt into the selective scheme will receive a screening. On the contrary, if $a < c$, none of the selective scheme patients get a screening.

Now, we assess the change in the number of amblyopia screenings after SCPC introduction, i.e. in the case where we have a parallel system of standard and selective contract schemes.

Furthermore, we have to make an assumption with respect to patients' parents probability to enter SCPC depending on their type realization θ. We assume that each type $\theta \in [0,1]$ has a strictly positive probability $p(\theta) > 0$ of entering the selective contract scheme. We make no further restrictions with respect to the probability density function, thereby allowing for possible non-observable self-selection effects among patients' parents, i.e. the possibility that patients' parents willingness-to-pay and their respective probability for opting into the scheme are in some way correlated. In the following analysis, we consider the most restrictive case, i.e. we assume the entry probability is equal to $p(\theta) = \varepsilon$ for all $\theta \in [0,1]$ and that $\varepsilon > 0$ is arbitrarily small.

We start with analyzing the case where $a \geq c$ holds: As pediatricians' utility-maximizing condition for choosing e^{opt} is identical before and after the introduction of SCPC, all patients with type $\theta \leq \hat{\theta}$ who are still in the standard scheme will receive the screening. Additionally, all patients who have opted into SCPC receive a screening regardless of their type θ. Thus, we have that for the parallel scheme all patients with type $\theta \leq \hat{\theta}$ receive a screening. Then, the expected number of screenings for types $\theta \in (\hat{\theta}, 1]$ is equal to

$$\int_{\hat{\theta}}^{1} g(\theta) * \varepsilon \, d\theta = [\varepsilon * G(\theta)]_{\hat{\theta}}^{1} = \varepsilon * \left(1 - \hat{\theta}\right) \quad (3)$$

Thus, under the parallel system $\hat{\theta} + \varepsilon * \left(1 - \hat{\theta}\right)$ patients receive a screening. This is strictly larger compared to the number of screenings under standard care (which is equal to $\hat{\theta}$) given that $0 \leq \hat{\theta} < 1$. Thus, we know that given not all patients have received a screening under standard care before SCPC introduction, i.e. $\hat{\theta} < 1$, the number of amblyopia screenings strictly increases after SCPC introduction.

In contrast, for $a < c$ we will have that pediatricians do not conduct any screenings for SCPC patients. Again, pediatricians' utility-maximizing condition with respect to e^{opt} does not change. For patients with type $\theta \leq \hat{\theta}$ only those, who have not opted in the SCPC, receive the screening and for patients with type $\theta > \hat{\theta}$, no one will receive the screening, regardless of whether individuals are in standard care or in SCPC. Thus, the expected number of screenings is equal to

$$\int_0^{\hat{\theta}} (1-\varepsilon) g(\theta)\, d\theta = (1-\varepsilon)[G(\theta)]_0^{\hat{\theta}} = (1-\varepsilon)\left(\hat{\theta} - 0\right) = (1-\varepsilon) * \hat{\theta} \quad (4)$$

This implies that under the parallel system $(1-\varepsilon) * \hat{\theta}$ patients receive a screening. Therefore, the number of screenings under the parallel scheme is strictly smaller than under standard care (which equals $\hat{\theta}$).

Under the relatively weak assumption that all types have an arbitrarily small probability for entering the scheme, the model's main implications are, that if pediatricians exhibit a relatively low degree of altruism (i.e. $a < c$), the total number of screenings provided is strictly smaller under the parallel system of SCPC and standard care than under standard care only. In contrast, given that altruism is sufficiently large, i.e. $a \geq c$, our model predicts that the total number of screenings will increase after SCPC introduction.

In the present context, we assume that $a \geq c$ is likely to hold as experimental findings suggest that altruism is particularly prevalent among pediatricians (Hojat et al., 2002; Li, 2018) and marginal cost for amblyopia screenings is relatively small (Matta et al., 2008). Thus, our model predicts that assuming pediatricians are sufficiently concerned about patients' well-being and the costs associated with screening provision are relatively small, reimbursement change from OOP-FFS to capitation will induce an increase in service provision. Therefore, we test the following hypothesis:

H_0: *Physicians responded to the introduction of a lump sum payment under SCPC with an increase in the number of amblyopia screening tests.*

3.3 Data and Methodology

To test our theoretical predictions empirically, we use a comprehensive data set, which contains information on pediatricians' insurance billing records, diagnoses and characteristics of both pediatricians and patients provided by a large public German sickness fund. The data set covers the period between 2011 and 2017. Our approach exploits a natural experiment arising from the introduction of SCPC in 2014. The program participation is voluntary for pediatricians and patients and since its introduction pediatricians enrolled successively over time. The panel structure of the data allows the implementation of a two-way fixed effects estimation approach, which allows controlling for time-invariant unobservable confounders and time effects.

We use the group of pediatricians who abstain from entering SCPC throughout the complete observation period and own a screening device as control for general time-trends affecting screening provision. In contrast, the treatment group comprises all physicians who opt into the selective contract scheme during the observation period and own a screening device. SCPC participation is voluntary implying that pediatricians have to opt actively into the scheme. This means we run into a potential problem of self-selection bias as assignment to the treatment and the control group is not random. Consequently, complications with respect to possible causal interpretations of findings might arise as observed changes in screening provision might not be attributable to changes provoked by SCPC participation but instead due to the fact that propensity of participating in SCPC is possibly correlated with the pediatricians' inclination of providing screenings. We will examine this concern more closely later in this section and argue why we believe that potential selection bias poses no substantial threat to our analysis.

Our unique data set contains information on diagnoses (ICD-10-GM codes) and characteristics of both physicians and patients. The data comprises information on patients' characteristics like age, nationality, place of residence (ZIP-code), regional structure, morbidity-adjusted payments and patient's and parental enrollment status in SCPC. With regard to pediatricians, we have information on SCPC enrollment status and physician's age. Finally, the data covers billing data submitted by pediatricians. For our analysis, we aggregate data on physician-level on a quarterly basis. The data exhibits panel structure as we have observations for the same physicians in several points of time.

In order to make meaningful comparisons between the screening behavior of pediatricians who enroll in SCPC and receive the capitation payment and pediatricians who remain in standard care, we have to assign pediatricians properly to the control or treatment group. Pediatricians qualify for the treatment group, if they a) charged the capitation fee at least once (implying they were SCPC enrollees and owned a suitable screening device) and b) had already owned a screening device before they entered SCPC. Requirements for assigning a pediatrician to the control group are a) that she remained in the ordinary scheme throughout the complete observation period and b) owned a suitable device before SCPC introduction in 2014. Here, we assume that given pediatricians invest in buying a costly screening device, they will keep it throughout the observation period. When assigning pediatricians to respective groups we run into two problems. First, pediatricians neither document screenings in the ordinary nor in SCPC scheme.[21] As our objective is the analysis of reimbursement-induced changes on pediatricians' screening rate, we have to identify an appropriate proxy for the number of conducted screenings. We use the number of amblyopia and amblyopia-related diagnoses following the simple logic that a higher number of screenings will imply a higher number of diagnosed cases. Furthermore, to avoid double counts we only take into account initial diagnoses made by a pediatrician. In the case of multiple relevant diagnoses for one child, we only count one as our focus lies solely on the detection of amblyopia and related risk factors.

[21] Under standard care, patients reimburse pediatricians directly via an out-of-pocket payment, which implies that screenings do not appear in pediatricians' insurance billing records. Under selective care, pediatricians receive an upfront payment per patient as soon as the ownership of a screening device has been proven and are not required to document conducted screenings.

Table 3.1 enlists all diagnoses we consider in the following analyses. Here, manufacturers' product information served as a basis to identify disorders, which are detectable with the screening device[22].

Table 3.1: Amblyopia and amblyopia-related diagnoses

ICD-10 code	**Diagnosis**
H53.0	*Amblyopia*
H52.3	*Anisometropia (two eyes have unequal refractive power)*
H52.2	*Astigmatism (refractive error in which eye does not focus evenly on retina)*
H52.1	*Myopia (short-sightness)*
H52.0	*Hyperopia (far-sightness)*
H50.4	*Strabismus (microstrabismus)*
H57.0	*Mydriasis (dilation of the pupil)*
Q13.2	*Anisocoria (unequal size of eyes' pupil)*
H18.6	*Keratoconus (progressive thinning of the cornea)*
H18.7	*Other deformities of the cornea*
Q13.3	*Congenital corneal opacity*
Q13.4	*Other congenital deformities of the cornea*
Q12.0	*Cataracta congenital*

Secondly, we have incomplete information on whether pediatricians own a suitable screening device. In order to provide screenings for pre-verbal children, pediatricians need a screening device which does not require active involvement of the child. We have to find a valid strategy to identify ownership status of such a device. Reliable information is only available for pediatricians who participate in SCPC, as proof of ownership is a prerequisite for charging the amblyopia screening capitation fee. Thus, we have to formulate criteria in order to identify non-participating pediatricians owning a screening device. As we only consider screening

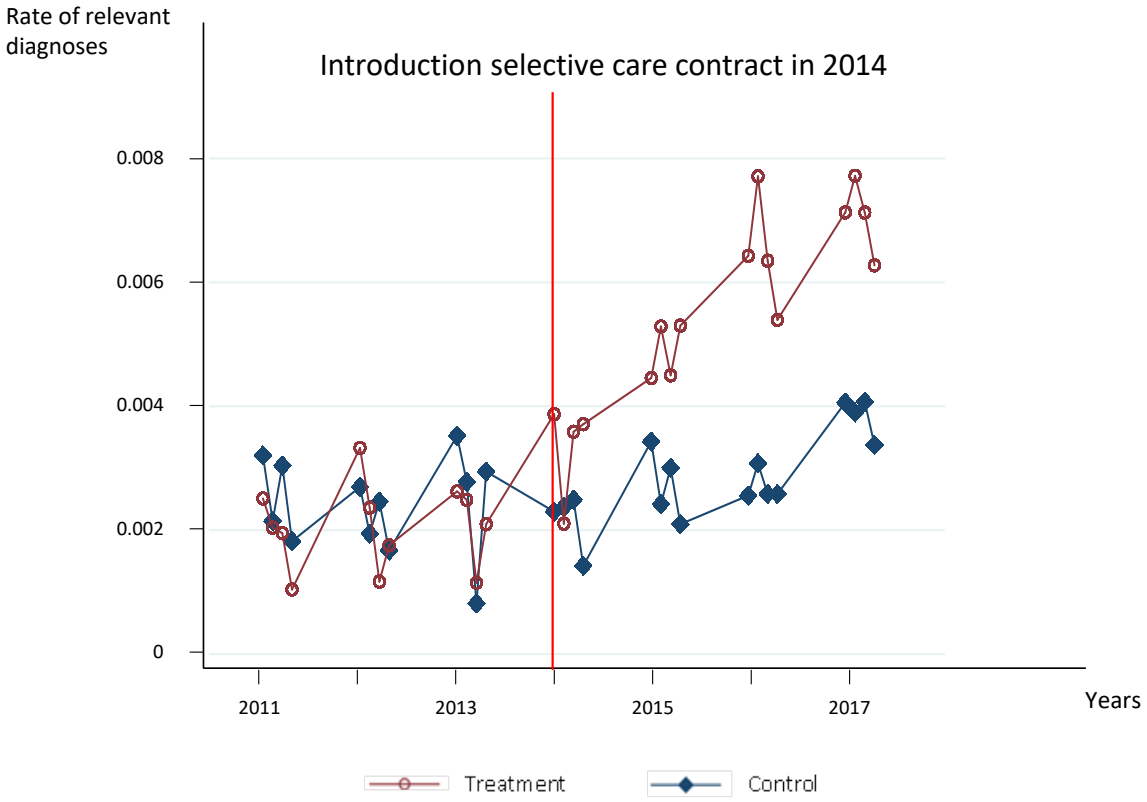
[22] See https://plusoptix.com/vision-screener/landing-google-eu?pk_campaign=GoogleAds_EU&gclid=Cj0KCQjwt-6LBhDlARIsAIPRQcKM3vqE4mjbjyDP1tgIXUrzph1H_e_l_RfOAG_sEkls49DfKvtUEHYaAmKVEALw_wcB

devices which enable the physician to conduct screenings contactless and without child's active cooperation, we filter for amblyopia or amblyopia-related diagnoses in sufficiently young children (≤ 3 years). Furthermore, we only include diagnoses, which require the usage of a screening device implying that health care providers cannot do them by just using naked eye diagnostics[23]. Our treatment group comprises 179 and the control group 156 pediatricians. This constitutes one of the largest samples to study outpatient-care provision under a specific selective contract in Germany so far.

To analyze the development of average screening rates for both treatment and control group over time, we take quarterly averages over diagnoses rates for both groups respectively. The quarterly diagnose rate is defined as the ratio between the number of relevant diagnoses and the total number of patients in the relevant age group (two to three years) who visited pediatrician's practice in the respective quarter. Figure 3.3 shows the development of diagnoses rates over time. Before the introduction of SCPC in 2014, the development of diagnoses rates over time for treatment and control group were relatively similar trend- as well as level-wise. After SCPC introduction we can see that control group's average level of diagnoses rates did not change notably, but that diagnoses rates for the treatment group increases substantially.

[23] For this reason we only considered microstrabismus and not strabismus as the latter is in many cases noticeable with naked eye diagnostics not requiring additional diagnostic tools.

Figure 3.3: Trends in physicians' quarterly average ratio between number of relevant diagnoses and number of patients aged two to three years over the observation period 2011-2017 - separately presented for treatment (red) and control group (blue).



3.3.1 Outcome Variable

The outcome variable of interest is the number of quarterly diagnoses per pediatrician. We adjust the number of diagnoses to the number of diagnoses the pediatrician possibly could have done. This ensures comparability of our outcome variable among pediatricians. The number of possible diagnoses equals the number of patients, who are in the relevant age group and who visited the pediatrician in the given quarter of interest. Thus, we deal with a fraction as we measure the occurrence of amblyopia and amblyopia related diagnoses per hundred patients. Figure 3.4 shows the distribution of "*fraction of diagnoses"* for the complete data set, the treatment and control group. For all three groups, the distribution is skewed to the right. To account for the skewed distribution, we will use a negative-binomial fixed-effects model to estimate the effect of SCPC on the average number of quarterly diagnoses per pediatrician. Technically, to reflect that our outcome variable describes a rate, we will introduce the exposure variable "*number patients"* in our estimation equations, measuring the number of possible diagnoses a pediatrician can possibly make in one quarter.

Figure 3.4: Distribution of relevant diagnoses per physician on quarterly basis for a) complete sample, b) control and c) treatment group.

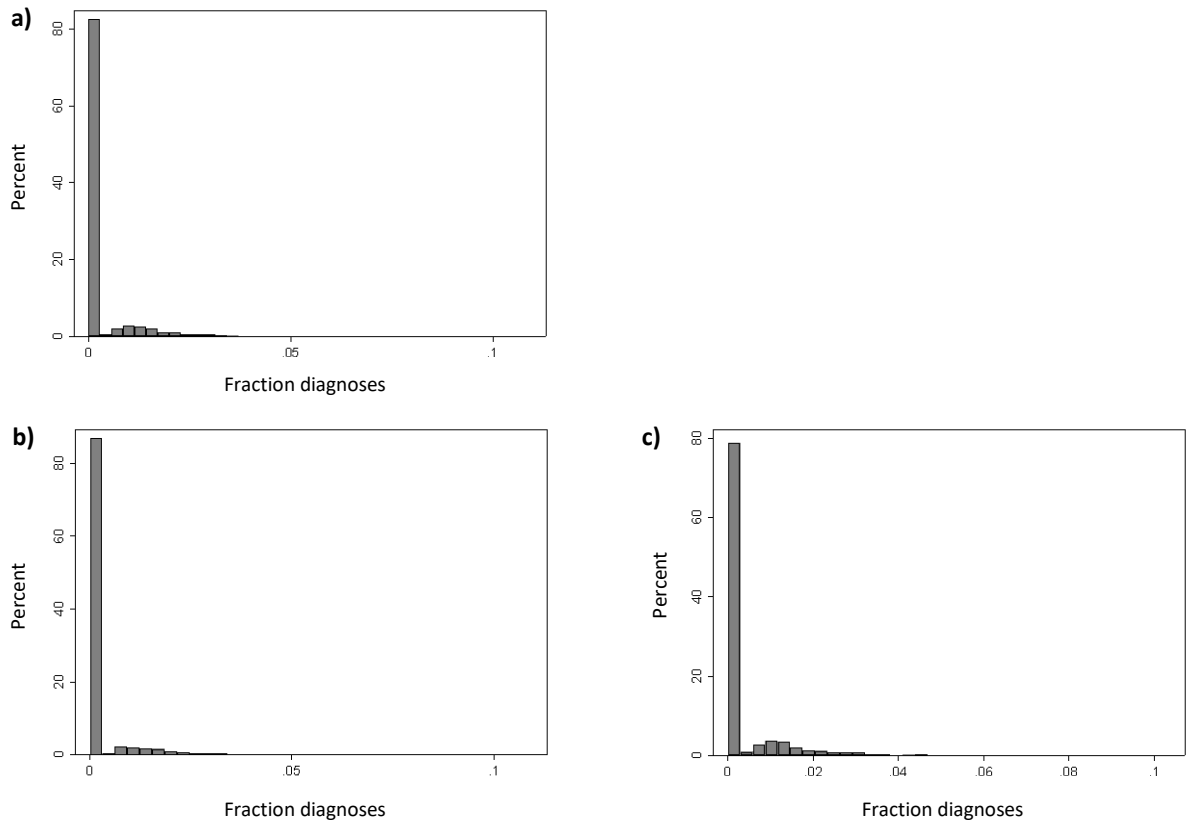


Table 3.2 shows the average number of quarterly diagnoses per one hundred patients for the treatment and the control group prior and post reimbursement change. As we can see, the average fraction of diagnoses before introduction of SCPC (2011-2013) is almost identical for the treatment and the control group (0.0020 vs. 0.0024). After SCPC introduction (2014-2017), the number of diagnoses per one hundred patients has more than doubled for pediatricians in the treatment group (0.0054 vs. 0.0020). In contrast, for pediatricians in the control group, fraction of diagnoses stagnates (0.0024 vs. 0.0028).

Table 3.2: Average "fraction of diagnoses" for treatment and control group pre and post SCPC introduction

	Treatment group			**Control group**		
Variable	Mean (SD)	Min.	Max.	Mean (SD)	Min.	Max.
Period	Jan 2011 to Dec 2013					
Obs.	2,060			1,840		
Fraction of diagnoses	0.0020 (0.0065)	0	0.0805	0.0024 (0.0083)	0	0.1111
Period	Jan 2014 to Dec 2017					
Obs.	2,784			2,233		
Fraction of Diagnoses	0.0054 (0.0119)	0	0.1053	0.0028 (0.0093)	0	0.1071

3.3.2 Control Variables

As covariates we use patient characteristics like nationality, patients' socioeconomic status (based on ZIP-code), regional structure[24] (based on ZIP-code), morbidity-adjusted payments[25], patient's and parental enrollment status in SCPC. We use ZIP-codes to control for patients' (i.e. patients' parents) socioeconomic status by implementing the German Index of Socioeconomic Deprivation. The index varies between 3 (= high socioeconomic status) and 21 (= low socioeconomic status)). Additionally, we add physicians' age[26] as another control variable. For our analysis, we aggregate data on physician-level on a quarterly basis. This means we average patient information over physicians, i.e. we calculate averages for the physicians' respective patient bases.

[24] The variation range is between 1 (= rural area) and 6 (= urban area).

[25] Sickness funds receive additional payments for cost-intensive chronic diseases. We add this covariate, which measures these additional payments to control for patients' morbidity. We do not use the Charlson Comorbidity Index because this measure was not designed with a focus on pediatric care and therefore offers only inappropriate assessment of children' comorbidity.

[26] Measured in age groups varying from 1 (age < 35 yrs.) to 10 (age between 75 and <80 yrs.).

Table 3.3: Descriptive statistics complete observation period

	Treatment group			**Control group**		
Period	Jan 2011 to Dec 2017					
Variable	Mean (SD)	Min.	Max.	Mean (SD)	Min.	Max.
Obs.	4,844			4,073		
Fraction of diagnoses	0.004 (0.01)	0	0.11	0.003 (0.01)	0	0.11
Regional structure	4.134 (1.30)	1.45	5.99	4.149 (1.35)	1.17	6
GISD	6.396 (0.36)	5.71	7.22	6.349 (0.38)	5.73	7.56
European	0.078 (0.04)	0	0.31	0.080 (0.04)	0	0.27
Non-European	0.181 (0.08)	0	0.51	0.177 (0.07)	0	0.41
Participation parents	0.318 (0.14)	0.02	1	0.274 (0.12)	0.01	0.67
Morbidity	384.745 (200.19)	0	2,108.37	386.641 (290.33)	0	10,391.96
Age pediatrician	4.959 (1.41)	1	9	5.667 (1.58)	2	10
Fraction enrolled patients	0.223 (0.30)	0	1	0.027 (0.07)	0	0.82

Table 3.3 shows the descriptive statistics for our outcome variable *fraction of diagnoses*, which measures the fraction of quarterly diagnoses per physician, and relevant covariates that we include in our estimation model. As covariates we use information on whether physician's patient-base lives on average in a more urban or rural region (*Regional structure*), information on average socio-economic status (*GISD*) of patient-base, percentage of patients with

European (*European*) and non-European citizenship (*Non-European*) and the percentage of patients whose parents participate themselves in a selective contract scheme (*Participation parents*). A comparison of the descriptive statistics shows that the average fraction of diagnoses per one hundred patients is higher for the treatment group compared to the control group (0.004 vs. 0.003). For physicians in the treatment group, the patient-bases exhibit on average a slightly higher fraction of parents enrolled in a selective contract scheme (0.318) compared to the control group (0.274). Rather straightforward, we observe a higher fraction of SCPC patients for pediatricians enrolled in SCPC (22.3% on average) compared to pediatricians not enrolled (2.7% on average). On average, pediatricians participating in SCPC are younger in comparison to non-participating physicians. With regard to nationality and socio-economic status as well as regional structure, the average patient-bases are quite similar for enrolled and non-enrolled pediatricians.

3.3.3 Estimation Strategy

In order to test our hypothesis that physicians responded to the introduction of a capitation fee (under SCPC) with an increase in the number of provided amblyopia screenings, we use a generalized difference-in-difference framework. We analyze both the extensive and intensive marginal effect of treatment. Eq. (5) estimates the extensive marginal effect of treatment or more specifically the general effect of SCPC participation on the screening rate. Here, T_{it} is a binary variable and indicates whether pediatrician i was enrolled in the SCPC scheme in quarter t. We employ Eq. (6) to estimate the intensive margin, which measures the effect of "treatment intensity" p_{it} on the outcome variable. This means we estimate the effect of a one-percentage point increase in the ratio between the number of eligible SCPC enrollees and the total number of patients in the relevant age group (represented by p_{it}) on the screening rate[27]. The reasoning is that the higher the fraction of selective contract patients is, the more

[27] The assignment to treatment and control group slightly differs between both estimation approaches. Assignment to groups for estimating extensive margin in equation (Eq. (5)) is according to explanations in the text. When estimating the intensive margin, we measure treatment intensity via the fraction $p_{it} = \frac{\textbf{Number SCPC patients aged 2–3 years}_{it}}{\textbf{Number all patients aged 2–3 years}_{it}}$. At this point we run into a problem, because we observe that SCPC patients also visit non-enrolled pediatricians. If this is the case, treatment and reimbursement takes place according to standard care terms. Although enrolled patients are contractually required to visit exclusively their contract SCPC pediatrician, no sanction or enforcement mechanisms are in place. Thus, we observe positive values $p_{it} > 0$ for non-enrolled pediatricians implying that some members of the control group are erroneously classified as treatment group members. To overcome the problem, we only identify observations with positive control group status and $p_{it} = 0$ as control group members.

affected is an enrolled pediatrician by SCPC incentives. We estimate the following estimation equations

$$ln\left(\frac{diag_{it}}{number\ patients_{it}}\right) = \alpha T_{it} + \beta' X_{it} + \gamma_i + \delta_t + u_{it} \quad (5)$$

$$ln\left(\frac{diag_{it}}{number\ patients_{it}}\right) = \alpha p_{it} + \beta' X_{it} + \gamma_i + \delta_t + u_{it}\ , \quad (6)$$

where $diag_{it}$ denotes the number of amblyopia and amblyopia-related diagnoses for pediatrician i in quarter t. Exposure variable $number\ patients_{it}$ reflects the number of possible diagnoses for pediatrician i in quarter t. The $1 \times N$ vector X_{it} compromises all covariates with information about pediatrician i's patient base and additionally pediatricians' age. We also control for time fixed effects δ_t and time-invariant pediatrician fixed effects γ_i. We employ a negative binomial fixed-effects model to derive estimates for the effect of SCPC on screening provision[28].

As already stated, we face the problem of a non-random assignment to the control and the treatment group as pediatricians self-selected into SCPC. The lack of randomization implies that we can only draw reliable conclusions with regard to the average treatment effect on the treated (ATET) and not with regard to the average treatment effect (ATE). Furthermore, self-selection might pose a threat to ATET estimator's unbiasedness as there might exist correlations between pediatricians' characteristics associated with SCPC participation and screening behavior. To counteract possible threats posed by selection bias when estimating ATET we employ a fixed-effects estimator to control for unobservable time-invariant confounders (e.g. pediatricians' baseline motivation to do screenings).[29]

[28] We use a fixed-effects negative binominal estimation model, as our outcome variable is a count with a distribution function skewed to the right (see Figure 3.4). To assess the possibility that the observed data exhibits overdispersion meaning that the outcome variable's variance is larger than the mean, we conduct a likelihood ratio test. Here, we test the null hypothesis that conditional mean and conditional variance are identical. We can reject the null hypothesis, suggesting that our outcome variable's distribution exhibits overdispersion. Therefore, we implemented the negative binominal model, which in contrast to a Poisson model accounts for overdispersion.

[29] FE-estimators do not compare differences in averages between individuals but instead only use within-variation of individuals, implying that individual-specific, time-invariant confounders cancel out and cannot bias the estimator.

Although we cannot estimate ATE, several aspects suggest that ATET might be a valid predictor for ATE. First, pre-intervention development of the outcome variable does not only exhibit similar trends for the control and the treatment group, but also similarity with respect to trend levels. This implies that no systematic differences in pediatricians' baseline screening behavior between treatment and control group is detectable (see Figure 3.3). This observation makes it plausible to assume that reimbursement change would have a similar effect on screening behavior of pediatricians in the control group. Additionally, SCPC led to a multitude of changes in pediatricians' reimbursement schemes where the change in amblyopia screening compensation is small both in relative and absolute terms. This implies that modified reimbursement in amblyopia screening is likely to have played a negligible impact on pediatricians' enrollment decision making it unlikely that baseline screening behavior and SCPC are directly associated. These considerations weaken the concern of potential self-selection bias. Furthermore, as we only consider pediatricians who owned a screening device before SCPC introduction in 2014, we know that the investment decision in a screening device took place independently from the decision to enter SCPC and thus from amblyopia reimbursement and its accompanying incentives. It is highly likely that only pediatricians who were sufficiently convinced by the additional value of early screenings invested in a device. Given all these considerations, device ownership status is likely to have a stronger association with screening provision behavior and baseline screening motivation than SCPC enrollment status.

3.4 Results

Table 3.4 shows the estimated extensive marginal effect of SCPC participation on the average number of quarterly diagnoses using the fixed-effects model specification in Eq. (5). The extensive margin measures the average effect of SCPC participation on screening provision. The coefficient of our treatment variable $Treatment$ is significantly positive (0.862) in the base model (1a) and remains significant and similar in magnitude when we add further covariates (columns 1a-1g).

For the fully specified model (1g), the estimated coefficient for $Treatment$ becomes 0.852. This means that enrolled pediatricians have on average a diagnosis rate 2.3 times higher than non-participating pediatricians have[30]. Furthermore, our estimation results indicate that patients' base characteristics like urbanization, socio-economic status, parents' enrollment status in selective schemes, citizenship and morbidity have no significant impact on diagnoses frequency.

Table 3.5 shows the results for the intensive margin estimator in estimation equation Eq. (6). Here, we estimated the coefficient for p_{it} which reflects the treatment intensity measured by the proportion of SCPC patients aged two and three years among all patients aged two and three years. For the baseline model, the coefficient estimator is positive (1.357) and highly significant. If we gradually add the full set of covariates, the magnitude of the estimator hardly changes and remains highly significant. For the fully specified model (2g) the estimator becomes 1.338 implying that, on average, if the fraction of eligible patients increases by 1 percentage point, the diagnoses rate increases by 1.3 percent.

Our results indicate that treatment intensity, which reflects the fraction of patients for whom pediatrician could only invoice the capitation fee without providing a screening service, exhibits positive correlation with pediatricians' screening rate. This further supports our hypothesis that SCPC will lead to an increase in screening provision and indicates that monetary incentives do not constitute the primary driver for pediatricians screening behavior. Given the case of a solely monetarily incentivized pediatrician, a higher fraction of eligible patients would imply a decrease in the screening rate, as service provision is costly and payment of the capitation fee takes place independently from actual provision. Instead, our estimation results indicate the very opposite. We find that an increase in the fraction of eligible patients is associated with increases in screening provision. Thus, our analysis indicates that concerns other than monetary ones strongly affect pediatricians' treatment decision.

[30] Since we have an exponential link function, the interpretation of the coefficients is not straightforward. The estimated coefficient for treatment is the difference in the natural logarithm of the outcome variable for participating physicians compared to non-participating physicians. In order to determine the incidence rate ratio, we take the exponential function of the coefficient.

Table 3.4: Effects of reimbursement reform on number of amblyopia diagnoses (extensive margin)

	Dependent variable						
	Ratio between number of relevant diagnoses and number of patients per quarter						
	(1a)	(1b)	(1c)	(1d)	(1e)	(1f)	(1g)
Treatment	0.862***	0.868***	0.868***	0.872***	0.872***	0.857***	0.852***
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Regional structure		0.085	0.065	0.034	0.037	0.025	0.007
		(0.09)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)
GISD			-0.126	-0.018	0.001	-0.050	-0.156
			(0.43)	(0.45)	(0.46)	(0.46)	(0.46)
European				2.375	2.376	2.594	2.530
				(1.58)	(1.58)	(1.59)	(1.59)
Non-European				1.534	1.510	1.631+	1.581
				(0.95)	(0.95)	(0.96)	(0.96)
Participation parents					0.165	0.214	0.171
					(0.64)	(0.64)	(0.64)
Morbidity						0.000	0.000
						(0.00)	(0.00)
Age pediatrician							0.063
							(0.06)
Constant.	-4.094	-4.433	-3.553	-4.549	-4.714	-4.282	-3.802
	(0.17)	(0.40)	(3.03)	(3.21)	(3.26)	(3.26)	(3.24)
ln(number patients)	1	1	1	1	1	1	1
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	8,917	8,917	8,917	8,917	8,917	8,917	8,917
Number of groups	*335*	*335*	*335*	*335*	*335*	*335*	*335*
	Observations per group						
Min	12	12	12	12	12	12	12
Average	26.6	26.6	26.6	26.6	26.6	26.6	26.6
Max	28	28	28	28	28	28	28
Wald chi2	394.99	394.95	394.84	400.24	400.64	403.79	403.90
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Standard errors are in parenthesis. + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 3.5: Effects of reimbursement reform on number of amblyopia diagnoses (intensive margin)

	Dependent variable						
	Ratio between number of relevant diagnoses and number of patients per quarter						
	(1a)	(1b)	(1c)	(1d)	(1e)	(1f)	(1g)
Fraction of enrolled patients	1.357*** (0.18)	1.370*** (0.19)	1.352*** (0.19)	1.355*** (0.19)	1.370*** (0.19)	1.351*** (0.19)	1.338*** (0.19)
Regional structure		0.060 (0.10)	0.147 (0.13)	0.117 (0.13)	0.104 (0.13)	0.100 (0.13)	0.086 (0.13)
GISD			0.615 (0.54)	0.794 (0.57)	0.727 (0.58)	0.606 (0.57)	0.514 (0.57)
European				1.974 (1.79)	1.961 (1.79)	2.250 (1.80)	2.183 (1.81)
Non-European				1.654 (1.02)	1.722+ (1.02)	1.917+ (1.03)	1.848+ (1.04)
Participation parents					-0.517 (0.71)	-0.469 (0.70)	-0.541 (0.71)
Morbidity						-0.0004* (0.00)	-0.0004* (0.00)
Age pediatrician							0.072 (0.07)
Constant.	4.152*** (0.18)	4.392*** (0.43)	-8.628* (3.74)	-10.090* (3.94)	-9.498* (4.04)	-8.606* (4.00)	-8.258* (3.94)
ln(number patients)	1	1	1	1	1	1	1
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7,273	7,273	7,273	7,273	7,273	7,273	7,273
Number of groups	*326*	*326*	*326*	*326*	*326*	*326*	*326*
	Observations per group						
Min	7	7	7	7	7	7	7
Average	22.3	22.3	22.3	22.3	22.3	22.3	22.3
Max	28	28	28	28	28	28	28
Wald chi2	356.49	356.37	358.68	362.94	361.91	367.68	367.79
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Standard errors are in parenthesis. + p<0.1, * p<0.05, ** p<0.01, *** p<0.001

3.5 Robustness

Our assessment of the validity of the parallel trend assumption relied on the visual inspection of the development of the average diagnoses rate – averaged over never enrolled (control group) and enrolled pediatricians (treatment group) respectively – in the years prior to SCPC introduction in 2014. In doing so, we disregard the fact that not all enrolled pediatricians entered SCPC simultaneously but rather entered successively. Possibly, pediatricians who entered SCPC immediately after introduction might exhibit a differing pre-SCPC screening behavior than pediatricians joining the program later or never do so. If this would be the case, generalizability of the main analyses results would be questionable. In order to test whether pre-treatment behavior varies systematically between non-, early- and late-enrollees, we assign the pediatricians to five different groups based on their enrollment status and date. Again, we have the standard control group consisting of all pediatricians who do not enroll during the complete observation period. Additionally, we split up the treatment group into four different groups based on the pediatricians' respective year of entry. In order to test the hypothesis that the pre-treatment trend between those groups does not differ significantly, we introduce estimation equation Eq. (7):

$$ln\left(\frac{diag_{it}}{number\ patients_{it}}\right) = \sum_g \xi_g * I(EnrollmentYear = g_i) * year_t + \beta X_{it} + \gamma_i + \delta_t + u_{it}$$
$$if\ year < EnrollmentYear \qquad (7)$$

Again, δ_t represents time fixed effects, γ_i represents time-invariant pediatrician fixed-effects and vector X_{it} compromises all covariates with information about pediatrician i's patient base as well as pediatricians' age. Furthermore, $g = \{2014, 2015, 2016, 2017\}$ denotes the year of program enrollment with g_i indicating program entry of pediatrician i. For example, if pediatrician i enrolled in $g_i = 2016$, the binary indicator variable $I(EnrollmentYear = 2016)$ becomes equal to one and for all other $g \neq 2016$ becomes equal to zero. The linear time trend we estimate for pediatrician i then becomes $\xi_g * year_t$, where only years before 2016 are taken into account as the aim is to estimate pre-enrollment trends. Our results shown in Table 3.6 suggest that none of the estimated linear time trend coefficients is significantly different from zero. Thus, we can assume that baseline pre-treatment screening behavior does not differ systematically between non-, early- and late-enrollees.

Table 3.6: Test of parallel pre-enrollment trends by time and group

	Dependent variable
Ratio between number of relevant diagnoses and number of patients per quarter	
Regional structure	-0.063
	(0.24)
GISD	-0.254
	(0.79)
European	4.311+
	(2.20)
Non-European	1.447
	(1.36)
Participation parents	0.669
	(0.92)
Morbidity	0.000
	(0.00)
Age pediatrician	-0.043
	(0.10)
*I(EnrollmentYear=2014)*year*	0.086
	(0.11)
*I(EnrollmentYear=2015)*year*	-0.167
	(0.11)
*I(EnrollmentYear=2016)*year*	0.119
	(0.10)
*I(EnrollmentYear=2017)*year*	-0.052
	(0.09)
cons.	-4.429
	(3.63)
ln(number patients)	1 (exposure)
Quarter fixed effects	Yes
N	6,340
Number of groups	*304*
Observation per group	
Min	3
Average	20.9
Max	28
Wald chi2(39)	92.32
Prob > chi2	0.0000
We use a negative binominal fixed effects model, standard errors are in parenthesis + p<0.1, * p<0.05, ** p<0.01, *** p<0.001	

As described in section 2, the data set does not contain information with respect to pediatricians' ownership status of amblyopia screening devices. To identify the ownership status, we used initial diagnoses in pre-verbal children and we included all pediatricians who had done an initial diagnosis before program introduction in 2014. However, using this approach we run into a potential problem, as we do not exactly know, when the pediatrician purchased the device. This implies that we possibly deal with an excess of zero diagnoses observations for pediatricians who purchased a device relatively late in the pre-treatment observation period. In this case, zero diagnoses observations in the periods in which the pediatrician not yet owned a screening device does not reflect a lack of screening provision but instead simply the lack of an adequate device. This might potentially lead to biased estimators. Due to the possible presence of excess zeroes, we might overestimate the treatment effect. Therefore, we implement a considerably more conservative approach by exclusively including observations for a specific pediatrician from the time point of his or her first amblyopia-associated initial diagnosis.

By doing so we, we eliminate the threat of excess zeroes and thereby the potential bias of overestimating the treatment effect. One problem in using the restrictive data sample is that the pendulum might swing in the opposite direction, as we possibly remove zero diagnoses observations, which are not excess zeroes but instead true zeroes. As incidence of amblyopia is relatively low among children, the average pediatrician, who purchases a device, is likely to experience a few quarters with zero diagnoses until making a first diagnosis. This might possibly lead to the problem that in reality, we have a higher number of true zero diagnoses than our restrictive data set indicates. A possible implication could be the underestimation of the treatment effect. To estimate the extensive and intensive margin of treatment for the more restrictive data set, we re-run the estimation equations Eq. (5) and Eq. (6) using a fixed-effects negative binominal estimation model. Table 3.7 shows the results for the fully specified models. The coefficients estimated for the extensive margin are almost identical for the restrictive sample (Table 3.7) and the full sample (Table 3.5) and are highly significant for both estimations. The same applies for the estimated coefficients for the intensive margins with 1.5 and 1.3 for the restrictive sample and the full sample respectively (Table 3.6 and Table 3.7). Again, both estimated coefficients are highly significant. This supports the results of our main analyses and lessens concerns of a possible overestimation of treatment effects.

Table 3.7: Effects of reimbursement reform on number of amblyopia diagnoses – restrictive sample

	Dependent variable		
	Ratio between number of relevant diagnoses and number of patients per quarter		
	Extensive		**Intensive**
Treatment	0.821***	***F**raction of enrolled patients*	1.473***
	(0.09)		(0.23)
Regional structure	0.210	*Regional structure*	0.235+
	(0.13)		(0.14)
GISD	-0.100	*GISD*	0.228
	(0.52)		(0.61)
European	3.872*	*European*	3.052
	(1.66)		(1.90)
Non-European	2.418*	*Non-European*	2.965**
	(0.99)		(1.09)
Participation parents	0.831	*Participation parents*	-0.211
	(0.67)		(0.74)
Morbidity	-0.000	*Morbidity*	-0.001**
	(0.00)		(0.00)
Age pediatrician	0.050	*Age pediatrician*	0.039
	(0.07)		(0.07)
cons.	-4.429	*cons.*	-6.318
	(3.63)		(4.19)
ln(number patients)	1	*ln(number patients)*	1
Quarter fixed effects	Yes	*Quarter fixed effects*	Yes
N	7,710	*N*	6,066
Number of groups	*335*	*Number of groups*	326
Observation per group		*Observation per group*	
Min	3	*Min*	*3*
Average	23.00	*Average*	18.60
Max	28	*Max*	28
Wald chi2	302.10	Wald chi2	285.86
Prob > chi2	0.0000	Prob > chi2	0.0000

We use a negative binominal fixed effects model, standard errors are in parenthesis
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

3.6 Conclusion

We analyze how a change in medical service reimbursement leads to changes in preventive service provision. Here, we use the introduction of a selective contract scheme regulating pediatricians' reimbursement in the German statutory health care system. In contrast to other study frameworks, we analyze an intervention affecting both pediatricians and patients simultaneously whereas the vast majority of previous studies mainly considered unilateral changes in physicians' reimbursement regimes. For the reform at hand, reimbursement changed from OOP-FFS to capitation fees. In general, research suggests that this kind of compensation change considerably affects supply-side incentives and is potentially associated with an under-provision of healthcare services (Hennig-Schmidt et al., 2011). This would be unambiguously the case if monetary incentives constitute the only determinant of physicians' treatment decisions. Whereas under fee-for-service each additional unit of service is associated with additional cost but also with additional revenue, under capitation, additional service provision has no impact on revenue but instead implies increasing costs for service provision (Pauly, 1995). In contrast, the situation we are looking at is a bit more complex, because not only does a change in pediatricians' reimbursement takes place, but also patients do not longer have to pay an OOP to receive the screening. Consequently, patients' willingness-to-pay and ability-to-pay might play a role in changes in service provision.

We set up a theoretical model to reflect these features and to make predictions with respect to future service provision. Our theoretical model predicts that, under standard care, where screening provision depends on patients' willingness-to-pay, a purely monetary-driven pediatrician would either reduce screening provision or keep provision constant. To incorporate experimental findings suggesting that not only monetary incentives drive physicians' treatment decisions, but instead also concerns about patients' well-being (Henning-Schmidt et al., 2011), we introduced a variable reflecting the utility gain from contributing to patient's health by conducting the amblyopia screening. Our model predicts that given the pediatrician is sufficiently interested in patients' well-being, i.e. if the marginal gain from another screening exceeds the additional costs associated of doing another screening, the total number of screenings will strictly increase.

Our empirical results are in line with our theoretical predictions. We find that screenings, measured by diagnoses per hundred patients, more than double (2.3 times higher) for pediatricians participating in the SCPC scheme. Additionally, our results show that average screening rates for physicians participating in SCPC increases with the number of enrolled patients. This indicates the importance of altruistic motives in physicians' treatment decision and the findings are also in accordance with experimental evidence, which shows that medical students exhibit a significantly higher level of altruism than those of other fields of studies (Coulter el al., 2007). This even more applies to aspiring pediatricians compared to aspirants of other medical specialties (Hojat et al., 2002; Li, 2018).

The analyzed reimbursement scheme has a simple structure and requires a low level of monitoring thereby offering participating pediatricians a high degree of freedom in making treatment decisions. Additionally, administrative effort is relatively small, as pediatricians do not have to document and bill screenings individually per patient. Our findings show that this approach led to the desired increase in screening provision. Our analysis shows that when policymakers design a reimbursement framework, which combines capitation fees and a relatively high degree of autonomy for the health service provider, several aspects have to be considered. As our findings suggest, under capitation the difference between marginal utility derived from doing the medical service in question and marginal cost of service provision has to be positive. This is the case, if the pediatrician is sufficiently altruistic and convinced about the beneficial effects of the treatment in question and marginal costs associated with the treatment are relatively low. Here, it is important to consider the target group as experimental evidence shows that specialist groups differ considerably with respect to their average degree of altruism (Li, 2018). Furthermore, it is important that health care providers and sickness fund discuss and align their objectives in order to create acceptance for the changes within both interest groups.

With respect to policy implications, our results suggest that capitated schemes do not necessarily decrease medical service provision. As many countries experienced steadily rising costs in the health care service sector during the past decades, alternative approaches for reimbursement in healthcare have been of major interest (Iverson & Lurås, 2000; Devlin & Sarma, 2008; Glazier et al., 2009). In Germany, stronger regulation is oftentimes the measure of choice when tackling problems of under- and overprovision in health care services. As this

approach limits physicians' autonomy in treatment decisions, it is prone to create resentment amongst service providers. Furthermore, it might entail several drawbacks, as it is associated with high administrative efforts and costs – for both the service provider and the statutory health insurers – in order to monitor adherence to regulatory guidelines. Additionally, high regulation can possibly compromise physicians' flexibility to provide optimal treatments, which is an especially important aspect in health care service provision as needs are highly patient-specific. This highlights the need for cost-effective reimbursement schemes offering the service provider a high degree in treatment flexibility. Our research shows that capitation might be a valuable tool, as it is can induce increases in health care service volume, but at the same time offers financial planning security for both service providers and sickness funds.

3.7 Limitations

As the number of actual amblyopia tests performed is not documented in the sickness fund's billing data, we had to use the number of documented amblyopia and amblyopia-related diagnoses as a proxy for the number of amblyopia screenings conducted implying possible inaccuracies in our analysis. Additionally, we only consider the effects of SCPC introduction on one specific health service and therefore, no conclusions can be drawn concerning the effects of SCPC introduction on other aspects of the selective contract scheme. Furthermore, we cannot make statements with respect to possible welfare effects associated with SCPC induced changes in amblyopia screening provision. As our results show an increase in the number of diagnoses, it might be possible to deduce that under the ordinary scheme, we see an underprovision of screenings. These aspects are beyond the scope of our research question and should be addressed in further research.

4 The Impact of Public Reporting and Reputational Concerns on Hospitals Quality Provision under Duopoly Competition[31]

Abstract

In this paper, I analyze a duopoly hospital market with fixed-prices where the health care service providers compete via quality. Policymakers publish outcome-based performance measures such as mortality, readmission and complication rates to reduce information asymmetry with respect to health care service quality between hospitals and patients. I assume that policymakers are able to influence the intensity with which public reporting of quality indicators affects hospitals' reputation (e.g. by legislative means of publication requirements) and consequently hospitals' respective market shares. I choose a dynamic framework reflecting that hospitals' reputation resembles a stock variable, which evolves over time and depends on hospital managers' quality decisions in the preceding periods. I find that strengthening the link between performance indicators' realization and hospitals reputation does not necessarily lead to stronger incentives for higher quality provision. In the case that marginal costs for increasing quality provision are sufficiently high and the level of competition is sufficiently low, hospitals react with decreasing quality provision to an intensified link between performance indicators and reputation. Finally, the stronger the correlation between quality and performance indicators' realization is, the lower will be the critical level of competition necessary to ensure that the effect of quality reporting is favorable on provided quality.

[31] This following chapter is a single authored manuscript by the candidate and yet unpublished.

4.1 Introduction and Background

To foster quality competition between hospitals and counter varying or insufficiently low quality provision in the health care sector, policymakers around the world have introduced public quality reporting schemes with the objective to make quality of health care services more transparent (Sinaiko et al., 2012). The underlying hope was to reduce information asymmetries between health service providers and potential patients thereby enabling the latter to make better-informed decisions when planning upcoming elective procedures (Berwick et al., 2003). In theory, well-informed patients favor hospitals with better quality reports over those with worse reports thereby incentivizing the latter to improve their service quality in order to increase their market shares and ultimately their profits (Berwick et al., 2003).Generally, several prerequisites need to be satisfied for public reporting to induce quality competition and consequently to have positive effects on quality provision. First, patients need to have access to information on health providers' quality and additionally must possess the ability to decode the given information properly in order to identify the hospital providing the highest quality services. Furthermore, health care providers need to have unused capacity at their disposal[32] and an incentive to provide additional treatments implying that marginal revenues exceed marginal costs. Assuming the latter two prerequisites satisfied, I will analyze whether intensifying the link between outcome-based performance indicators and hospitals' reputations will increase health care service providers' quality efforts.

Empirical evidence indicates that public reporting of outcome-based performance indicators such as readmission and mortality rates affect patients' hospital choices for elective procedures considerably (Varkevisser et al., 2012; Dranove, 2008; Emmert and Schlesinger, 2017). However, research also shows that patients have problems to decode provided quality information properly preventing them from rationally adapting their demand behavior to available quality information. Gourevitch et al. (2019) conducted a randomized experiment to analyze the impact of additional quality information on patients' hospital decisions. The authors find that although study participants were aware of differences in performance indicators' realization between different health care providers, study participants were not

[32] This is more likely to be the case in countries which are characterized by a less regulated hospital sector (like the US and Germany) compared to countries with higher degrees of regulation which are typically characterized by excess demand and waiting lists for procedures (e.g. GB and Italy) (Brekke et al. 2010).

able to apply the reported information when assessing their own expected risk of experiencing suboptimal health outcomes as they preferred hospitals with less favorable quality reports over those with better quality reports[33]. In line with these findings, Emmert et al. (2014) and Sinaiko et al. (2012) find that public quality records oftentimes lack general comprehensibility and that patients' level of health literacy is oftentimes not sufficient to use reported quality information appropriately. Additionally, Emmert and Wiener (2017) show that patients who integrate reports on quality indicators in their decision process are on average younger and better educated indicating that quality reports are likely to miss important target groups such as elderly and low socio-economic status patients with both groups being prone to suffer from poor health conditions (Emmert et al., 2014). These findings illustrate, that policymakers, who intend to introduce public quality reporting as a mean to reduce information asymmetries and foster hospital competition, have to pay close attention to information accessibility and comprehensibility. First, an effective implementation of public quality reporting requires potential patients to be aware of the reports' existence (Emmert et al., 2014). Second, easy accessibility to quality information supports the usage of quality reports (Emmert et al., 2014). Furthermore, policymakers can raise transparency and comprehensibility by presenting quality indicators in a target-group-specific way (Emmert et al., 2014).

Another important aspect of public quality reporting is the inherent incentive for risk selection behavior on the side of service providers. Adequate adjustments to performance indicators reflecting differences in characteristics of treated patients become therefore necessary, as otherwise service providers might have incentives to engage in risk selection resulting in hospitals' avoidance of high-risk patients in order to achieve better health outcomes in patients and consequently higher scores in performance indicators (Schneider and Epstein, 1996). Put differently, hospitals might avoid treatment of severely ill patients, who have a higher risk of mortality and/ or for experiencing complications in order to prevent negative effects on hospitals' performance indicators (Dranove et al., 2003). Empirical evidence is ambiguous on this subject. Dranove et al. (2003) analyzed the introduction of public quality

[33] Gourevitch et al. (2019**)** analyzed whether women, who were soon to be expecting child delivery, change their hospital decision when confronted with information regarding hospitals' Caesarean sections rates. The authors found that participating women were aware of the differences in the Caesarean section rates between hospitals, but that this knowledge did not lead to significantly different hospital decisions compared to the control group were the participating women did not receive quality reports.

reporting on cardiovascular procedures in New York and Pennsylvania in the early 1990s and found evidence that surgeons practicing in either of those two states were more likely to shun high-risk patients compared to surgeons working in a state without reporting requirements. In contrast, Chen and Meinecke (2012) and Chou et al. (2014) also analyzed the introduction of public quality reporting on cardiovascular procedures in Pennsylvania, but could not find evidence for risk selection[34]. Vallance et al. (2018) examined the effects of publicly reported health outcomes for elective colorectal cancer surgeries on hospitals' selection behaviors and found no evidence for hospitals to engage in risk selection after public quality reporting introduction. A theoretical analysis of health care providers' incentives for risk selection under public reporting schemes by Chen (2011) showed that incentives for risk selection depend on the distribution of patients' illness severity. Only if providers face different distributions of patient types, high quality providers might have an incentive to avoid high-risk patients in order to signal high quality. In order to abstract from possible risk selection behavior, I will use a framework, which assumes completely symmetric competitors, who face identically distributed patient types.

In this paper, I address the question of whether policymakers' engagements in strengthening the link between quality reports and hospitals' reputations necessarily promotes increases in hospitals' quality provision. Thus, I contribute to the question whether it is desirable to promote both availability and comprehensibility of quality reporting schemes such as Emmert et al. (2014) suggest. So far, empirical evidence on the effect of public reporting on hospitals' quality provisions has been ambiguous. Campanella et al. (2016) did a meta-analysis of the empirical evidence on the effects of quality reporting on quality provision, which comprises the findings of 27 studies published between 1994 and 2014. The majority of included studies analyzed data on cardiac surgeries conducted in US or Canadian hospitals. Empirical evidence is inconclusive as slightly more than half of the considered studies found a positive correlation between public reporting and quality provision, one study found a negative correlation and

[34] Contrary to Dranove et al. (2003), who compared providers' behavior before and after the actual release of public reports in 1993, Chen and Meinecke (2012) studied hospitals' behavior before and after the point of time when surgeons learnt that the regulator planned to publish data on medical outcomes in 1990. The approach by Chen and Meinecke (2012) tries to disentangle the effects of patient and provider selection as patients had no access to the quality information in the chosen observation period.

the rest found no significant effect at all.[35] Another meta-analysis conducted by Prang et al. (2021) focusing on more recently published studies (between 2000 and 2020) came to similar conclusions.

In a theoretical approach, Gravelle and Sivey (2010) analyze whether an increase in information precision with respect to hospitals' service qualities induces hospitals to increase quality efforts. They use a fixed-price duopoly framework, where hospitals compete via quality for patients and allow for asymmetric cost structures in hospitals' quality provisions. Their key result is that if competitors face only small differences in their respective costs for quality provision, the hospitals' optimal level of provided quality strictly increases in information precision. For sufficiently large differences in costs for quality provision, the opposite is true and optimal quality decreases monotonically in information precision. In a model extension, Gravelle and Sivey (2009)[36] also analyze the impact of transportation costs on quality provision using a spatial Hotelling model framework. They find that for sufficiently small transportation cost, quality increases monotonically in information precision. When transportation costs reach a critical threshold, a higher information precision does not lead to increases in quality provision anymore and the latter remains constant.

Similar to Gravelle and Sivey (2009, 2010), I study the impact of the degree of quality information (via quality reporting) on providers' choices of quality provision. I also assume a market with fixed prices and employ a spatial Hotelling model with hospitals competing via quality efforts for market shares. Contrary to Gravelle und Sivey (2009. 2010), who implement a static approach, where reputation is not explicitly modeled, I consider a dynamic framework where reputation is a stock variable, which evolves over time in dependence of hospital managers' quality decisions in preceding periods. In the framework of Gravelle and Sivey patients receive a quality signal and the policy maker is assumed to determine the precision of this quality signal directly. In contrast, I consider a framework where patients base their

[35] Almost all considered studies analyzed changes in mortality rates (22 out of the 27). Twelve studies identified a positive effect (i.e. a decrease in mortality rates), one study found a negative effect (i.e. an increase in mortality rate), seven studies did not detect any significant effect at all and two studies reported mixed effects. In this context, mixed effects refer to the result that respective studies analyzed mortality rates of different procedures and found different effects on mortality rate for different procedures. Noticeably less studies (8 out of 27) consider health outcomes other than or additional to mortality rates to assess the effect of public reporting on quality improvement. Again, the results are similar with slightly more than half of the studies finding a positive association between public reporting and health outcomes whereas the rest identifies no significant effects.

[36] This is a working paper version of the paper by Gravelle and Sivey (2010).

decisions on hospitals' reputation, where reputation is determined by observable performance indicators. Policymakers are assumed to be able to intensify the link between outcome-based performance indicators and reputation, which can be achieved via means of improved public access to performance indicators and enhanced target group-specific addressing.

My results deviate from Gravelle and Sivey (2009, 2010) in so far, as I am showing that for symmetric competitors, a more intense link between outcome-based performance indicators and hospitals' reputation does not necessarily lead to an increase in equilibrium quality provision. If hospitals' marginal costs for quality provision and patients' costs for traveling to the respective hospital are sufficiently high (i.e. if competition between hospitals is sufficiently low), fostering the link between outcome-based performance indicators and reputation results in a decrease in equilibrium quality provision. As empirical evidence on the effect of public quality reporting on health care providers' quality efforts is ambiguous, this finding might offer additional starting points for future research. Furthermore, I find that the critical threshold for transportation costs - such that an intensified link between quality-reporting and reputation induces a decrease in quality provision - becomes strictly larger in the degree of correlation between quality and performance indicator realization. This stresses the importance of choosing suitable outcome-based performance indicators, meaning that the latter have to be highly correlated with quality provision in order to avoid adverse effects of quality reporting. Furthermore, it is important to note that in the following analysis, quality constitutes the choice variable over which competing hospitals maximize their expected profits. Policymakers are able to determine the intensity of the link between performance indicators and hospitals' reputation. The present framework does not seek to make statements about welfare implications, as the objective is solely to get a better understanding of the relationship between public reporting and hospitals' quality provision.

4.2 Model

Like other differential game approaches analyzing quality competition in a duopolistic hospital market (Brekke et al., 2010; Cellini and Lisi, 2020), I model demand for medical services via a spatial Hotelling model framework (Hotelling, 1929) where hospitals compete on the dimension of service quality. I presume that potential patients are not able to observe

treatment quality directly but instead are only able to assess provided quality via hospitals' reputations. The latter develop in dependence of reported performance indicators. I assume that performance indicators are at least partially informative with respect to treatment quality, which is in line with empirical evidence showing that frequently used outcome-based performance indicators such as readmission and mortality rates are correlated with the quality of health care services (Shahian et al., 2012; Halfon et al., 2006).

Two hospitals are located at the ends of a unit line $S = [0,1]$. Patients are uniformly distributed over S and each patient demands one unit of medical service. A patient located at $x \in S$ realizes utility U when getting a treatment in hospital i characterized by reputation r_i and located at z_i:

$$U(x, z_i) = v + kr_i - \tau|x - z_i| \tag{1}$$

Here, v denotes the gross value from medical treatment, parameter k measures patients' marginal utility from another unit of reputation (perceived quality) and $\tau > 0$ denotes patients' traveling costs associated with having treatment in hospital i. The patient located at D^* is indifferent between hospital i and hospital j if:

$$v + kr_i - \tau D^* = v + kr_j - \tau(1 - D^*) \tag{2}$$

Thus, demand $D_i^*(r_i, r_j)$ becomes

$$D_i^*(r_i, r_j) = \frac{1}{2} + k\frac{(r_i - r_j)}{2\tau} \tag{3}$$

This implies that given hospital i has a higher reputation than hospital j, the market share of hospital i is larger than one-half. The magnitude of the excess demand depends positively on the product of excess reputation and patients' marginal utility for reputation. In contrast, patients' costs for traveling τ affect demand negatively.

The present framework models reputation as a stock variable implying that hospitals are not able to affect the realization of reputation on period t directly, but instead only indirectly via

quality choice and resulting realization of the performance indicators. Reputation accumulation takes place continuously over time and depends on realization of reputation in the previous periods. I assume that two factors drive how reputation changes over time. Firstly, hospitals' reputation depends on reported performance indicators and secondly accumulated stock of reputation depreciates over time due to patients' tendency to forget about information regarding hospitals' quality as time goes by (Hibbard et al., 2005).

The following law of motion describes the development of reputation over time:

$$\frac{\partial r(t)}{\partial t} \equiv \dot{r}_i = \Big(1 - \varphi\big(1 - \delta q_i(t)\big)\Big) - \omega r_i \tag{4}$$

Here, the term $\big(1 - \delta q_i(t)\big) \in [0,1]$ denotes hospital i's outcome for a defined performance indicator.[37] In this framework, I focus on performance indicators measuring unfavorable outcomes like mortality, readmission or complication rates. I do so, because these are measures widely implemented in actual reporting systems (Fischer et al., 2014; Birkmeyer and Dimick, 2004). As literature suggests that the aforementioned measures depend negatively on quality (Jha et al., 2007), I model the relationship between quality and performance indicators inversely proportional, i.e. increases in quality induce decreases in all of the aforementioned rates. For the sake of simplicity, I will only refer to mortality rate in the remainder of the paper.

The expression $\Big(1 - \varphi\big(1 - \delta q_i(t)\big)\Big)$ reflects the impact of performance indicators on reputation, where parameter $\varphi \in (0,1)$ measures how sensitive reputation reacts towards changes in the outcomes of performance indicators.[38] Finally, ω reflects people's tendency to forget about information regarding hospitals' reputations over time. This is in line with empirical findings indicating that individuals recall information less accurately with regard to hospitals' reputations after time has passed (Hibbard et al., 2005).

These assumptions imply that the reputation rate increases by $\Big(1 - \varphi\big(1 - \delta q_i(t)\big)\Big)$ and decreases by ωr_i. The term $\Big(1 - \varphi\big(1 - \delta q_i(t)\big)\Big)$ is separable into two parts: The first part

[37] The assumption $\big(1 - \delta q_i(t)\big) \in [0,1]$ is associated with parameter constraints. In section 3, I will derive the corresponding parameter constraints, which ensure that $\big(1 - \delta q_i(t)\big) \in [0,1]$ is satisfied in the steady state.

[38] I define link intensity $\varphi \in (0,1)$ on an open interval in order to rule out the unrealistic cases for which realization of performance indicators have no impact at all on reputation, i.e. $\varphi = 0$, or perfectly translate into reputation, i.e. $\varphi = 1$.

reflects increases in hospitals' reputations due to general medical advancements (improvements in medical technologies, better-educated medical staff and scientific progress) and affects all hospital to the same extent. For the sake of simplicity it is normalized to one. In contrast, the second part of the term given by $\varphi\big(1-\delta q_i(t)\big)$ is hospital-specific and describes how the realization of a given performance indicator affects reputation. The degree to which the reported performance indicator influences reputation is modeled via parameter φ. Given that performance indicators only have a very small impact on reputation, i.e. $\varphi \to 0$, the hospital-specific reputation component becomes very small with $\varphi\big(1-\delta q_i(t)\big) \to 0$. Consequently, reputation changes with rate $\dot{r}_i \xrightarrow{\varphi \to 0} 1-\omega r_i$ and solely depends on the general effect of technological advancements in the health care sector. In the contrary case, when hospital-specific performance indicators have a maximum impact on reputation, i.e. $\varphi \to 1$, the reputation growth rate becomes $\dot{r}_i \xrightarrow{\varphi \to 1} \delta q_i(t)-\omega r_i$, which is strictly smaller than $\dot{r}_i = 1-\omega r_i$. This implies that for a given quality level $q_i(t)$, the rate with which reputation changes over time is monotonically decreasing in φ. As many published performance indicators reflect adverse outcomes like mortality, readmission or complication rates, the reasoning is that a larger φ implies a general rise in patients' awareness about the possibility of unfavorable outcomes. Thus, the higher the awareness about unfavorable outcomes, the stronger the impact of individuals' assessment on the probability of experiencing adverse outcomes and therefore the lower hospital's reputation. It is important to note, that the general awareness about unfavorable outcomes evoked by a larger φ affects both hospitals symmetrically and is reflected in the term $\varphi\big(1-\delta q_i(t)\big)$ by the minuend φ.

For a given φ, r_i strictly increases in q_i implying that for a fixed link intensity between performance indicators and reputation, reputation strictly increases in the level of provided quality. This is straightforward, as this framework assumes unfavorable performance indicators to decrease monotonically in quality efforts therefore resulting in an increased growth rate of reputation. This reflects the assumption that hospitals are able to improve their reputation by raising quality provision.

Hospitals are assumed to maximize the present value of their respective profit and hospital i's instantaneous expected profit becomes:

$$E(\pi_i(t)) = p\left(\frac{1}{2} + k\frac{\left(r_i(t) - r_j(t)\right)}{2\tau}\right) - \frac{\beta}{2}\left(\frac{1}{2} + k\frac{\left(r_i(t) - r_j(t)\right)}{2\tau}\right)^2 - \frac{\theta}{2}\left(q_i(t)\right)^2 + \vartheta q_i(t) r_i(t) - F \tag{5}$$

Here, p denotes the payment the hospital receives per patient for a given treatment. I assume that hospitals are price-takers and that reimbursement takes place via a prospective diagnosis-related group system where reimbursement p is determined by an average-cost rule.[39] Hospital i faces demand $D = \frac{1}{2} + k\frac{\left(r_i(t) - r_j(t)\right)}{2\tau}$, where D depends on hospital i's and hospital j's reputation. Costs are assumed to increase convexly in the number of treatments and in the level of provided quality. Additionally, hospitals face fixed costs F. In this framework, I assume that reputation does not only affect demand but also affects costs associated with providing higher quality (i.e. costs associated with recruiting skilled staff or raising financial means for investments in buildings and technical equipment). Literature indicates that reputation affects hiring costs and makes it easier to acquire investors and to raise funds in general (Turban and Cable, 2003; Hibbard et al., 2005; Helm, 2007). Consequently, hospital i faces costs of $\left(\frac{\theta}{2}\left(q_i(t)\right)^2 - \vartheta q_i(t) r_i(t)\right)$ when providing quality level $q_i(t)$, implying that costs monotonically decrease in reputation. To ensure that quality provision is always associated with positive costs even for very high levels of reputation I assume that $\left(-\frac{\theta}{2}q_i^2 + \vartheta q_i r_i\right) < 0$ holds.[40]

[39] The reimbursement p hospitals receive in period t is determined by the average treatment costs of hospitals in period $(t-1)$. Thus, the regulator only has limited tools to affect hospitals' quality incentives, as he is not able to determine prices with which hospitals are reimbursed directly.

[40] The assumption $\left(-\frac{\theta}{2}q_i^2 + \vartheta q_i r_i\right) < 0$ is associated with parameter constraints. In section 3, I will derive the corresponding parameter constraints, which ensure that the aforementioned condition is satisfied in the steady state.

Then, hospital i's maximization problem becomes

$$max_{q_i} \int_0^{\infty} E\big(\pi_i(t)\big)e^{-\rho t}dt \tag{6}$$

subject to

$$\dot{r}_i = \big(1 - \varphi(1 - \delta q_i)\big) - \omega r_i \tag{7}$$

$$\dot{r}_j = \Big(1 - \varphi\big(1 - \delta q_j\big)\Big) - \omega r_j \tag{8}$$

$$r_i(0) = r_j(0) = r(0) > 0 \tag{9}$$

where ρ denotes the rate of hospital i's time preference and q_i is the control variable. I assume that all introduced parameters are strictly larger than zero.

4.3 Open-loop equilibrium

The open-loop equilibrium concept assumes that competitors know the initial state of the system, i.e. $r(0)$ and choose their optimal strategy (i.e. each hospital choose the optimal time path for their control variable q) at the beginning of the game in $t = 0$ and stick to this strategy for the remainder of the game. This equilibrium concept is either appropriate if players are not able to observe the development of their competitor's stock variable over time (which is not realistic in this case as reputation is easily observable), or if the plan for the control variable requires a high degree of commitment. In health care service provision, quality often depends on long-term investment plans, which hospital management cannot easily revise during the game. Thus, I consider the open-loop equilibrium in the remainder of the paper.

The Langrarian associated with hospital i's maximization problem is:

$$\begin{aligned} L = \\ \int_0^{\infty} e^{-\rho t}\left(p\left(\frac{1}{2} + k\frac{\big(r_i(t) - r_j(t)\big)}{2\tau}\right) - \frac{\beta}{2}\left(\frac{1}{2} + k\frac{\big(r_i(t) - r_j(t)\big)}{2\tau}\right)^2 - \frac{\theta}{2}\big(q_i(t)\big)^2 + \vartheta q_i(t) r_i(t)\right)dt \\ + \int_0^{\infty} e^{-\rho t}\mu_i(t) * \big(\dot{r}_i - 1 + \varphi - \varphi\delta q_i(t) + \omega r_i(t)\big)dt \\ + \int_0^{\infty} e^{-\rho t}\mu_j(t) * \Big(\dot{r}_j - 1 + \varphi - \varphi\delta q_j(t) + \omega r_j(t)\Big)\,dt \end{aligned} \tag{10}$$

Here, μ_i and μ_j denote the current value co-state variables associated with the two law-of-motion equations for reputation stock for hospital i and hospital j respectively. The co-state variables can be interpreted as Lagrange multipliers belonging to the state equations (7) and (8). Partially integrating and re –writing yields:

$$L = \int_0^{\infty} e^{-\rho t}\left(p\left(\frac{1}{2}+k\frac{(r_i-r_j)}{2\tau}\right)-\frac{\beta}{2}\left(\frac{1}{2}+k\frac{(r_i-r_j)}{2\tau}\right)^2-\frac{\theta}{2}\left(q_i(t)\right)^2+\vartheta q_i r_i\right)dt +$$
$$\int_0^{\infty} e^{-\rho t}\mu_i(t)(\omega r_i-\varphi\delta q_i-1+\varphi)dt+[e^{-\rho t}\mu_i(\infty)r_i(\infty)-$$
$$e^{-\rho t}\mu_i(0)r_i(0)]_0^{\infty}-\int_0^{\infty}(-\rho e^{-\rho t}\mu_i(t)+e^{-\rho t}\dot{\mu}_i)\,r_i dt+\int_0^{\infty}e^{-\rho t}\mu_j(t)*\left(\dot{r}_j-1+\varphi-\right.$$
$$\left.\varphi\delta q_j+\omega r_j\right)dt \qquad (11)$$

Then, the first-order conditions for a local maximum (q_i^*, r_i^*) are:[41]

$$\frac{\partial L}{\partial q_i(t)} = -\theta q_i+\vartheta r_i-\varphi\delta\mu_i(t)=0 \qquad (12)$$

$$\frac{\partial L}{\partial r_i(t)} = \frac{pk}{2\tau}-\left(\frac{1}{2}+k\frac{(r_i-r_j)}{2\tau}\right)\frac{\beta k}{2\tau}+\vartheta q_i+\mu_i(t)(\omega+\rho)-\dot{\mu}_i(t)=0 \qquad (13)$$

$$\frac{\partial L}{\partial \mu_i(t)} = \dot{r}_i-1+\varphi-\varphi\delta q_i+\omega r_i=0 \qquad (14)$$

$$\frac{\partial L}{\partial \mu_j(t)} = \dot{r}_j-1+\varphi-\varphi\delta q_j+\omega r_j=0 \qquad (15)$$

Solving (12) for $\mu_i(t)$:

$$\mu_i(t) = -\frac{\theta}{\varphi\delta}q_i+\frac{\vartheta}{\varphi\delta}r_i \qquad (16)$$

[41] In order to satisfy the second-order condition for a local maximum (q_i^*, r_i^*), the Hamiltonian $H_i = p\left(\frac{1}{2}+k\frac{(r_i-r_j)}{2\tau}\right)-\frac{\beta}{2}\left(\frac{1}{2}+k\frac{(r_i-r_j)}{2\tau}\right)^2-\frac{\theta}{2}{q_i}^2+\vartheta q_i r_i+\mu_i(t)(1-\varphi+\varphi\delta q_i-\omega r_i)$ associated with the dynamic game, has to be concave in the control variable q_i and the state variable r_i implying that $H_{qq}<0$ and $H_{rr}<0$ has to hold. Furthermore, the determinant of the associated Hessian matrix has to be larger than zero. Here,we have that $H_{qq}=-\theta<0$ and $H_{rr}=-\frac{\beta k^2}{4\tau\tau^2}<0$, implying that the Hamiltonian is concave in both the control and state variable. Furthermore, the determinant of the Hamiltonian's Hessian matrix is positive whenever $\theta\beta > \frac{4\tau^2\vartheta^2}{k^2}$ which is satisfied for a sufficiently convex cost function.

Totally differentiating (16) with respect to time yields:

$$\frac{\partial \mu}{\partial t} = \dot{\mu}_i(t) = -\frac{\theta}{\varphi\delta}\dot{q}_i + \frac{\vartheta}{\varphi\delta}\dot{r}_i \tag{17}$$

Substituting (14) into (17) yields:

$$\dot{\mu}_i(t) = -\frac{\theta}{\varphi\delta}\dot{q}_i + \frac{\vartheta}{\varphi\delta}(1 - \varphi + \varphi\delta q_i - \omega r_i) \tag{18}$$

Finally, substitution (16) and (18) into (13) and solving for $\dot{q}_i$ gives:

$$\dot{q}_i = -\frac{\left(k\varphi\delta\left(p-\frac{\beta}{2}\right)-2\tau\vartheta(1-\varphi)\right)}{2\tau\theta} - r\frac{\vartheta(2\omega+\rho)}{\theta} + q(\omega+\rho) \tag{19}$$

The dynamics of reputation is given by:

$$\dot{r}_i = 1 - \varphi + \varphi\delta q_i - \omega r_i \tag{20}$$

Thus, (19) and (20) describe the dynamics of the system. As I assume symmetric competitors, implying that $q_i = q_j$ and $r_i = r_j$, the dynamics around the steady state can be written in matrix form as:

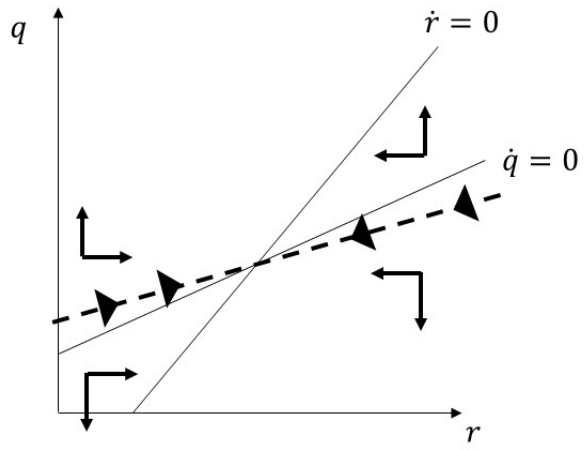
$$\begin{bmatrix}\dot{q}\\ \dot{r}\end{bmatrix} = \begin{bmatrix}(\omega+\rho) & \left(-\frac{\vartheta(2\omega+\rho)}{\theta}\right)\\ \varphi\delta & (-\omega)\end{bmatrix} * \begin{bmatrix}q\\ r\end{bmatrix} + \begin{bmatrix}-\frac{\left(k\varphi\delta\left(p-\frac{\beta}{2}\right)-2\tau\vartheta(1-\varphi)\right)}{2\tau\theta}\\ (1-\varphi)\end{bmatrix} \tag{21}$$

The two-by-two matrix is the Jacobian J, which determines the dynamics of the underlying game. The trace of the system's Jacobian is always positive, as $tr(J) = \rho$ and $\rho > 0$ by assumption. Given that θ is sufficiently large, i.e. $\theta > \frac{\varphi\delta\vartheta(2\omega+\rho)}{(\omega+\rho)\omega}$, the Jacobian's determinant is smaller than zero (i.e. $det(J) < 0$), implying that the equilibrium is stable in the saddle-path sense. In Figure 4.1, the steepest line represents all combinations of q and p for which $\dot{r} = 0$, i.e. for which reputation does not change and remains constant over time. The second steepest line gives all combinations of q and p for which the change rate of quality is equal to zero.[42] The intersection of both lines indicates the steady state, where both quality

[42] The line representing all combinations of q and p for which $\dot{r} = 0$ holds is steeper compared to the line representing all combinations of q and p for which $\dot{q} = 0$, because in order to ensure the existence of a saddle-path equilibrium $det(J) < 0$ has to hold implying that $\theta > \frac{\varphi\delta\vartheta(2\omega+\rho)}{(\omega+\rho)\omega}$.

and reputation remain constant over time. The dashed line indicates the saddle-path. For all combinations of q and p on the dashed line located to the right of the intersection, both q and p are decreasing thereby converging towards the steady state on the saddle-path. The opposite is true for combinations of q and p located on the dashed line to the left of the equilibrium. Here, both q and p increase over time thus converging on the saddle-path towards the equilibrium.

Figure 4.1: Saddle path dynamics



In steady state, both reputation and quality remain constant over time implying that $\dot{r} = 0$ and $\dot{q} = 0$ has to hold. Thus, solving the matrix system for r^{OL} and q^{OL} yields:

$$r^{OL} = \frac{(1-\varphi)}{\omega} + \frac{\varphi\delta}{\omega} q^{OL} \tag{22}$$

$$q^{OL} = \frac{\left(\omega k \varphi \delta\left(p - \frac{\beta}{2}\right) + 2\tau\vartheta(1-\varphi)(\omega+\rho)\right)}{2\tau\left((\omega+\rho)\theta\omega - \vartheta\varphi\delta(2\omega+\rho)\right)} \tag{23}$$

The steady state quality level $q^{OL} > 0$ is strictly larger than zero whenever the condition for the existence of a saddle-path equilibrium $\theta > \frac{\varphi\delta\vartheta(2\omega+\rho)}{(\omega+\rho)\omega}$ is satisfied. Furthermore, $q^{OL} > 0$ ensures that the performance measure at hand is strictly larger than zero, i.e. that $(1 - \delta q^{OL}) < 1$.[43] From (23) it is immediately apparent that the steady state quality strictly

[43] The condition $(1 - \delta q^{OL}) > 0$ is satisfied whenever $k < \frac{2\tau\left((\omega+\rho)\theta\omega - \vartheta\delta(\varphi\omega+(\omega+\rho))\right)}{\omega\varphi\delta^2\left(p-\frac{\beta}{2}\right)}$.

increases in k, the marginal impact of hospital's reputation on patients' utility. Furthermore, steady state quality increases in parameter δ, which measures how intensely quality affects performance indicators. This implies that the higher the marginal impact of quality on realization of performance indicator is the higher becomes equilibrium quality. Also higher reimbursement for treatments (represented by p) and stronger cost-reducing effects of reputational effects on quality investments (measured by ϑ) induce higher quality levels in equilibrium. Contrary, both higher marginal costs for providing treatments and improving quality induce lower levels of equilibrium quality. Taking the first derivative of q^{OL} with respect to patients' transportation costs τ yields:

$$\frac{\partial q^{OLS}}{\partial \tau} = \frac{\left(\delta\varphi k\omega 2\left(p-\frac{\beta}{2}\right)\right)}{\left(4\tau^2\left(\delta\varphi\vartheta(\rho+2\omega)-\theta\omega(\rho+\omega)\right)\right)} < 0 \tag{24}$$

This implies that equilibrium quality strictly decreases in transportation costs[44], i.e. the less competitive the hospital market the smaller becomes equilibrium quality.

Because I assume that reimbursement is determined via an average-cost rule implying that policymakers are not able to determine reimbursement directly, the aforementioned only have limited tools to affect hospitals' incentives for quality provision. Thus, the link between reported performance indicators and hospitals' reputation is of special interest as policymakers have means to influence this relation actively. Because the impact of parameter φ, representing the link intensity, on q^{OL} is not directly apparent from (23), a further analysis of the first derivative of q^{OLS} with respect to φ becomes necessary:

$$\frac{\partial q^{OLS}}{\partial \varphi} = \frac{(\rho+\omega)\left(\delta\omega\left(8\tau\vartheta^2+2k\theta\omega\left(p-\frac{b}{2}\right)\right)-4\tau\vartheta(\theta\omega(\rho+\omega)-\delta\rho\vartheta)\right)}{\left(4\tau\left(\theta\omega(\rho+\omega)-\delta\varphi\vartheta(\rho+2\omega)\right)^2\right)} \gtreqless 0 \tag{25}$$

Additionally, I assume that costs associated with quality provision are strictly positive in steady state, i.e. that the condition $\left(\frac{\theta}{2}(q^{OL})^2 - \vartheta q^{OL} r^{OL}\right) > 0$ has to hold, which is the case for parametrization $k > \frac{2\vartheta\tau(1-\varphi)(\theta(\omega+\rho)-2\vartheta\varphi\delta)}{(\theta\omega-2\vartheta\varphi\delta)\varphi\delta\left(p-\frac{\beta}{2}\right)}$ and $\vartheta < \frac{\theta\omega}{2\varphi\delta}$.

Thus, in the remainder of the analysis I consider the interval $\frac{2\tau\left((\omega+\rho)\theta\omega-\vartheta\delta(\varphi\omega+(\omega+\rho))\right)}{\omega\varphi\delta^2\left(p-\frac{\beta}{2}\right)} > k > \frac{2\vartheta\tau(1-\varphi)(\theta(\omega+\rho)-2\vartheta\varphi\delta)}{(\theta\omega-2\vartheta\varphi\delta)\varphi\delta\left(p-\frac{\beta}{2}\right)}$, where $\vartheta < \frac{\theta\omega}{2\varphi\delta}$ ensures that the lower bound of the interval is strictly larger than zero.

[44] As the condition $\theta > \frac{\varphi\delta\vartheta(2\omega+\rho)}{(\omega+\rho)\omega}$ has to hold to ensure the existence of a saddle-path equilibrium, the denominator is strictly smaller than zero.

Here, the direction of the relationship between φ on q^{OL} depends on the realization of parameters. A marginal increase in φ leads to an increase in equilibrium quality, i.e. $\frac{\partial q^{OLS}}{\partial \varphi} > 0$ whenever

$$\theta < \bar{\theta} = \frac{\vartheta\delta(2\omega+\rho)}{\omega(\rho+\omega)} \tag{26}$$

This implies that given marginal costs for quality provision are sufficiently small, i.e. $\theta < \bar{\theta}$, an increase in the link between performance indicators and reputation will always lead to an increase in quality.

In contrast, for sufficiently large θ, i.e. for all $\theta > \frac{\vartheta\delta(2\omega+\rho)}{\omega(\rho+\omega)}$, the direction of the marginal impact of φ on q^{OLS} depends on the magnitude of transportation costs.

Equilibrium quality decreases in φ, i.e. $\frac{\partial q^{OLS}}{\partial \varphi} < 0$, whenever

$$\tau > \bar{\tau} = \frac{k\theta\omega\left(p-\frac{b}{2}\right)}{2\vartheta\left(\theta\omega(\rho+\omega)-\vartheta\delta(2\omega+\rho)\right)} \tag{27}$$

with $\bar{\tau} > 0$. This implies that for sufficiently high transportation costs (which reflects a low degree of competition between hospitals) a further strengthening of the link between reported performance indicators and reputation induces a decrease in equilibrium quality. The opposite is true whenever $\tau < \bar{\tau}$, because in this case, the first derivative of equilibrium quality with respect to φ is strictly positive. This leads to *Proposition 1*.

Proposition 1:

Given that marginal costs for quality provision are sufficiently small, intensifying the link between performance indicator outcomes and reputation will always have a positive effect on equilibrium quality. However, the effect of link intensity on quality becomes ambiguous as soon as marginal costs for quality provision become sufficiently large. In this case, the effect of link intensity on q^{OLS} depends on the magnitude of transportation costs. Given that the latter exceed the critical threshold $\bar{\tau}$, strengthening the aforementioned link induces a decrease in equilibrium quality.

The intuition behind proposition 1 is as follows: The relationship between φ and reputation r is negative (as this framework assumes performance indicators measuring adverse outcomes),

implying that, ceteris paribus, an increase in φ leads to a decrease in r. Hospitals have an incentive to counter this decrease in reputation by raising quality efforts q to improve reputation and achieve higher market shares. This incentive for raising quality exists as long as the marginal profit from providing additional treatments is positive, i.e. the difference between the marginal revenue from another treatment and the associated costs (consisting of costs for providing an additional unit of treatment and investments in quality improvement) is positive. For sufficiently low marginal costs for quality provision θ, hospitals always have an incentive to offset the loss in r provoked by an increase in φ as the marginal gain from increasing their market share will always exceed associated marginal costs. In the contrary case, if marginal costs for quality provision are relatively high, it is not necessarily true that hospitals have an incentive to offset the loss in reputation induced by an increasing φ. In this case, hospitals' response to an increasing φ depends on transportation costs τ. In highly competitive market environments (characterized by low transportation costs τ), hospitals' market shares react strongly to reputation differences between competitors. In this case, the marginal gain from additional quality provision exceeds the associated costs and expanding quality efforts will be profitable. However, hospitals might have no incentives to raise quality in response to an increased φ if transportation costs are relatively high. In this case, reputation loss has only a minor impact on future hospitals' revenues, because patients' demand is relatively insensitive towards quality differences between hospitals. Therefore, marginal revenue generated by an increase in quality via increased reputation and future market share is smaller than the marginal costs of treatment and quality efforts.

These findings imply that policymakers, who aim to increase health care quality by means of public reporting, have to make sure that either costs for quality efforts are sufficiently small or, if this is not the case, patient mobility and thereby hospital competition is sufficiently high.

It is also interesting to note that, ceteris paribus, the critical threshold $\bar{\tau}$ monotonically increases in δ which measures the marginal impact of quality on the performance indicators. Thus, given that $\theta > \bar{\theta}$ is true, the larger the marginal impact of quality on performance indicators is, the higher becomes the largest possible level of transportation cost such that $\frac{\delta q^{OLS}}{\delta \varphi} > 0$ still holds.

Proposition 2:

The stronger the impact of quality on performance indicators' outcomes δ, the higher becomes the maximum level of transportation cost $\bar{\tau}$ such that a positive impact of strengthening the link between performance indicators' outcomes and reputation still has a positive impact on equilibrium quality.

Proposition 2 underlines the importance of formulating appropriate performance indicators when designing a public reporting system. The effect of δ on q^{OL} is twofold: As (23) shows, a stronger link between quality and performance indicators directly incentivizes higher equilibrium quality levels. An increase in δ leads to an increase in the future market share and thereby in the marginal gain from quality service provision. Consequently, hospitals' optimal level of quality provision increases. Additionally, a larger δ leads, ceteris paribus, to an increase in the critical threshold for transportation cost $\bar{\tau}$. This implies that an adverse (negative) effect of φ on q^{OL} starts to arise for higher costs of transportation τ. The intuition driving this result is as follows: A higher δ is associated with a higher degree of dependence between quality and reputation. Thus, an increase in δ is associated with an increased marginal gain for a given level of quality, because quality has, ceteris paribus, a higher impact on reputation and therefore on future market shares. As costs associated with quality improvements remain unaffected from δ, the level of q increases such that marginal gain from implementing quality measures equals marginal costs. Therefore, policymakers have to ensure that implemented performance indicators exhibit a strong correlation with quality efforts in order to minimize unintended adverse effects on quality provision.

Given that marginal costs associated with quality efforts are sufficiently high, equilibrium quality increases or decreases monotonically in φ depending on the model parameterization. A regulator seeking to increase health care service quality will choose the following solutions. In case that $\tau < \bar{\tau}$, the regulator will choose $\varphi \to 1$ because equilibrium quality q^{OLS} strictly increases in φ. The opposite is true if $\tau > \bar{\tau}$, then the first derivative of equilibrium quality strictly decreases in φ and policymakers achieve maximum q^{OLS} with implementing $\varphi \to 0$. Therefore, it is crucial for policymakers to evaluate the degree of patient mobility and resulting hospital competition when deciding about the introduction of a public reporting system.

4.4 Conclusion

In this framework, I analyzed how the link between performance indicators' outcomes and reputation affects hospitals' incentives for quality provision. With many countries having introduced quality reports in the in- and outpatient sector in the hope to foster quality competition, it is important to better understand how public reporting affects quality provision incentives. The main result of this paper is that if marginal costs for additional quality efforts are sufficiently low, intensifying the link between quality reports and reputation will always result in increased quality provision. The effect of the aforementioned link is ambiguous if marginal costs for additional quality efforts are sufficiently high. In this case, an increase in link intensity only has a positive effect if competition between hospitals is sufficiently high. If the opposite is the case, intensifying the link results in adverse effects on quality provision.

The second central finding is that the higher the correlation between quality and performance measure outcomes is, the lower becomes the minimum competition level required to ensure that an increased link between public reporting and reputation has a positive effect on provided quality. This implies that a sensible choice of performance measures can minimize the risk of adverse effects of public reporting on quality provision provoked by a lack of competition.

First, these results might explain ambiguous empirical results regarding the effect of quality reporting on hospitals' quality efforts and offer guidelines for the design of future empirical studies. Particularly by stressing the need to control for hospitals' costs for investments in quality efforts and the degree of competition intensity.

With respect to policy implications, the costs hospitals face when investing in additional quality efforts are crucial. If hospitals operate within an economy where costs for quality provision are relatively low, adverse effects provoked by intensifying the link between performance indicator outcomes and reputation are unlikely to occur. This implies that policymakers wanting to improve health care quality by means of public reporting also have to ensure that costs for quality investments are relatively low (e.g. by improving health care workers' education, fight possible shortages in health care workers labor supply and facilitating access to financial means in general). Given that investment costs are relatively

high, policymakers have to assess the level of competition between hospitals in order to evaluate whether fostering the impact of public reporting on reputation is a promising tool to raise health service quality.

This finding is also of relevance in the context of political efforts to foster a higher specialization of hospitals. For several years, German sickness funds have been demanding a reduction in the number of hospitals and a stronger focus on hospital specialization (Augurzky et al. 2020). The underlying assumption is that for a given hospital the outcome volume increases the smaller the total number of hospitals becomes and that higher volumes lead to learning effects, which finally result in higher service quality (Augurzky et al. 2020). Again, if policymakers ensure that investing in additional quality efforts is associated with low costs, simultaneous fostering of hospital specialization and link intensity of reporting on reputation does not lead to adverse incentives for quality provision. However, for economies characterized by high costs for quality improvement efforts (e.g. countries characterized by a shortage in health care labor supply), the picture looks different. In this case a simultaneous effort to foster link intensity between public reporting and reputation and to promote hospital specialization might lead to adverse effects in quality provision. In this case, policymakers have to consider and evaluate the compatibility of public reporting schemes and higher specialization of hospitals carefully, as a higher degree of specialization is associated with a lower number of hospitals and therefore less competition.

It is important to note that I do not consider welfare implications in this framework. Clearly, creating, aggregating, evaluating and providing data on hospitals' performances as well as enhancing patients' health literacy is associated with costs. These costs are borne by hospitals, policymakers and other (public) institutions involved in the reporting process. Consequently, to get a better understanding on the desirability of public reporting from the societal welfare's perspective, further research is necessary.

5 General Conclusion

This thesis has analyzed the introduction of different incentives schemes implemented in the in- and out-patient sector aiming at improving cost-effectiveness in health care provision and at raising service quality. The key results of the undertaken empirical and theoretical analyses show how monetary and non-monetary incentives affect service providers' behavior.

The analysis in Chapter 2 indicates that the introduction of a surgical suite governance document is associated with significant reductions in first case delays. Hence, a surgical suite governance document seems to offer a promising way to incentivize health care workers' behavior to a more effective use of costly resources like surgery capacities. As the governance document formulated only a very few guidelines, it shows that OR staff already is aware about process deficiencies and is able to autonomously find ways to overcome these. Thus, a surgical suite governance document constitutes an inexpensive way to overcome process deficiencies with a low risk to provoke OR personal resistance as being to limiting or interfering.

Chapter 3 considers the change from OOP-FFS to capitation in the in-patient sector. This change affected both, the health care providers and patients simultaneously, thereby making predictions regarding physicians' service provision behavior difficult. In standard economic theory, transitions from FFS to capitations are associated with a decline in services. A theoretical analysis predicts that given health care providers are sufficiently interested in patients' well-being, the introduction of capitations will be associated with an increase in services. The empirical analysis based on routine data provided by a large German sickness fund, confirms this prediction. This indicates that physicians are not only incentivized by monetary incentives. Hence, capitations can be a valuable tool to ensure planning security in sickness funds' expenses and health care providers' revenues without jeopardizing the quality of health care provision.

The analysis in Chapter 4 focuses on the impact of public reporting on quality incentives for hospitals. The main finding is that strengthening the link between performance indicators' realization and hospitals reputation does not necessarily lead to stronger incentives for quality provision. For a sufficiently low level of competition and a sufficiently high level of costs associated with quality provision, an intensified link between performance indicators and

reputation can possibly induce a decrease in quality provision. This finding offers practical relevance as more and more countries have implemented quality reporting schemes. Additionally, many countries aim at higher degrees of hospitals specialization, which is necessarily associated with a lower intensity of competition. Thus, a simultaneous encouragement of hospital specialization and intensified efforts to foster the importance of public reporting might result in an unintended decrease in quality provision.

In summary, these findings show that policymakers have to assess the respective situation and affected target group carefully when designing incentive schemes. As health care provision has to account for the high degree of heterogeneity in patients' needs, incentive schemes have to maintain health care providers' flexibility in treatment decisions. Furthermore, possible interdependencies and conflicts with other policy measures have to be considered carefully.

6 Literature

Aldebasi, Y. H. (2014). Prevalence of correctable visual impairment in primary school children in Qassim Province, Saudi Arabia. *Journal of Optometry*, *7*(3), 168-176.

Arrow, KJ. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review, 358*. 941-973.

Augurzky, B., Hentschker, C., Pilny, A., & Wübker, A. (2020). Krankenhausreport 2018: Schriftenreihe zur Gesundheitsanalyse.

Berry, M., Schüpfer, G., Martin, J., Bauer, M., Denz, C., Bender, H. J., & Schleppers, A. (2008). Controlling and reporting in operating room management: analysis of the state of development. *Der Anaesthesist*, *57*(3), 269-274.

Berwick, D. M., James, B., & Coye, M. J. (2003). Connections between quality measurement and improvement. *Medical Care*, I30-I38.

Birkmeyer, J. D., & Dimick, J. B. (2004). Potential benefits of the new Leapfrog standards: effect of process and outcomes measures. *Surgery*, *135*(6), 569-575.

Bohnenkamp, B., Braun, S. (2017). Arbeitsabläufe: Effizientes OP-Management – Worauf es wirklich ankommt. *Deutsches Ärzteblatt, 114*(2),35-36.

Brekke, K. R., Cellini, R., Siciliani, L., & Straume, O. R. (2010). Competition and quality in health care markets: a differential-game approach. *Journal of Health Economics, 29*(4), 508-523.

Campanella, P., Vukovic, V., Parente, P., Sulejmani, A., Ricciardi, W., & Specchia, M. L. (2016). The impact of public reporting on clinical outcomes: a systematic review and meta-analysis. *BMC Health Services Research*, *16*(1), 1-14.

Cardoen, B., Demeulemeester, E., & Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, *201*(3), 921-932.

Chen, Y. (2011). Why are health care report cards so bad (good)?. *Journal of Health Economics*, *30*(3), 575-590.

Chen, Y., & Meinecke, J. (2012). Do healthcare report cards cause providers to select patients and raise quality of care?. *Health Economics*, 21, 33-55.

Chou, S. Y., Deily, M. E., Li, S., & Lu, Y. (2014). Competition and the impact of online hospital report cards. *Journal of Health Economics*, 34, 42-58.

Coulter, I. D., Wilkes, M., & Der-Martirosian, C. (2007). Altruism revisited: a comparison of medical, law and business students' altruistic attitudes. *Medical Education*, *41*(4), 341-345.

Darwish, A., Mehta, P., Mahmoud, A., El-Sergany, A., & Culberson, D. (2016). Improving operating room start times in a community teaching hospital. *Journal of Hospital Administration*, *5*(3), 33.

Dexter F. Epstein R (2009). Typical savings from each minute reduction in tardy first case of the day starts. *Anesthesia & Analgesia*. 108. 1262-1267.

Does, R. J., Vermaat, T. M., Verver, J. P., Bisgaard, S., & Van Den Heuvel, J. (2009). Reducing start time delays in operating rooms. *Journal of Quality Technology*, *41*(1), 95-109.

Dranove, D., Kessler, D., McClellan, M., & Satterthwaite, M. (2003). Is more information better? The effects of "report cards" on health care providers. *Journal of Political Economy*, 111(3), 555-588.

Dranove, D., & Sfekas, A. (2008). Start spreading the news: a structural estimate of the effects of New York hospital report cards. *Journal of health economics*, *27*(5), 1201-1207.

Emmert, M., Hessemer, S., Meszmer, N., & Sander, U. (2014). Do German hospital report cards have the potential to improve the quality of care?. *Health Policy*, *118*(3), 386-395.

Emmert, M., & Schlesinger, M. (2017). Patients' awareness, usage and impact of hospital report cards in the US. *The Patient-Patient-Centered Outcomes Research*, *10*(6), 729-738.

Emmert, M., & Wiener, M. (2017). What factors determine the intention to use hospital report cards? The perspectives of users and non-users. *Patient education and counseling*, *100*(7), 1394-1401.

Elflein, H. M., Fresenius, S., Lamparter, J., Pitz, S., Pfeiffer, N., Binder, H., ... & Mirshahi, A. (2015). The prevalence of amblyopia in Germany: data from the prospective, population-based Gutenberg Health Study. *Deutsches Ärzteblatt International*, *112*(19), 338.

Elflein, H. M., & Pitz, S. (2015). Amblyopie. *Monatsschrift Kinderheilkunde*, *163*(3), 223-229.

Ellis, R. P., & McGuire, T. G. (1993). Supply-side and demand-side cost sharing in health care. *Journal of Economic Perspectives*, *7*(4), 135-151.

Ellis, R. P., & McGuire, T. G. (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of health economics*, *5*(2), 129-151.

Ernst, C., Szczesny, A., Soderstrom, N., Siegmund, F., & Schleppers, A. (2012). Success of commonly used operating room management tools in reducing tardiness of first case of the day starts: evidence from German hospitals. *Anesthesia & Analgesia*, *115*(3), 671-677.

Faghihi, M., Hashemi, H., Nabovati, P., Saatchi, M., Yekta, A., Rafati, S., ... & Khabazkhoob, M. (2017). The prevalence of amblyopia and its determinants in a population-based study. *Strabismus*, *25*(4), 176-183.

Fischer, C., Lingsma, H. F., Marang-van de Mheen, P. J., Kringos, D. S., Klazinga, N. S., & Steyerberg, E. W. (2014). Is the readmission rate a valid quality indicator? A review of the evidence. *PloS one*, *9*(11), e112282.

Flynn, J. T., Woodruff, G., Thompson, J. R., Hiscox, F., Feuer, W., Schiffman, J., ... & Smith, L. K. (1999). The therapy of amblyopia: an analysis comparing the results of amblyopia therapy utilizing two pooled data sets. *Transactions of the American Ophthalmological Society*, *97*, 373.

Flynn, J. T., Schiffman, J., Feuer, W., & Corona, A. (1998). The therapy of amblyopia: an analysis of the results of amblyopia therapy utilizing the pooled data of published studies. *Transactions of the American Ophthalmological Society*, *96*, 431.

Fronius, M., Cirina, L., Ackermann, H., Kohnen, T., & Diehl, C. M. (2014). Efficiency of electronically monitored amblyopia treatment between 5 and 16 years of age: new insight into declining susceptibility of the visual system. *Vision Research*, *103*, 11-19.

Fu, J., Li, S. M., Liu, L. R., Li, J. L., Li, S. Y., Zhu, B. D., ... & Anyang Childhood Eye Study Group. (2014). Prevalence of amblyopia and strabismus in a population of 7th-grade junior high school students in Central China: the Anyang Childhood Eye Study (ACES). *Ophthalmic epidemiology*, *21*(3), 197-203.

Ganekal, S., Jhanji, V., Liang, Y., & Dorairaj, S. (2013). Prevalence and etiology of amblyopia in Southern India: results from screening of school children aged 5–15 years. *Ophthalmic epidemiology*, *20*(4), 228-231.

Geldner, G., Eberhart, L. H., Trunk, S., Dahmen, K. G., Reissmann, T., Weiler, T., & Bach, A. (2002). Efficient OP management. Suggestions for optimisation of organisation and administration as a basis for establishing statutes for operating theatres. *Der Anaesthesist*, *51*(9), 760-767.

Gfrörer, R., Schüpfer, G., Schmidt, C. E., & Bauer, M. (2005). Teamwork in the operating theatre. Effect on quality of decision-making. *Der Anaesthesist*, *54*(12), 1229-1234.

Glazier, R. H., Klein-Geltink, J., Kopp, A., & Sibley, L. M. (2009). Capitation and enhanced fee-for-service models for primary care reform: a population-based evaluation. *CMAJ*, *180*(11), E72-E81.

Godager, G., & Wiesen, D. (2013). Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics*, *32*(6), 1105-1116.

Gourevitch, R. A., Mehrotra, A., Galvin, G., Plough, A. C., & Shah, N. T. (2019). Does comparing cesarean delivery rates influence women's choice of obstetric hospital?. *The American Journal of Managed Care*, *25*(2), e33.

Gravelle, H., & Sivey, P. (2009). *Imperfect quality information in a quality-competitive hospital market*. Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Gravelle, H., & Sivey, P. (2010). Imperfect information in a quality-competitive hospital market. *Journal of Health Economics*, *29*(4), 524-535.

Halfon, P., Eggli, Y., Pêtre-Rohrbach, I., Meylan, D., Marazzi, A., & Burnand, B. (2006). Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. *Medical Care*, 972-981.

Hashemi, H., Pakzad, R., Yekta, A., Bostamzad, P., Aghamirsalim, M., Sardari, S., ... & Khabazkhoob, M. (2018). Global and regional estimates of prevalence of amblyopia: A systematic review and meta-analysis. *Strabismus*, *26*(4), 168-183.

Hassing, L. B., Grant, M. D., Hofer, S. M., Pedersen, N. L., Nilsson, S. E., Berg, S., ... & Johansson, B. (2004). Type 2 diabetes mellitus contributes to cognitive decline in old age: a longitudinal population-based study. *Journal of the International Neuropsychological Society*, *10*(4), 599-607.

Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation. *Journal of Health Economics*, *30*(4), 637-646.

Hennig-Schmidt, H., & Wiesen, D. (2014). Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students. *Social Science & Medicine*, *108*, 156-165.

Hotelling, H. (1929). Stability in competition. *Economic Journal*, *153*(39), 41-57.

Hojat, M., Gonnella, J. S., Nasca, T. J., Mangione, S., Vergare, M., & Magee, M. (2002). Physician empathy: definition, components, measurement, and relationship to gender and specialty. *American Journal of Psychiatry*, *159*(9), 1563-1569.

Howell, E. A., & Zeitlin, J. (2017). Improving hospital quality to reduce disparities in severe maternal morbidity and mortality. *Seminars in Perinatology* , *41*(5), 266-272.

Iversen, T., & Lurås, H. (2000). The effect of capitation on GPs' referral decisions. *Health Economics*, *9*(3), 199-210.

Jha, A. K., Orav, E. J., Li, Z., & Epstein, A. M. (2007). The inverse relationship between mortality rates and performance in the Hospital Quality Alliance measures. *Health Affairs*, *26*(4), 1104-1110.

Klora, M., Zeidler, J., May, M., Raabe, N., & von der Schulenburg, J. M. G. (2017). Evaluation of family doctor-centred health care in Germany based on AOK Rheinland/Hamburg claims data. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, *120*, 21-30.

Lagrèze, W. A. (2010). Vision screening in preschool children: do the data support universal screening?. *Deutsches Aerzteblatt International*, *107*(28-29), 495.

Li, J. (2018). Plastic surgery or primary care? Altruistic preferences and expected specialty choice of US medical students. *Journal of Health Economics*, *62*, 45-59.

Ma, C. T. A. (1994). Health care payment systems: cost and quality incentives. *Journal of Economics & Management Strategy*, *3*(1), 93-112.

Macario, A. (2010). What does one minute of operating room time cost?. *Journal of Clinical Anesthesia*, *4*(22), 233-236.

Marć, M., Bartosiewicz, A., Burzyńska, J., Chmiel, Z., & Januszewicz, P. (2019). A nursing shortage–a prospect of global and local policies. *International Nursing Review*, *66*(1), 9-16.

Marcon, E., & Dexter, F. (2006). Impact of surgical sequencing on post anesthesia care unit staffing. *Health Care Management Science*, *9*(1), 87-98.

Marešová, P., Mohelská, H., & Kuča, K. (2015). Economics aspects of ageing population. *Procedia Economics and Finance*, *23*, 534-538.

Matta, N. S., Singman, E. L., & Silbert, D. I. (2008). Performance of the Plusoptix vision screener for the detection of amblyopia risk factors in children. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *12*(5), 490-492.

McIntosh, C., Dexter, F., & Epstein, R. H. (2006). The impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: a tutorial using data from an Australian hospital. *Anesthesia & Analgesia*, *103*(6), 1499-1516.

Merchant, R. M., Yang, L., Becker, L. B., Berg, R. A., Nadkarni, V., Nichol, G., ... & Groeneveld, P. W. (2012). Variability in case-mix adjusted in-hospital cardiac arrest rates. *Medical Care*, *50*(2), 124.

Moghaddam, A. A. S., Kargozar, A., Zarei-Ghanavati, M., Najjaran, M., Nozari, V., & Shakeri, M. T. (2012). Screening for amblyopia risk factors in pre-verbal children using the Plusoptix photoscreener: a cross-sectional population-based study. *British Journal of Ophthalmology*, *96*(1), 83-86.

Newhouse, J. P. (1970). Toward a theory of nonprofit institutions: An economic model of a hospital. *The American Economic Review*, *60*(1), 64-74.

Overdyk, F. J., Harvey, S. C., Fishman, R. L., & Shippey, F. (1998). Successful strategies for improving operating room efficiency at academic institutions. *Anesthesia & Analgesia*, *86*(4), 896-906.

Oscar, A., Cherninkova, S., Haykin, V., Aroyo, A., Levi, A., Marinov, N., ... & Spitzer, J. (2014). Amblyopia screening in Bulgaria. *Journal of Pediatric Ophthalmology & Strabismus*, *51*(5), 284-288.

Park, K. W., & Dickerson, C. (2009). Can efficient supply management in the operating room save millions?. *Current Opinion in Anesthesiology*, *22*(2), 242-248.

Pauly, M. V. (1995). Paying Physicians as Agents: Fee-for-Service, Caption, or Hybrids?. *Health Care Policy and Regulation* (pp. 163-180). Springer, Boston, MA.

Pfannstiel, M. A. (2014). Steigerung der Dienstleistungsproduktivität im OP-Bereich von Krankenhäusern. *Dienstleistungsmanagement im Krankenhaus II* (pp. 229-247). Springer Gabler, Wiesbaden.

Plusoptix GmbH (2021). Product information. Last time accessed: 29.10.2021 via https://plusoptix.com/vision-screener/landing-google-eu?pk_campaign=GoogleAds_EU&gclid=Cj0KCQjwt-6LBhDlARIsAIPRQcIBJGHqdIxS2S5OU3xqn2ePMTisifryUii84k1ci6JTpxImzyAxOsUaAhaeEALw_wcB .

Prang, K. H., Maritz, R., Sabanovic, H., Dunt, D., & Kelaher, M. (2021). Mechanisms and impact of public reporting on physicians and hospitals' performance: A systematic review (2000–2020). *PloS one*, *16*(2), e0247297.

Raphael, I. J., Parmar, M., Mehrganpour, N., Sharkey, P. F., & Parvizi, J. (2013). Obesity and operative time in primary total joint arthroplasty. *The Journal of Knee Surgery*, *26*(02), 095-100.

Scalea, T. M., Carco, D., Reece, M., Fouche, Y. L., Pollak, A. N., & Nagarkatti, S. S. (2014). Effect of a novel financial incentive program on operating room efficiency. *JAMA surgery*, *149*(9), 920-924.

Schiff, R. L., & Emanuele, M. A. (1995). The surgical patient with diabetes mellitus. *Journal of General Internal Medicine*, *10*(3), 154-161.

Schuster, M., Pezzella, M., Taube, C., Bialas, E., Diemer, M., & Bauer, M. (2013). Delays in starting morning operating lists: an analysis of more than 20 000 cases in 22 German hospitals. *Deutsches Aerzteblatt International*, *110*(14), 237.

Schuster, M., Wicha, L. L., & Fiege, M. (2007). Key performance indicators of OR efficiency. Myths and evidence of key performance indicators in OR management. *Der Anaesthesist*, *56*(3), 259-271.

Schneider, E. C., & Epstein, A. M. (1996). Influence of cardiac-surgery performance reports on referral practices and access to care—a survey of cardiovascular specialists. *New England Journal of Medicine*, *335*(4), 251-256.

Sinaiko, A. D., Eastman, D., & Rosenthal, M. B. (2012). How report cards on physicians, physician groups, and hospitals can have greater impact on consumer choices. *Health Affairs*, *31*(3), 602-611.

Singman, E., Matta, N., Tian, J., Brubaker, A., & Silbert, D. (2013). A comparison of the PlusoptiX S04 and A09 photoscreeners. *Strabismus*, *21*(2), 85-87.

Shahian, D. M., Meyer, G. S., Mort, E., Atamian, S., Liu, X., Karson, A. S., ... & Zheng, H. (2012). Association of National Hospital Quality Measure adherence with long-term mortality and readmissions. *BMJ Quality & Safety*, *21*(4), 325-336.

Sorenson, C., Drummond, M., & Khan, B. B. (2013). Medical technology as a key driver of rising health expenditure: disentangling the relationship. *ClinicoEconomics and Outcomes Research: CEOR*, *5*, 223.

Jacques, P. J. S., Patel, N., & Higgins, M. S. (2004). Improving anesthesiologist performance through profiling and incentives. *Journal of Clinical Anesthesia*, *16*(7), 523-528.

Szczesny, A., & Ernst, C. (2016). The role of performance reporting system characteristics for the coordination of high-cost areas in hospitals. *European Accounting Review*, *25*(4), 635-660.

Truong, A., Tessler, M. J., Kleiman, S. J., & Bensimon, M. (1996). Late operating room starts: experience with an education trial. *Canadian Journal of Anaesthesia*, *43*(12), 1233.

Tsai, T. C., Joynt, K. E., Orav, E. J., Gawande, A. A., & Jha, A. K. (2013). Variation in surgical-readmission rates and quality of hospital care. *New England Journal of Medicine*, *369*(12), 1134-1142.

Turner, T., Cull, W. L., Bayldon, B., Klass, P., Sanders, L. M., Frintner, M. P., ... & Dreyer, B. (2009). Pediatricians and health literacy: descriptive results from a national survey. *Pediatrics*, *124*(Supplement 3), S299-S305.

Unger, J., Schuster, M., Bauer, K., Krieg, H., Müller, R., & Spies, C. (2009). Time delay in beginning first OR positions in the morning. *Der Anaesthesist*, *58*(3), 293-8.

Vallance, A. E., Fearnhead, N. S., Kuryba, A., Hill, J., Maxwell-Armstrong, C., Braun, M., ... & Walker, K. (2018). Effect of public reporting of surgeons' outcomes on patient selection,"gaming," and mortality in colorectal cancer surgery in England: population based cohort study. *BMJ*, *361*.

Van Veen-Berkx, E., Elkhuizen, S. G., Kalkman, C. J., Buhre, W. F., & Kazemier, G. (2014). Successful interventions to reduce first-case tardiness in Dutch university medical centers: results of a nationwide operating room benchmark study. *The American Journal of Surgery*, *207*(6), 949-959.

Van Leeuwen, R., Eijkemans, M. J., Vingerling, J. R., Hofman, A., de Jong, P. T., & Simonsz, H. J. (2007). Risk of bilateral visual impairment in individuals with amblyopia: the Rotterdam study. *British Journal of Ophthalmology*, *91*(11), 1450-1451.

Varkevisser, M., van der Geest, S. A., & Schut, F. T. (2012). Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands. *Journal of Health Economics*, *31*(2), 371-378.

Vitez, T. S., & Alex Macario, M. D. (1998). Setting performance standards for an anesthesia department. *Journal of Clinical Anesthesia*, *10*(2), 166-175.

Williams, C., Northstone, K., Harrad, R. A., Sparrow, J. M., & Harvey, I. (2002). Amblyopia treatment outcomes after screening before or at age 3 years: follow up from randomised trial. *BMJ*, *324*(7353), 1549.

Williams, C., Northstone, K., Harrad, R. A., Sparrow, J. M., & Harvey, I. (2003). Amblyopia treatment outcomes after preschool screening v school entry screening: observational data from a prospective cohort study. *British Journal of Ophthalmology*, *87*(8), 988-993.

Wong, J., Khu, K. J., Kaderali, Z., & Bernstein, M. (2010). Delays in the operating room: signs of an imperfect system. *Canadian Journal of Surgery*, *53*(3), 189.

Wooldridge, J.M.: Introductory econometrics. A modern approach. South-Western Cengage Learning, Mason (2013).

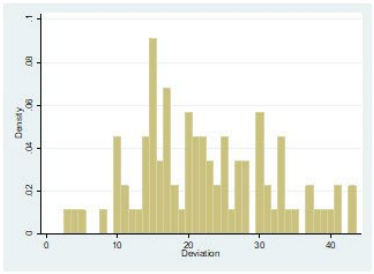
Wright, J. G., Roche, A., & Khoury, A. E. (2010). Improving on-time surgical starts in an operating room. *Canadian Journal of Surgery*, *53*(3), 167.

Yan, X. R., Jiao, W. Z., Li, Z. W., Xu, W. W., Li, F. J., & Wang, L. H. (2015). Performance of the Plusoptix A09 photoscreener in detecting amblyopia risk factors in Chinese children attending an eye clinic. *PLoS One*, *10*(6), e0126052

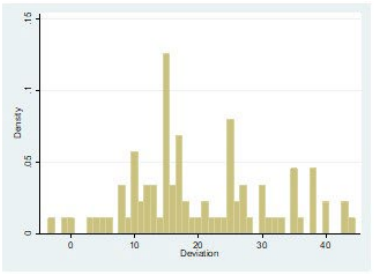
7 Appendix

Appendix I:

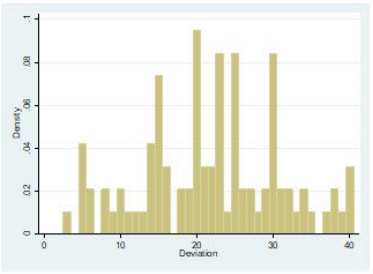
Figure 7.1: Distribution of deviations from targeted incision time for half-year intervals– Site A



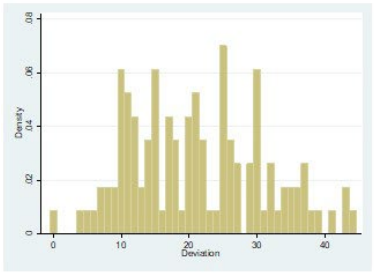
Half year 1



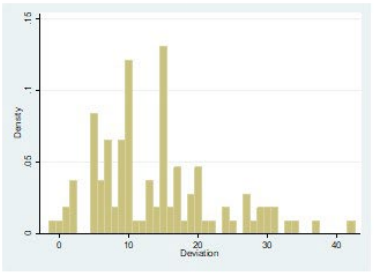
Half year 2



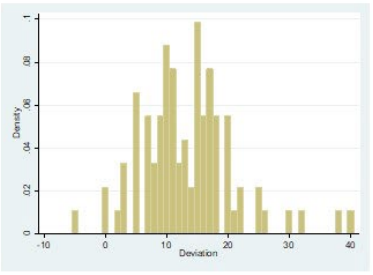
Half year 3



Half year 4

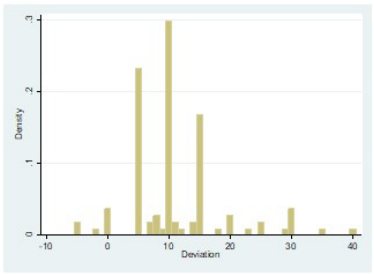


Half year 5

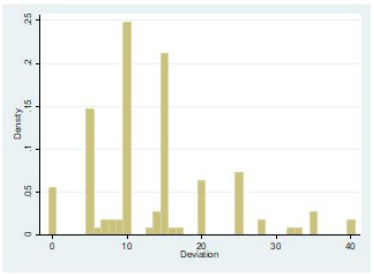


Half year 6

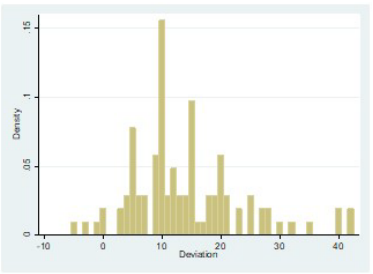
Figure 7.2: Distribution of deviations from targeted incision time for half-year intervals– Site B



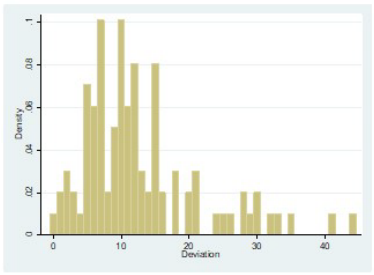
Half year 1



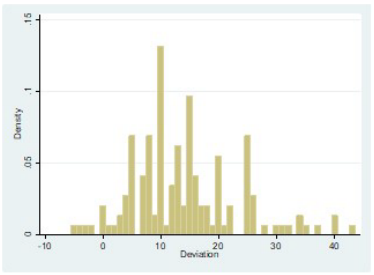
Half year 2



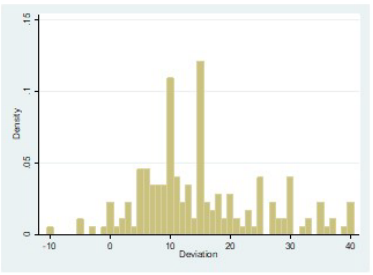
Half year 3



Half year 4



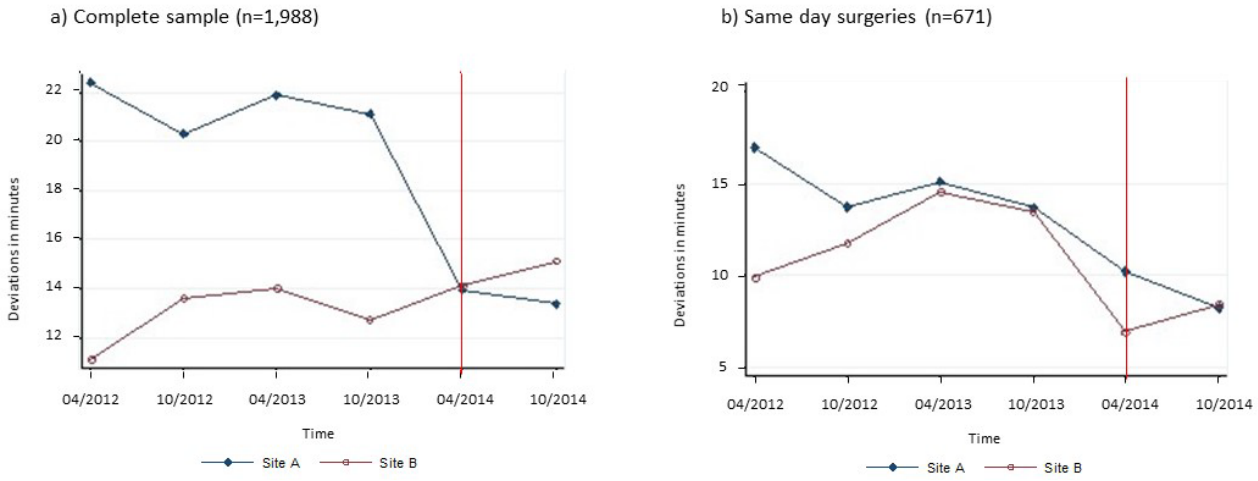
Half year 5



Half year 6

Appendix II:

Figure 7.3: Time trends deviation from targeted incision time from April 2012 to February 2015 measured in half-year time intervals (where the sixth (final) interval encompasses only five months due to data structure)



Appendix III: Placebo treatments

In order to assess the validity of the common trend assumption, we add pre- and post-intervention interaction terms between the treated hospital (site A) and half-year dummies to estimation equation Eq. (1). Thereby we simulate placebo treatments for site A at different points in time. If we cannot detect significant differences between both hospital sites in the pre-treatment periods, credibility of the parallel trend assumption increases. Thus, if all estimated coefficients for pre-treatment placebo interaction terms do not significantly deviate from zero, the validity of the common trend assumption becomes more likely. Furthermore, the coefficients for post-intervention placebo treatments are also of interest to gain further insights on how sustainable possible effects of OR charter introduction are. To create placebo treatments we introduce leads and lags of the treatment to the basic difference-in-difference analysis (estimation equation Eq. (1)). Then, the model we are going to estimate becomes:

$$y_{itn} = \beta_0 + \beta_2 SiteA_{itn} + \delta_t + \sum_{t=1}^{6} \varphi_t \delta_t SiteA_{itn} + \beta_3 ward_{itn} + \beta_4 pain_{itn} + X_{itn}\gamma + u_{itn} \quad (3)$$

Here, notation corresponds to estimation equation Eq.(1), but instead of only having the actual treatment term $post_{itn} * SiteA_{itn}$ we introduce interaction terms between treated site and all half-year dummies $\sum_{t=1}^{6} \varphi_t \delta_t SiteA$, where half year dummy $t = 4$ (which is the last half year dummy before OR charter is introduced in $t = 5$) serves as reference category. In Table 7.1, we see that all estimates for the pre-treatment interaction term coefficients are insignificant and the coefficients for the two post-treatment periods are highly significant. Thus, the results support the validity of the parallel pre-trend assumption.

Table 7.1: Placebo effects

	Dependent variable
	Dependent variable: Deviation from scheduled incision time
Age65	-0.006
	(0.52)
Female	0.763
	(0.47)
Obesity	1.552*
	(0.66)
Diabetes	1.931**
	(0.71)
Surgtime	0.068***
	(0.01)
Ward	3.255***
	(0.65)
Pain	5.082***
	(0.56)
Half year dummies	Yes
SiteA	8.041***
	(1.13)
Half Year 1×Site A	2.346
	(1.59)
Half Year 2×Site A	-1.805
	(1.69)
Half Year 3×Site A	-0.653
	(1.63)
Half Year 5×Site A	-7.637***
	(1.54)
Half Year 6×Site A	-8.909***
	(1.57)
Constant	4.579***
	(1.03)
N	1317
r2	0.304

Robust standard errors are in parentheses. * p<0.05, ** p<0.01, *** p<0.001

Appendix IV. Linear time-trends

To assess the robustness of the parallel trend assumption further, we test for linear time trends by adding site-specific linear trends to estimation equation (1). The aim is to control for unobservable site-specific time trends driven by factors independent from actual OR charter introduction. The estimation equation then becomes:

$$y_{itn} = \beta_0 + \beta_1 SiteA_{itn} + \beta_2 SiteA_{itn} * post_{it} + \beta_3 ward_{itn} + \beta_4 pain_{itn} + \beta_5 SiteA_{itn} * t + \delta_t + X_{itn}\gamma + u_{itn} \quad (4)$$

Again, δ_t denotes half year fixed effects dummies, i.e. general shocks occurring over time affecting delays in site A and site B to the same extent. We added the interaction term $\beta_5 SiteA_{itn} * t$ to our original difference-in-difference models in Eq. (1). Here, t measures time (in half-year intervals) as a continuous variable and coefficient β_5 reflects the site-specific linear time trend. As Table 7.2 shows, the coefficient estimate indicates a decreasing linear time trend for deviations. We find no significant linear time trend. This finding indicates that for non-same day surgeries no linear time trends are detectable, supporting the validity of the parallel trend assumption.

Our additional analyses supports the credibility of parallel pre-treatment development in the outcome variable, as we are able to reject the null hypothesis of systematic differences between both sites before treatment (placebo treatment dummies) and of linear time trends. This indicates validity of estimation results derived for orthopedic non-same day surgeries.

Table 7.2: Linear time trends

	Dependent variable
	Deviation from scheduled incision time
Treatment	-6.345***
	(1.72)
ge65	0.011
	(0.52)
Female	0.753
	(0.47)
Obesity	1.641*
	(0.66)
Diabetes	1.894**
	(0.71)
Surgtime	0.068***
	(0.01)
Ward	3.251***
	(0.65)
Pain	5.096***
	(0.56)
SiteA	9.625***
	(1.36)
SiteA × Half year	-0.640
	(0.48)
Constant	4.114***
	(0.99)
Half year dummies	Yes
r2	0.300
N	1317
Robust standard errors are in parentheses. * p<0.05, ** p<0.01, *** p<0.001	

p