Learning Curve Effects in Hospitals as Highly Specialized Expert Organizations

Dissertationsschrift
zur Erlangung des Grades
eines Doktors der Wirtschaftwissenschaften (Dr. rer. pol.)
an der Wirtschaftswissenschaftlichen Fakultät
der Universität Würzburg

vorgelegt von
Carsten Bauer

2022

Erstgutachterin: Prof. Dr. Andrea Szczesny

Zweitgutachter: Prof. Dr. Martin Kukuk

Acknowledgments

I wrote the majority of this dissertation during my employment as a research and teaching assistant at the Chair of Management Accounting and Control at University of Würzburg. I am grateful to Prof. Dr. Andrea Szczesny for giving me this opportunity and the freedom to work on my research projects as well as for her support when it was time to wrap things up. I want to thank Prof. Dr. Martin Kukuk who instantly agreed on becoming my second supervisor and provided me with valuable suggestions and comments regarding econometrics. I am grateful to my coauthors Nele Möbs, Dr. Oliver Unger, Dr. med. Johannes Taeger, Prof. Dr. Andrea Szczesny, Prof. Dr. Christian Ernst, Prof. Dr. med. Dr. Martin Holderried, and PD Dr. med. Kristen Rak. I also appreciate the support by PD Dr. Dr. Stefan Schenk, Dr. Tamara Schamberger, Dr. Manuel Rademaker, Dr. med. univ. Tobias Spingler, Jun.-Prof. Dr. Daniel Schaupp, Hanna Jenzen, and Mareike Seeger. I also thank my colleagues from the Faculty of Medicine, especially Prof. Dr. Rüdiger Pryss, as well as the colleagues from the university hospital Würzburg, first and foremost Ralph Keim. I would also like to thank my colleagues at the Chair of Management Accounting and Control, Lorenz Heil, Prof. Dr. Sven Hörner, Maya Michels, Felix Otto, and Simone Wittmann. Finally, I am very grateful to my family and friends for their support, especially in challenging phases of this dissertation. Without them, the accomplishment of this project would not have been possible.

Carsten Bauer
Würzburg, September 2022

Contents

List of Tables

List of Figures

Zusammenfassung (German Summary)

Die vorliegende Arbeit befasst sich mit Lernkurveneffekten in Kliniken als hochspezialisierte Expertenorganisationen und umfasst vier Aufsätze, die sich jeweils mit einem anderen Aspekt des Themas beschäftigen. Der Fokus liegt dabei in den ersten drei Artikeln auf Lernkurveneffekten von Ärzten bei chirurgischen Eingriffen, im vierten Beitrag steht das Team der Notaufnahme als Ganzes im Fokus.

Das Vorwort beleuchtet in kompakter Form die stetig steigenden Gesundheitsausgaben, den damit steigenden Kostendruck, die Krankenhauslandschaft in Deutschland sowie ihre Entwicklung. Ferner wird das Vergütungssystem mittels Fallpauschalen wie auch die Besonderheiten des Gesundheitssektors, der stark staatlich reguliert ist und in dem ethische Aspekte omnipräsent sein müssen, umrissen. Nicht zuletzt wird der Nutzen des Wissens um Lernkurveneffekte skizziert, um Kosten zu senken und zudem die Qualität konstant zu halten oder sogar zu verbessern.

Der erste Aufsatz des Sammelbandes untersucht die Lerneffekte in einer Klinik, die sich auf endoprothetische Eingriffe (Hüft- und Kniegelenksersatz) spezialisiert hat. Es werden dabei zwei große Bereiche untersucht, der spezialisierte sowie der nicht-spezialisierte Bereich. Es werden keine Kosten direkt untersucht, sondern Kostenindikatoren verwendet. Der Indikator für die Kosten in der kurzen Frist sind die OP-Zeiten. Der Indikator für die mittlere bis lange Frist ist die Qualität. Sie wird über im Aufwachraum möglicherweise auftretende Komplikationen operationalisiert. Die Untersuchung selbst erfolgt mit Regressionsmodellen (KQ und Logit). Die Ergebnisse zeigen, dass die Spezialisierung Vorteile durch Lerneffekte in Form von kürzeren OP-Zeiten und geringeren Komplikationsraten bei den endoprothetischen Eingriffen mit sich bringt. Für die nicht-spezialisierten OPs ergeben sich dieselben Resultate. Auch hier sinken die OP-Zeiten sowie die Komplikationen. Es kommt demnach nicht zu möglicherweise negativen Auswirkungen der Spezialisierung auf den nicht-spezialisierten Bereich, sondern es liegen vorteilhafte Spillover-Effekte vor. Insgesamt ist die Spezialisierung damit als sinnvoll zu betrachten, da sie die kurz-, mittel- und langfristigen Kosten für alle Operationen senkt. Die Autoren sind Carsten Bauer, Nele Möbs, Oliver Unger, Andrea Szczesny und Christian Ernst.

Im zweiten Artikel stehen die Unterschiede zwischen den Lernkurveneffekten von Ärzten, die teilweise im Team operieren und denen, die stets alleine operieren, im Vordergrund. Die Untersuchung verbindet somit Lernkurveneffekte mit Teamüberlegungen, welche in den letzten Jahren zunehmend diskutiert werden. Bei den betrachteten Eingriffen handelt es sich um Tonsillektomien (Entfernung der Gaumenmandeln), einen Standardeingriff. Die Indikatoren für die Kosten in der kurzen bzw. mittleren bis langen Frist sind auch hier die OP-Zeiten der Operationen bzw. Komplikationen als Maß für die Qualität. Die Komplikationen sind in diesem Fall Nachblutungen, die meist wenige Tage nach der Operation auftreten. Die Untersuchung selbst erfolgt mit Regressionsmodellen (KQ und Logit). Die Ergebnisse der Untersuchung zeigen, dass die OP-Zeiten mit zunehmender Erfahrung der Ärzte sinken. Ärzte, die auch im Team arbeiten, lernen dabei schneller als Ärzte, die stets alleine operieren. So sind die OP-Zeiten der Ärzte, die auch im Team arbeiten, geringer als die der anderen. Als Besonderheit stehen im verwendeten Datensatz die Fallkosten zur Verfügung, womit die Kostenindikatoren validiert werden können. Aus den Ergebnissen ergibt sich die Empfehlung, dass Assistenzärzte an Team-Operationen teilnehmen sollten. Die Autoren sind Carsten Bauer, Oliver Unger und Martin Holderried.

Der dritte Aufsatz widmet sich der Stapesplastik, mithilfe derer Schallleitungsschwerhörigkeit aufgrund von Otosklerose (überschießende Knochenbildung) behandelt werden soll. Die Eingriffe werden als ideale Untersuchungsmöglichkeit von Lernkurveneffekten in der Chirurgie angesehen, da sie konzeptionell einfach, jedoch technisch schwierig sind. Es wird eine möglichst umfassende Betrachtung angestrebt, indem die OP-Zeiten als kurzfristiger und die Qualität als mittel- bis langfristiger Kostenindikator herangezogen werden. Um Qualität zu operationalisieren, wird die postoperative Differenz zwischen Luft- und Knochenleitungsschwelle sowie die Kombination aus dieser Differenz mit der Abwesenheit von Komplikationen verwendet. Auch hier werden verschiedene Regressionsmodelle (KQ und Logit) geschätzt. In diesem Artikel wird neben der Klinikebene auch die Individualebene betrachtet, d.h. OP-Zeiten und Qualität für einzelne Ärzte untersucht, was den Vergleich individueller Lernkurven verbessert, da die Ärzte alle unter weitgehend identischen Bedingungen gearbeitet haben. Es zeigt sich, dass die OP-Zeiten mit zunehmender Erfahrung zunächst sinken. Der marginale Effekt von weiterer Erfahrung wird dabei mit zunehmender Erfahrung geringer bis sich die Richtung des Effektes ändert und die OP-Zeiten mit weiter zunehmender Erfahrung – vermutlich aufgrund der Allokation der schwierigeren Fälle auf die erfahrensten Ärzte – wieder steigen. Bezüglich Qualität sind keine Lernkurveneffekte feststellbar. Die Autoren sind Carsten Bauer, Johannes Taeger und Kristen Rak.

Der vierte Beitrag ist ein systematischer Literaturüberblick zu Lerneffekten bei der Behandlung von ischämischen Schlaganfällen. Bei einem Schlaganfall zählt jede Minute, weswegen die inhärente Notwendigkeit besteht, die Dauer vom Auftreten der Symptome bis zur Behandlung zu verkürzen. Der Artikel befasst sich mit der Verkürzung der Dauer vom Eintreffen der Patienten im Krankenhaus bis zur Behandlung mittels Thrombolyse, der sogenannten „Door-to-Needle Time". In der Literatur gibt es hierzu Untersuchungen von Lernen im weiteren Sinne durch ein Qualitätsverbesserungsprogramm und Lernen im engeren Sinne, bei dem Lernkurveneffekte evaluiert werden. Daneben werden Studien ausgewertet, die sich mit den unterschiedlichen Zeiten zwischen Krankenhäusern mit niedrigen und hohen Fallzahlen befassen, da diese Unterschiede wahrscheinlich das Ergebnis von Lernen und Skaleneffekten sind. Nahezu alle der 165 ausgewerteten Artikel berichten von Verbesserungen bezüglich der Dauer bis zur Behandlung. Zudem unterstreichen die klinischen Ergebnisse die gängige Auffassung, dass eine kürzere Zeit vom Eintreffen im Krankenhaus bis zur Behandlung mit einem besseren Ergebnis einhergeht. Der Literaturüberblick diskutiert zudem die ökonomischen Implikationen der Ergebnisse. Der Autor ist Carsten Bauer.

Im Nachwort kommt u.a. zur Sprache, dass für die Nutzung der Lernkurveneffekte zur Effizienzsteigerung nach der Messung der Lerneffekte weitere Anstrengungen unternommen werden müssen, da die Thematik keine einfachen, standardisierten Lösungen zulässt. Zudem wird die Bedeutung der Mehrperspektivität in der Forschung für das Behandlungsergebnis des Patienten, das Gesundheitssystem und die Gesellschaft hervorgehoben.

Preface

Health care costs have been increasing for decades. For example, German health expenditures as well as hospital costs increased by about 80 % just from 2000 to 2018. Health expenditures accounted for 11.7 % of the gross domestic product (GDP) in 2019. In 1970, it had been just 5.7 % (OECD, 2020). Hospitals are a considerable factor in health care; 23.5 % of total German health expenditures were ascribed to them in 2018 (Gesundheitsberichterstattung des Bundes, 2020a,b). Figure P.1 illustrates the development of health expenditures and hospital costs in Germany. With 1,175,000 people, 20.7 % of people in German health care worked in hospitals in 2018 (Gesundheitsberichterstattung des Bundes, 2020d).[1] Highly qualified staff plays a major role in hospitals. It is the crucial

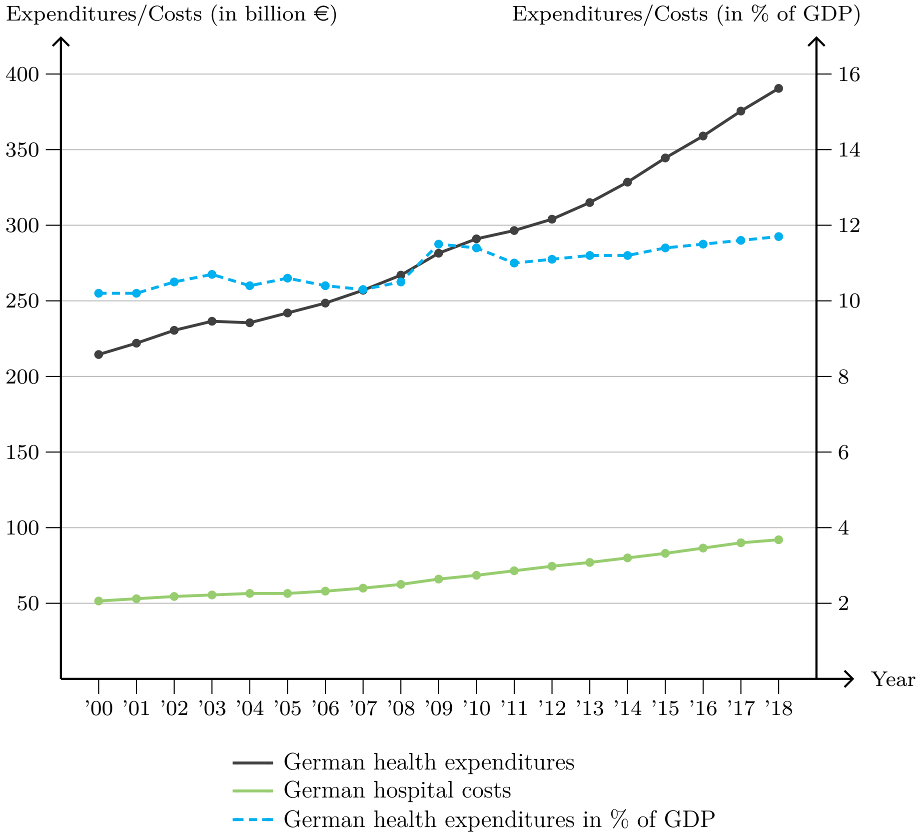

Figure P.1: Development of Health Expenditures and Hospital Costs
(Data: Gesundheitsberichterstattung des Bundes, 2020a,b)

---

[1] Overall, 12.7 % of people in gainful employment work in health care (Gesundheitsberichterstattung des Bundes, 2020d; Statistisches Bundesamt, 2020a).

resource for the quality of the services provided by the hospital and it represents the primary cost factor. Staff costs accounted for 61.1 % of total German hospital costs in 2018 (Statistisches Bundesamt, 2020b).

To get an overview about hospitals in Germany: In 2018, there were 1,925 hospitals, 552 were community-owned, 650 were owned by nonprofit organizations, and 723 were privately owned. Altogether, they provided 498,192 beds and had 19,392,000 patient cases. Community-owned hospitals had 238,907 beds and nonprofit organizations 164,081 beds. 95,204 beds were provided by privately owned hospitals. The number of hospitals as well as the total number of beds they provide has been steadily decreasing in recent times. In 2000, there had still been 2,242 hospitals with 559,651 beds.[2] Though, bed occupancy rate slightly decreased in the same period (0.819 to 0.771), as length of stay did (9.7 to 7.2 days). Furthermore, not all hospitals are equally affected by the structural change. The number of smaller hospitals and the total number of beds they provide decreased, whereas the number of hospitals with more than 500 beds and their total number of beds increased. While the numbers were declining with community-owned and nonprofit hospitals, the number of privately owned hospitals as well as their total number of beds rallied (Gesundheitsberichterstattung des Bundes, 2020c). Besides their number of beds, hospitals can typically be categorized based on the level of care they provide, too. The first level ensures local primary health care. The second level represents specialized medical methods in regional care. The third level stands for maximum care with a comprehensive range of differentiated services, which are provided interregionally.

Along with health care costs, the economic pressure on hospitals has been steadily increasing. Plain reimbursement of costs incurred by the hospital has been gradually replaced by more incentive-based systems. In the early 2000s, the Diagnosis-Related Group (DRG) system was introduced in Germany. This prospective payment system strongly incentivizes hospitals to cut costs, since reimbursement depends on the DRG a case is assigned to and is independent of actual costs incurred by the hospital. Since 2020, reimbursement consists of a combination of a DRG lump-sum compensation and a compensation for nursing staff, i.e. nursing has been separated from the lump-sum compensation and actual costs are reimbursed. Furthermore, payments for DRGs are calculated annually based on past average cost values. This results in a steadily intensifying reimbursement

---

[2] Beds in intensive care units are an exception to the overall trend. The number markedly increased from 23,113 in 2002 (no data available for 2000) by 18.8 % to 27,463 in 2018.

system. In 2019, 44 % of hospitals had an annual shortfall.[3] Smaller hospitals with less than 300 beds had the second highest percentage with 43.8 %, larger hospitals with 600 beds and more the highest one with 54.2 % (Blum et al., 2020).[4] The data thus correspond to the closing of smaller hospitals in recent times. In 46 %, the hospital's annual yield was lower than the year before (Blum et al., 2020).[5] To sum up, cost containment is more important than ever before to ensure the hospital's economic survival.

There are numerous ways to react to an intensified cost pressure: A manifest reaction would be a simple cost containment strategy, this means a smaller staff has to deal with an unchanged number of patients. This strategy almost inevitably comes along with a lower quality of hospital services. Some hospitals might try to get a more "advantageous" patient portfolio by referring "disadvantageous" patients, i.e. patients who will need disproportionately many resources compared to the reimbursement payment to the hospital, to hospitals of higher level of care. Apparently, this "profitability gain" would happen at the expense of the hospitals the patients are referred to. Other hospitals might specialize on certain surgeries or treatments in order to benefit from economies of scale. This strategy will be the starting point for the study in chapter 1. Smaller hospitals might not be able to make use of economies of scale like larger ones are, which might explain the closing of smaller hospitals in recent times. But there are further ways to counter the economic pressure: Hospitals might utilize learning curve effects in order to cut costs while keeping their quality stable or even improving it.[6] When hospitals have detailed information, they can try to adjust processes in order to optimally utilize learning curve effects and be more efficient. For example, the paper in chapter 2 offers some suggestions on how to do this with surgeons working in teams. A decision between costs and quality of services, as it would be the case with a simple cost containment strategy, is not necessary. Though, small hospitals might not be able to make use of learning curve effects like larger ones are, since they probably have considerably less similar cases: For simplicity, it is assumed a small and a larger hospital incur the same costs for a specific surgery and both can reduce their costs to the same extent for each doubling of surgeries. Both hospitals have

---

[3] Numbers based on a representative sample of 438 German hospitals with 100 beds and more. For comparison: In 2018, „only" 40.2 % had an annual shortfall (Blum et al., 2019).

[4] For comparison: In 2018, smaller hospitals had the highest percentage with 43.5 %, larger hospitals the smallest one with 33.3 % (Blum et al., 2019).

[5] For comparison: In 2018, the hospital's annual yield was lower than the year before even in 51.8 % (Blum et al., 2019).

[6] Specialization also benefits from learning curve effects, since specialized hospitals can reach e.g. the next doubling of specialized cases faster, since they just treat more such cases. However, specialization first and foremost relies on economies of scale which make use of a higher number of similar cases within a given period of time.

had one surgery of such kind yet. The cost situation will be getting more disadvantageous for the small hospital over time, since the larger hospital will reach the the next doubling faster than the small one.

While taking into account possible cost containment strategies, one must bear in mind that the health care industry is not a common industry. Medical activities must be oriented toward the patients' health and ethical considerations must be omnipresent. In the past, in a system of plain reimbursement of costs incurred by hospitals, there was little potential of ethical conflicts with regard to economic aspects. However, progressive commodification of health care increasingly poses ethical conflicts. Reimbursement based on DRGs provides different compensations for different diagnoses. This aspects might interfere when making the patient's diagnosis. There might emerge an incentive to "make the patient ill", i.e. the patient is diagnosed with a more serious disease in order to conduct further treatments which are not necessary from a medical perspective based on the "true diagnosis", but profitable. This procedure is called "upcoding". Although economic ways of thinking are not a threat to ethical action per se, they are to be understood rather as a means of an ethically necessary efficient use of available resources in order to maximize the medical services for patients and contributors of health care given finite resources. (Wehkamp, 2002; Staudt, 2020). Another aspect for ethical conflicts in DRG context is the rule of minimum volumes for several DRGs. It is supposed to ensure quality of medical treatments. Hospitals which fall below the required quantity for a specific DRG are not allowed to treat patients with the respective diagnosis anymore.[7] Though, there can originate undesirable incentives for facilities that do not reach the required numbers. They might appear in form of non-indicated treatments and surgeries which lead to higher, unnecessary costs as well as avoidable risks for the patients. To make matters even more complicated, medical decisions often have to be made rapidly. In the papers, these considerations concerning ethics will be addressed only indirectly by checking outcome quality. Doing so, the studies can control for possibly disadvantageous effects of cost containment on patients. If outcome quality is not negatively affected by cost containment measures using learning curve effects, there should not be major ethical problems with these measures.

The health sector is also a special one because it is strongly regulated. The following paragraph is supposed to provide a general idea by giving some examples of major

---

[7] There are some exceptions, e.g. for emergencies or if the transfer to another hospital, which reaches the minimum volume, is not justifiable.

regulations. For example, as already mentioned, German hospitals cannot set prices for their services, but receive DRG compensations that have been determined by a third party. Hospitals are also not able to determine the services they provide completely on their own as their services have to be part of the respective federal state's hospital plan.[8] Another restriction to the services hospitals provide are minimum volumes for several DRGs.[9] Further, there are limitations regarding investments due to the dual hospital financing. With DRG compensations, health insurers are supposed to only bear current costs. Investment costs are incurred by the respective federal state. Thus, hospitals are restricted in their investment opportunities. Hospital staff is also gripped by regulations, e.g. physicians' professional regulations or the ones regarding the qualification of nursing staff.

The collection at hand is concerned with learning curve effects in hospitals as highly specialized expert organizations and comprises four papers, three of them concerned with surgery, but focusing on a different aspect of the topic and one concerned with conservative treatment. The title of the collection emphasizes the relevance of highly qualified and specialized experts in hospitals. The learning curve effects with this expert knowledge is in the center of interest. As staff is the overall crucial hospital resource, learning curve effects might be detected in various groups of staff in various environments within the hospital. Three papers comprised in this collection are concerned with learning curve effects among surgeons in the operating room (in the following "OR" for short). The OR is one of the most resource-utilizing facilities in the hospital, which makes it especially interesting to investigate learning curve effects there. These three papers study German hospitals; however, their insights are not limited to them. The basic idea of these articles is the following: Having gained insights into learning curve effects, hospitals can utilize these to decrease their costs. The first one deals with learning curve and spillover effects in a hospital which has specialized on endoprosthetic surgery. The second paper investigates learning curve effects in a teamwork vs. an individual work setting in the OR. The focus of the third paper in this collection is on learning curve effects in stapes surgery. This kind of surgery is regarded as the optimum to study learning curves in surgery because it is conceptually simple, but technically difficult. The forth paper is a systematic review of research in learning in ischemic stroke treatment which becomes manifest in the reduction of time between arrival at the hospital and treatment.

---

[8] This plan contains the level of care, number of beds, medical disciplines, etc. of each hospital.

[9] See previous paragraph for details.

1 Spillover Effects of Specialization Strategies in Hospitals

Carsten Bauer, Nele Möbs, Oliver Unger, Andrea Szczesny and Christian Ernst[10]

1.1 Introduction

In most developed countries, the economic pressure on hospitals to improve their efficiency has increased considerably over the past 30 years. The once prevalent system of simple cost reimbursement for hospital care has almost universally been superseded by much more high-powered economic incentive-based systems such as capped budget systems or prospective payment systems. These systems offer powerful incentives to reduce expenses, particularly in high-cost areas such as the operating room (OR). Specialization is one prominent response strategy to such pressure. To set valuable decisions for operative and strategic questions, it is as important for hospitals as it is for every company to obtain detailed information about learning effects and the costs they influence. In the health care industry, Clark and Huckman (2012) examine the impact of the degree of specialization of a hospital on its clinical performance in specialized and related areas of surgery and find that greater operational focus has a positive effect on the quality of medical procedures. Ernst and Szczesny (2006) show that the learning effects in specialized cases lead to an improved resource allocation in the OR and, as a result, to lower costs. Although these studies clearly show lower expenditures in high-cost areas such as the OR, there is little evidence regarding how such isolated cost-reduction strategies affect subsequent cost developments.

In particular, the question of whether cost-reduction efforts by specialization in the OR deploy spillover effects is in the focus of interest. In order to offer a comprehensive investigation, the effects of specialization on specialized cases are evaluated, too. This

---

[10] We thank participants of 2017 VHB conference as well as participants of Hohenheim-Tübingen Workshop on OR management for helpful comments and suggestions. We thank Dr. Tamara Schamberger, Dr. Manuel Rademaker, Dr. med. univ. Tobias Spingler and especially PD Dr. Dr. Stefan Schenk for their support.

means the learning effects connected to specialization are also analyzed. It has not been investigated yet whether this cost-saving strategy affects later events in the patient's case or the overall costs of a case. Thus, these spillover effects are examined on a larger scale that considers the entire patient's case, since the cost-reduction efforts may actually be counterproductive because of lower quality of surgeries as well as increased costs in the medium to long term of the patient's case for both specialized and non-specialized cases. Cost-reduction strategies in an isolated area such as the OR might have negative consequences during later treatment, procedures or recovery processes. For example, consider two methods of anesthetizing a patient, where one method is much more expensive than the other one, but leads to considerably fewer problems regarding patient sickness and complications in the post-anesthesia care unit (PACU, formerly known as the recovery room). Clearly, choosing the supposedly less expensive option could lead the hospital to incur much higher costs later on and therefore to overcompensate initial savings. For instance, Beldi et al. (2009) find that hastiness in the OR leads to an increase in postoperative infections, suggesting that such a narrow focus may pose a problem in reality, as the increased follow-up costs due to post-operative infections eliminate any gains from the decreased OR costs attributable to shorter surgery time and increase overall costs per case in the end. It is thus far from obvious whether isolated cost containment measures remain beneficial in a larger perspective that considers the entire patient's case. In summary, the rarely debated issue is whether a more sophisticated specialization-based cost-reduction strategy at one point in the patient's treatment process (here in the OR) with its possible spillover effects can result in higher costs at later points in the patient's case. The entire patient's case in this context denotes a patient's treatment from the surgical intervention to the treatment of possible complications in the PACU.[11]

In this analysis, unique data from the health care industry are used, covering one point in time, 1996, in which economic pressure on hospitals dramatically increased[12], and the focal hospital reacted by becoming specialized in endoprosthetic surgery (total knee and hip replacement). The data stem from a period prior to the introduction of Diagnosis-Related Groups (DRGs) in Germany and therefore allow studying the unique situation of a hospital specialization unaffected by yearly adjustments common under the DRG system. It is also unique in the sense that micro-level hospital data of similar detail to the one at hand are virtually non-existent in Germany for the period prior to 2003/2004.

---

[11] Due to the data set, a broader definition of the patient's case is not possible. See section 1.4.3 for a more detailed discussion of this aspect.

[12] In 1996, there was a switch to a rigidly capped budgets system in Germany.

The time from incision to suture (in the following “OR time” for short) of specialized and non-specialized interventions is used as an indicator of learning (specialized cases) and spillover effects (non-specialized cases) which influence the short-term costs related to the OR. The proxy for the follow-up costs of a patient’s case is the probability of a patient experiencing a complication in the PACU. Because detailed micro-cost data on medium- and long-term quality are generally unavailable, “quality” is used in a short-term design to act as an indicator for costs incurred at later points in the treatment process, i.e. the short-term quality measure serves as a proxy for medium- to long-term follow-up costs. Decreasing surgery quality typically results in higher follow-up costs because of more (intensive) complications or other adverse events, a higher amount of required medication and/or the need for additional procedures and perhaps a readmission.

The remainder of this paper is organized as follows: In section 1.2, the impact of the reimbursement reforms on hospitals is discussed, possible reactions are described, and the hypotheses are developed based on the literature. The data set as well as the research models are described in section 1.3. A discussion of the results and their limitations is presented in section 1.4, followed by some considerations about the robustness of estimations in section 1.5. The paper closes with a short summary and outlines some suggestions for further research in section 1.6.

1.2 Background and Hypotheses

1.2.1 OR Times as a Proxy for Costs

The OR times are used as an indicator of short-term costs. Given shorter OR times, more interventions are feasible within a fixed period of time, which leads to a fixed costs degression. Besides shorter OR times, Ernst and Szczesny (2006) find that there are also labor cost savings related to learning effects as well as an enhanced resource allocation. Cost savings might negatively affect the quality performance. If it were so, cost reductions due to shorter OR times would be attended by lower quality which in turn probably would increase follow-up costs.[13] Due to possible adverse effects of shorter OR times in the medium and long term, the complications as an indicator for medium- to long-term follow-up costs are also investigated. Only taking into consideration both indicators the OR times and the complications, the economic implications of the specialization in the OR on the entire patient’s case can be properly evaluated.

---

[13] For a more detailed discussion of the relation between cost savings and quality, see section 1.2.2.

1.2.2 Complications as a Proxy for Costs

Whereas the OR times are used as an indicator for costs in the short term only, the complications serve as a medium- to long-term cost indicator. The possible effects of a specialization strategy in the OR on follow-up costs can be measured in a variety of ways. Ideally, actual cost data are analyzed, but these data are rarely made available to researchers, particularly in one-site studies such as the one at hand. An alternative is to use veridical proxies for costs, following Clark and Huckman's (2012) use of patient mortality in their article on hospital specialization. In this case, the probability of a patient suffering from a complication in the PACU is employed to measure the quality consequences of a specialization strategy in the OR. This quality proxy thus accomplishes two objectives. First, it can answer the question how the specialization and its effects influence short-term outcomes with regard to quality. Second, it can link these results to the follow-up costs of the patient's case.

Quality is linked to treatment costs because negative outcomes (such as complications or adverse events) are typically associated with higher resource use. There is distinct evidence in the literature that complications are a proxy for higher costs. Kalish et al. (1995) depict an impressive increase in costs of $16,023 if patients experienced complications. They also show that complications extend the length of stay (LOS), which is consistent with the result of higher costs. In line with this, Collins et al. (1999) as well as Zhan and Miller (2003) find significant cost increases when the LOS increases. Khan et al. (2006) find a cost increase of 78 % when the patient suffers from a postoperative complication after non-cardiac surgery. Dimick et al. (2004) find a cost increase of $9,607 when a minor complication occurs and $23,869 in case of a major one.[14] All in all, research describes complications as well as longer LOS as strong markers of resource consumption and therefore as cost drivers. It also indicates a clear negative correlation between quality of health care and costs per case, this means quality and costs can be understood as complements. For example, Dimick et al. (2006) offer support for a highly positive correlation between complications (poor quality) and costs. They find that reimbursement for patient care without complications exceeds hospital average costs, resulting in a profit margin of 23 % for the hospital, which collapses to a mere 3.4 % if complications occur. Jha et al. (2009) use the inverse approach by hypothesizing that hospitals with lower costs may be more efficient and thus may provide higher quality than hospitals with higher costs. Even if it

---

[14] Major complications are defined as those that are considered significant enough to result in prolonged LOS or the need for additional interventions. This shows that even minor complications that do not result in a prolonged LOS lead to additional costs.

is assumed that the immediate cost effects of these complications are unlikely to be large, it has been shown that even minor complications may have a large impact on medium- to long-term and therefore overall costs. Other effects such as the patient's need for additional interventions and/or possibly a readmission support the argument that short-term quality measured by complications in the PACU can indeed be linked to medium- and long-term quality and therefore affects costs in the patient's case in a meaningful way.

1.2.3 Specialization and the Effect on Specialized Cases

Studies focusing on both manufacturing (e.g. Fisher and Ittner, 1999) and services (e.g. Huckman and Zinner, 2008) offer empirical support for the benefits of greater specialization. Applied to the health care industry, Hyer et al. (2009) analyze the improved performance of what they refer to "as a *focused hospital unit* [...]" (Hyer et al., 2009, p. 204). Clark and Huckman (2012) examine the impact of the degree of specialization of a hospital on its clinical performance in specialized and related areas of cardiovascular surgery and find that greater operational focus has a positive effect on the quality of medical procedures. The literature supports the notion that hospitals can reduce their costs by limiting the array of services they provide (e.g. Tiwari and Heese, 2009). Consistently, Ernst and Szczesny (2006) report intensified specialization efforts of a German hospital after the tightened capped budget systems had been introduced in 1996. These authors found a learning effect occurring in the OR with respect to specialized cases, resulting in a reduction of OR times and labor costs. These results appear to support the notion that specialization strategies of hospitals have achieved their main objectives. However, it is necessary to analyze whether isolated cost containment measures targeted at the OR impact follow-up costs. For instance, there is some concern that shorter OR times turned the OR into a more fast-paced and therefore potentially more mistake-prone working environment. Savings made in the OR and the outcome quality of the hospital may act as conflicting objectives (i.e. substitutes). If so, this might lead to more mistakes and complications that have adverse cost consequences which may overcompensate the cost savings gained in the OR. However, the literature provides strong evidence that a higher rate of specialized surgeries together with an improved allocation of resources leads to decreased complication rates and fewer mistakes in the OR (e.g. Clark and Huckman, 2012). Here, it must also be considered that more experienced surgical teams generate improved outcomes due to learning effects (Contreras et al., 2011). Thus, the quality proxy for follow-up costs is likely to provide some preliminary answers to cost effects, although no actual cost data are available.

Based on the reasoning, and supported by the literature, the hypothesis regarding OR times reads:

H_1: *Ceteris paribus, specialization in the OR leads to a decrease in the OR time for specialized interventions.*

With regard to complications, as already stated, the following hypothesis results:

H_2: *Ceteris paribus, specialization in the OR leads to a decrease in the probability of complications in the PACU for specialized interventions.*

1.2.4 Specialization and the Effect on Non-Specialized Cases

A specialization strategy with its learning effects in the OR leads to the question whether these effects in the specialized field may not only appear with specialized cases, but might also affect the non-specialized cases by means of spillover effects. Huesch and Sakakibara (2009) show that knowledge spillovers are a common finding in models of innovation and have been applied in health economics research in the last decade. They emphasize that based on surgeons directly interacting with team members and the daily exchange of experience, spillovers may occur subsequently to a specialization strategy. Other than that, experience gained through learning can improve processes of care, refine the standardization of procedures as well as policies and lead to improved quality control. To put it in other words: By implementing a specialization strategy and undertaking the attendant adjustments to processes and allocation plans, a form of organizational learning might begin. Consider adjustments to the workflow made to increase surgical and process efficiency to allow for the increased numbers of specialized cases.[15] A decreasing probability of complications for all surgeries might be a possible result of these efforts originally geared toward the endoprosthetic procedures. To measure the effects of specialization and co-specialization[16] on quality performance, Clark and Huckman (2012) use patients' mortality rate. However, they fail to provide evidence for positive spillover effects between a focal activity and what they call related activities.[17] Regarding the specialization efforts and the higher volume of specialized cases, Com-Ruelle et al. (2008) show that outcomes can be improved by increasing activity volumes. Dudley et al. (2000) review this effect in

---

[15] See table 1.4.

[16] Medical fields near to the field of specialization.

[17] The authors define positive spillovers as the extent to which a co-specialization in areas related to the cardiovascular procedures directly influences quality performance on cardiovascular patients. Areas related to cardiovascular procedures are identified by focusing on secondary diagnoses of primary cardiovascular patients being named in a sufficient number of cases (20 %).

more detail and show that mortality as the quality proxy was lower in hospitals with a greater number of elective surgeries. The translation into the specialization/quality language allows concluding that a specialization strategy in the OR and the gained learning experience in the specialized field, lead to efficiency gains appearing in other parts of daily work—called organizational learning or spillovers. Because the OR is a knowledge-intensive setting in which repetition and experience breed competence (Skinner, 1974), it is reasonable to assume that experience gained is likely to lead to economies of scale in shared resources (Clark and Huckman, 2012), and spillover effects on non-specialized cases can result.

Since the presence of such spillover effects is assumed, the hypotheses read:

H_3: *Ceteris paribus, specialization in the OR leads to a decrease in the OR time for non-specialized interventions.*

H_4: *Ceteris paribus, specialization in the OR leads to a decrease in the probability of complications in the PACU for non-specialized interventions.*

1.3 Data and Estimation Models

1.3.1 Information on the Data

In this analysis, the same anonymous data set is used as it has been by Ernst and Szczesny (2008). The data stem from the anesthesiology department of a small German hospital that provides basic care and keeps its patients in approximately 100 beds. For the entire study period, the hospital was community-owned by the county and operated as a not-for-profit entity.

This type of detailed German micro-level hospital data from a period prior to the introduction of DRGs in Germany (2003/2004) is virtually non-existent and allows studying a specialization strategy unaffected by the yearly adjustments that are common under the DRG system. Previous research has documented that the focal hospital had pursued a specialization strategy in endoprosthetic surgeries in response to the 1996 reimbursement reform, i.e. the introduction of tightened capped budgets (Ernst and Szczesny, 2008). The data set contains information on a short-term measure of outcome quality: the occurrence of complications in the PACU. Complications are documented beginning in 1994, which allows examining the period from 1994 to 1998, which contains the period with the biggest increase in the specialization (1996-1998). 6,491 observations are used to examine the effects of the specialization strategy.[18]

The total number of cases per year fluctuates around 1,300 with the number of specialized interventions greatly increasing from 241 (around 18 %) in 1994 to 433 (around 36 %) in 1998. A major shift from non-endoprosthetic to endoprosthetic surgeries is observed from 1996 to 1997. While the number of endoprosthetic interventions is sharply rising from 277 in 1996 to 364 in 1997, there is a strong decline in the number of non-endoprosthetic ones from 1,116 to 851 over the same period. This shift is regarded as the result of the specialization strategy. In the years following the study period, the number of non-endoprosthetic surgeries stabilizes at around 1,200 (around 70 %) per year. It is worth mentioning that the absolute number of surgeries considerably rises, reaching a maximum of 569 endoprosthetic procedures in 2001 and 2002. This second increase is thus not realized at the expense of non-specialized surgeries, but by cashing in on the investment in the specialization strategy, leading to a higher number of specialized interventions, a higher degree of capacity utilization and an enhanced fixed costs degression.

1.3.2 Estimation Models – OR Times

In order to address the first hypothesis of the present study, the following regression model, trying to explain the OR times of the endoprosthetic procedures (ORT_ENDO), was set up:[19]

$$\begin{aligned}\text{ORT\_ENDO} = {} & \beta_0 + \beta_1 \cdot \Sigma\,\text{ORT\_ENDO} \\ & + \beta_2 \cdot \text{ASA}_2 + \beta_3 \cdot \text{ASA}_3 + \beta_4 \cdot \text{ASA}_{4/5} \\ & + \beta_5 \cdot \text{BLOSS} + \beta_6 \cdot \text{OOR} + \beta_7 \cdot \text{ORW} + \beta_8 \cdot \text{AGE} + \beta_9 \cdot \text{AGE}^2 \\ & + \beta_{10} \cdot \text{SUR}_2 + \cdots + \beta_{19} \cdot \text{SUR}_{11} \\ & + \beta_{20} \cdot 1995 + \cdots + \beta_{23} \cdot 1998 + \varepsilon \end{aligned} \tag{1.1}$$

In model 1.1, the experience in the specialized cases is measured via cumulated OR times and used as the regressor Σ ORT_ENDO. Because hospitals tend to have more precise information regarding expected severity than insurers, they tend to replace difficult cases with lighter cases while holding total patient numbers (capacity) stable (Ellis, 1998). Since the occurrence of complications is highly correlated with the American Society of

---

[18] Table 1.4 shows the descriptive statistics for all relevant variables.

[19] See table 1.1 for a short description of all the variables. Models are estimated by Ordinary Least Squares (OLS) regression unless otherwise specified.

Anesthesiologists (ASA) classification of patients (e.g. Hautmann et al., 2010), a different patient portfolio has an impact on overall quality. On average, fewer patients with high severity lead to fewer complications and thus causes quality measures such as complications or mortality rates to improve (i.e. to be positively biased). To obtain meaningful results, it is therefore essential to control for the risk classification respectively the medical condition of the patients. To do so, the ASA score is employed. The literature confirms this score as a high-quality and appropriate predictor for complications and postoperative outcomes (e.g. Arvidsson et al., 1996; Hall and Hall, 1996; Wolters et al., 1996). The observation's ordinal ASA score is translated into dummy variables ASA_i ($i = 1, 2, ..., 5$), which respectively represent the classification of the ASA scores (1 for a normal healthy patient and 5 for a moribund one). $\mathrm{ASA}_{4/5}$ is the dummy for the patient classified as ASA 4 or 5.[20] The complication rate is expected to increase with an increasing ASA score. It is reasonable to have a closer look at the mean ASA scores in (non-)endoprosthetic cases.[21] Mean ASA scores are higher for endoprosthetic cases than for non-endoprosthetic ones in each considered year. The mean ASA score for endoprosthetic cases slightly decreases with fluctuations over time. For non-endoprosthetic cases, the mean ASA score even increases slightly with fluctuations, which explains the increase in the overall mean ASA score. To conclude, there cannot be found a tendency neither for endoprosthetic, nor for non-endoprosthetic cases. It does not seem that the hospital implemented a certain selection strategy as a reaction to increased cost pressure.

In addition to the ASA classification, the blood loss (BLOSS) and the patient's age (AGE) are assumed to be signifiers of the complexity of a case, possibly extending OR times. Although AGE does not signify complexity itself, it can be used as a proxy for various unobservable signifiers of complexity. As there might be a disproportionately high increase in complexity with increasing age, the squared age is used as an additional regressor (AGE^2) in order to represent a possibly nonlinear relation between the patient's age and the OR time. Furthermore, the waiting time resulting from operating activities (ORW) is used as a further indicator of complexity by identifying the time required for additional unplanned actions during procedures.[22] OOR is a count variable of other operations which are conducted simultaneously besides the main operation. SUR_i is a dummy

---

[20] For a description of the ASA scores, see table 1.3. ASA 1 is the reference category, and therefore, ASA_1 is not included in the regression model. Due to the small number of observations with ASA 5, there is a single dummy for ASA 4 and 5.

[21] For the mean ASA scores, see table 1.4.

[22] For example, it includes consultations with another physician or the preparation of surgical instruments for an unexpected higher scope of a procedure.

variable for the i-th surgeon in the data set.[23] Using these dummies, it can be controlled for unobservable differences between surgeons. It may be the case that a certain surgeon is generally faster or slower than others. 1995 to 1998 are dummy variables for the individual years to control for potential unobserved changes over time. The term ε represents the error term.

Since the dummies for the individual years do not have an influence significantly different from zero (in the following "significant" for short),[24] and the variance inflation factors (VIFs) are in parts considerably above 10 for the dummies,[25] these are replaced in model 1.2 by the dummy POST95, which signals whether an observation is from 1996 onwards (specialization period) to control for potential differences between the period before and the period of specialization.

$$\begin{aligned}\text{ORT\_ENDO} = {} & \beta_0 + \beta_1 \cdot \Sigma\,\text{ORT\_ENDO} \\ & + \beta_2 \cdot \text{ASA}_2 + \beta_3 \cdot \text{ASA}_3 + \beta_4 \cdot \text{ASA}_{4/5} \\ & + \beta_5 \cdot \text{BLOSS} + \beta_6 \cdot \text{OOR} + \beta_7 \cdot \text{ORW} + \beta_8 \cdot \text{AGE} + \beta_9 \cdot \text{AGE}^2 \\ & + \beta_{10} \cdot \text{SUR}_2 + \cdots + \beta_{19} \cdot \text{SUR}_{11} \\ & + \beta_{20} \cdot \text{POST95} + \varepsilon \end{aligned} \tag{1.2}$$

For the reason of the present study, the following regression model, trying to explain the OR times of the non-endoprosthetic procedures ($\text{ORT\_}\overline{\text{ENDO}}$), was set up:

$$\begin{aligned}\text{ORT\_}\overline{\text{ENDO}} = {} & \beta_0 + \beta_1 \cdot \Sigma\,\text{ORT\_ENDO} + \underbrace{\beta_2 \cdot \Sigma\,\text{ORT\_}\overline{\text{ENDO}}}_{\text{omitted}} \\ & + \beta_3 \cdot \text{ASA}_2 + \cdots + \beta_5 \cdot \text{ASA}_{4/5} \\ & + \beta_6 \cdot \text{BLOSS} + \beta_7 \cdot \text{OOR} + \beta_8 \cdot \text{ORW} + \beta_9 \cdot \text{AGE} + \beta_{10} \cdot \text{AGE}^2 \\ & + \beta_{11} \cdot \text{SUR}_2 + \cdots + \beta_{20} \cdot \text{SUR}_{11} \\ & + \beta_{21} \cdot 1995 + \cdots + \beta_{24} \cdot 1998 + \varepsilon \end{aligned} \tag{1.3}$$

In model 1.3, the experience in the specialized as well as non-specialized cases is measured via cumulated OR times and used as the regressors $\Sigma\,\text{ORT\_ENDO}$ (specialized/endoprosthetic experience) and $\Sigma\,\text{ORT\_}\overline{\text{ENDO}}$ (non-specialized experience).

In order to obtain more specific results, the regressor $\Sigma\,\text{ORT\_}\overline{\text{ENDO}}$ in model 1.3

---

[23] The first surgeon serves as the reference category and is not included in the regression model.

[24] See table 1.6. The dummies for individual years have also been tested for joint significance using an F-test. P-value = 0.0974.

[25] VIFs indicate possible problems arising from a problematically high level of multicollinearity. A regression without the dummies yields a significant coefficient of $\Sigma\,\text{ORT\_ENDO}$.

is replaced by $\Sigma\,\text{ORT\_AREA}$ in model 1.4, i.e. the experience in the respective OR area (abdomen, extremities, neck, head, proctology, thorax, urology)[26] is used. Likewise, the OR times in the OR areas, i.e. ORT_AREA, serves as the regressand instead of $\text{ORT\_}\overline{\text{ENDO}}$. This means the estimation has been brought from the hospital to the OR area level.

$$\begin{aligned}\text{ORT\_AREA} = {} & \beta_0 + \beta_1 \cdot \Sigma\,\text{ORT\_ENDO} + \underbrace{\beta_2 \cdot \Sigma\,\text{ORT\_AREA}}_{\text{omitted}} \\ & + \beta_3 \cdot \text{ASA}_2 + \cdots + \beta_5 \cdot \text{ASA}_{4/5} \\ & + \beta_6 \cdot \text{BLOSS} + \beta_7 \cdot \text{OOR} + \beta_8 \cdot \text{ORW} + \beta_9 \cdot \text{AGE} + \beta_{10} \cdot \text{AGE}^2 \\ & + \beta_{11} \cdot \text{SUR}_2 + \cdots + \beta_{20} \cdot \text{SUR}_{11} \\ & + \beta_{21} \cdot 1995 + \cdots + \beta_{24} \cdot 1998 + \varepsilon\end{aligned} \tag{1.4}$$

Since almost no dummy for an individual year has a significant influence in model 1.4,[27] these are replaced by the dummy POST95 in model 1.5.

$$\begin{aligned}\text{ORT\_AREA} = {} & \beta_0 + \beta_1 \cdot \Sigma\,\text{ORT\_ENDO} + \underbrace{\beta_2 \cdot \Sigma\,\text{ORT\_AREA}}_{\text{omitted}} \\ & + \beta_3 \cdot \text{ASA}_2 + \cdots + \beta_5 \cdot \text{ASA}_{4/5} \\ & + \beta_6 \cdot \text{BLOSS} + \beta_7 \cdot \text{OOR} + \beta_8 \cdot \text{ORW} + \beta_9 \cdot \text{AGE} + \beta_{10} \cdot \text{AGE}^2 \\ & + \beta_{11} \cdot \text{SUR}_2 + \cdots + \beta_{20} \cdot \text{SUR}_{11} \\ & + \beta_{21} \cdot \text{POST95} + \varepsilon\end{aligned} \tag{1.5}$$

However, there are problems with the regression models due to strong multicollinearity with regard to $\Sigma\,\text{ORT\_ENDO}$ and $\Sigma\,\text{ORT\_}\overline{\text{ENDO}}$ (model 1.3) as well as $\Sigma\,\text{ORT\_AREA}$ (model 1.4/1.5). VIFs are often above 200, indicating that there are problematically high levels of multicollinearity. Regressions with only one of these regressors mostly yield significant coefficients. When regressing $\Sigma\,\text{ORT\_}\overline{\text{ENDO}}$ and $\Sigma\,\text{ORT\_AREA}$ on $\Sigma\,\text{ORT\_ENDO}$, it is possible to explain more than 95 % of the variance of the auxiliary regressand. Due to multlicollinearity, the non-endoprosthetic experience therefore has to be omitted from the models.

---

[26] The OR area codes used in this paper can be found in table 1.2.

[27] See table 1.8 and 1.9. The dummies for individual years have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors if necessary (ABD, EXT, NECK, URO). P-value < 0.05 / rejection of "no joint significance" only for one OR area (PRO).

1.3.3 Estimation Models – Complications

To test the hypotheses regarding the probability of a patient experiencing a complication in the PACU, different maximum likelihood models for dichotomous dependent variables (binary logit models) are estimated. The models examine the influence of a set of independent variables on the probability of a complication $P(\text{COMPL})$.

The dummy variable COMPL, which reveals if there is a complication in the PACU, is used. The regression results enable to make statements regarding how the probability of experiencing a complication changes, if the independent variables in the regression model alter. Based on the assumptions and the clinical information about the various independent variables, the following logit model is established:[28]

$$\begin{aligned}
&\mathrm{E}(\text{COMPL} \mid \boldsymbol{x}) = P(\text{COMPL} = 1 \mid \boldsymbol{x}) = \int_{-\infty}^{z} f(t)dt = \int_{-\infty}^{z} \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \\
&z = \boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \text{ENDO} + \beta_2 \cdot \text{ENDOSHARE} \\
&\qquad + \beta_3 \cdot \text{ASA}_2 + \beta_4 \cdot \text{ASA}_3 + \beta_5 \cdot \text{ASA}_{4/5} \\
&\qquad + \beta_6 \cdot \text{BLOSS} + \beta_7 \cdot \text{OOR} + \beta_8 \cdot \text{ORT} + \beta_9 \cdot \text{ORW} \\
&\qquad + \beta_{10} \cdot \text{URG} + \beta_{11} \cdot \text{AGE} + \beta_{12} \cdot \text{AGE}^2 \\
&\qquad + \beta_{13} \cdot \text{SUR}_2 + \cdots + \beta_{22} \cdot \text{SUR}_{11} \\
&\qquad + \beta_{23} \cdot 1995 + \cdots + \beta_{26} \cdot 1998
\end{aligned} \tag{1.6}$$

To depict the influence of the specialization strategy in endoprosthetic surgeries and to test the hypotheses, the dummy variable ENDO and the continuous variable ENDOSHARE are deployed. By indicating if the procedure is an endoprosthetic one, ENDO controls for potential quality differences between endoprosthetic and other interventions. ENDOSHARE reflects the three-month moving average of the percentage of endoprosthetic cases.[29] To represent complexity, the OR time (ORT) is also used, since more complex interventions supposedly take longer. OOR is employed since it could be that the probability of a complication is higher if there were other operations conducted simultaneously. Another regressor is URG, the urgency of the intervention, since it is assumed that the probability of a complication is larger for an emergency surgery than for an elective surgery planned a certain time in advance.

To be able to distinguish between potentially differing effects of the specialization strat-

---

[28] For a short description of all the variables, see table 1.1.

[29] For example, it takes the value of 16.29 for an observation in March 1994 because 16.29 % of all surgeries from January to March 1994 were endoprosthetic interventions.

egy on specialized and non-specialized interventions, which is not possible with model 1.6, the interaction term ENDO_ENDOSHARE is additionally included in model 1.7. It reflects the interaction between ENDO and ENDOSHARE.[30] By the help of the interaction term, the overall effects observed in model 1.6 can be split into separate effects for specialized and non-specialized procedures. It is written:

$$\begin{aligned} &\mathrm{E}\,(\mathrm{COMPL} \,|\, \boldsymbol{x}) = P(\mathrm{COMPL} = 1 \,|\, \boldsymbol{x}) = \frac{1}{1+e^{-z}} \\ &z = \boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \mathrm{ENDO} + \beta_2 \cdot \mathrm{ENDOSHARE} + \beta_3 \cdot \mathrm{ENDO\_ENDOSHARE} \\ &\quad + \beta_4 \cdot \mathrm{ASA}_2 + \beta_5 \cdot \mathrm{ASA}_3 + \beta_6 \cdot \mathrm{ASA}_{4/5} \\ &\quad + \beta_7 \cdot \mathrm{BLOSS} + \beta_8 \cdot \mathrm{OOR} + \beta_9 \cdot \mathrm{ORT} + \beta_{10} \cdot \mathrm{ORW} \\ &\quad + \beta_{11} \cdot \mathrm{URG} + \beta_{12} \cdot \mathrm{AGE} + \beta_{13} \cdot \mathrm{AGE}^2 \\ &\quad + \beta_{14} \cdot \mathrm{SUR}_2 + \cdots + \beta_{23} \cdot \mathrm{SUR}_{11} \\ &\quad + \beta_{24} \cdot 1995 + \cdots + \beta_{27} \cdot 1998 \end{aligned} \tag{1.7}$$

Since the dummies for 1996, 1997, and 1998 are significant, but the dummy for 1995 (before specialization) is not,[31] the dummies for the individual years are replaced by the dummy POST95. Therefore, the regression models are:

$$\begin{aligned} &\mathrm{E}(\mathrm{COMPL} \,|\, \boldsymbol{x}) = P(\mathrm{COMPL} = 1 \,|\, \boldsymbol{x}) = \frac{1}{1+e^{-z}} \\ &z = \boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \mathrm{ENDO} + \beta_2 \cdot \mathrm{ENDOSHARE} \\ &\quad + \beta_3 \cdot \mathrm{ASA}_2 + \beta_4 \cdot \mathrm{ASA}_3 + \beta_5 \cdot \mathrm{ASA}_{4/5} \\ &\quad + \beta_6 \cdot \mathrm{BLOSS} + \beta_7 \cdot \mathrm{OOR} + \beta_8 \cdot \mathrm{ORT} + \beta_9 \cdot \mathrm{ORW} \\ &\quad + \beta_{10} \cdot \mathrm{URG} + \beta_{11} \cdot \mathrm{AGE} + \beta_{12} \cdot \mathrm{AGE}^2 \\ &\quad + \beta_{13} \cdot \mathrm{SUR}_2 + \cdots + \beta_{22} \cdot \mathrm{SUR}_{11} \\ &\quad + \beta_{23} \cdot \mathrm{POST95} \end{aligned} \tag{1.8}$$

---

[30] The interaction term therefore equals zero if it is a non-specialized procedure and ENDOSHARE if it is a specialized one.

[31] See table 1.12. The dummies for the individual years 1996-1998 have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors. P-value = 0.0000.

$$\begin{aligned} \mathrm{E}(\mathrm{COMPL} \mid \boldsymbol{x}) &= P(\mathrm{COMPL} = 1 \mid \boldsymbol{x}) = \frac{e^z}{1+e^z} \\ z = \boldsymbol{x}'\boldsymbol{\beta} &= \beta_0 + \beta_1 \cdot \mathrm{ENDO} + \beta_2 \cdot \mathrm{ENDOSHARE} + \beta_3 \cdot \mathrm{ENDO\_ENDOSHARE} \\ &\quad + \beta_4 \cdot \mathrm{ASA}_2 + \beta_5 \cdot \mathrm{ASA}_3 + \beta_6 \cdot \mathrm{ASA}_{4/5} \\ &\quad + \beta_7 \cdot \mathrm{BLOSS} + \beta_8 \cdot \mathrm{OOR} + \beta_9 \cdot \mathrm{ORT} + \beta_{10} \cdot \mathrm{ORW} \\ &\quad + \beta_{11} \cdot \mathrm{URG} + \beta_{12} \cdot \mathrm{AGE} + \beta_{13} \cdot \mathrm{AGE}^2 \\ &\quad + \beta_{14} \cdot \mathrm{SUR}_2 + \cdots + \beta_{23} \cdot \mathrm{SUR}_{11} \\ &\quad + \beta_{24} \cdot \mathrm{POST95} \end{aligned} \tag{1.9}$$

1.4 Results

1.4.1 OR Times

Before dealing with the estimation results, it is reasonable to have a look at figure 1.1 which illustrates the development of OR times over time. There can be detected declining OR

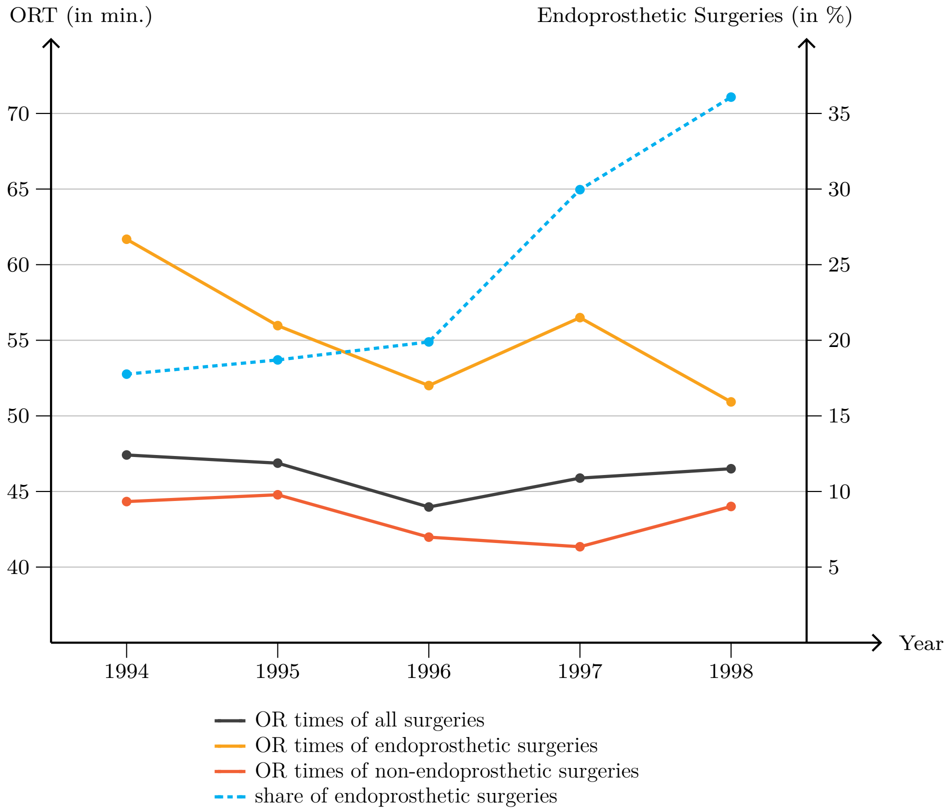

Figure 1.1: Average OR Times of Surgeries

times of endoprosthetic interventions over time. With regard to non-endoprosthetic ones, there is no clear trend. Interestingly, the OR times of endoprosthetic surgeries already decline when the share of endoprosthetics has not risen considerably yet. Obviously, the mere preparation of a specialization on endoprosthetic surgeries reduces endoprosthetic OR times. A possible explanation for this phenomenon might be a thorough inspection, reorganization, and optimization of OR processes in the early stages of the specialization process. Generally, processes and structures are clearly more in the focus of interest in this transition period. The Hawthorne effect might also be an approach: As the surgeons anticipate that the OR as a whole is under observation in this specialization process and so they are, too, their productivity improves.

Table 1.6 depicts the estimation results of model 1.1. The experience in endoprosthetic procedures does not have an influence on the OR times of endoprosthetic surgeries. However, it has to be remembered that there have been problems with the model due to multicollinearity.[32] In order to solve these problems, model 1.2 has been set up. The estimation results can also be found in table 1.6. Implied by the negative[33] coefficient of Σ ORT_ENDO, the experience in endoprosthetic procedures reduces OR times of endoprosthetic interventions. This supports H_1, and therefore H_1 cannot be rejected.

Regarding the patient's risk classification, the results do not confirm the expectations. For model 1.2, ASA_i do not have an influence on the OR times of endoprosthetic surgeries. A rising blood loss increases OR time, as it is with other operations executed besides the main intervention. The other control variables ORW and AGE do not have an influence on the endoprosthetic OR times.[34] There cannot be detected differences between surgeons.[35]

Table 1.7 depicts the estimation results of model 1.3. Implied by the negative coefficient of Σ ORT_ENDO, the experience in endoprosthetic procedures reduces OR times of non-endoprosthetic surgeries. This means there is a spillover effect from specialized to non-specialized interventions, manifesting in shorter OR times. In the following, the estimation

---

[32] See section 1.3.2.

[33] All coefficients which are interpreted are significantly different from zero.

[34] The two regressors representing age have also been tested for joint significance using an F-test. P-value = 0.3. The variance in AGE is comparatively low in model 1.2. The standard deviation (SD) is 9.87; in model 1.3, the SD of AGE is 24.19. This complicates the detection of a possibly existing effect.

[35] VIFs for model 1.2 are considerably below 10, indicating that there are no severe problems arising from a problematically high level of multicollinearity. For AGE and AGE^2, VIFs are larger, obviously resulting from a high correlation between these two regressors. Though, AGE^2 has not been dropped from the regression equations due to the theoretical foundation. Further exceptions are some surgeon dummies. The correlation matrix of the variables is also provided in table 1.5.

results of model 1.4 and 1.5 are explained which allow more specific statements about the spillover effect.

Table 1.8 and 1.9 contain the estimated coefficients of model 1.4. The overall spillover effect on non-endoprosthetic interventions in model 1.3 is specified here: There is not a spillover effect on interventions in all OR areas, but extremities (EXT), proctology, and urology. The result with regard to EXT is the most obvious one, since endoprosthetics is a field within EXT. But there is also a spillover effect on interventions in OR areas which are “further away” from the endoprosthetic field.

Table 1.10 and 1.11 depict the estimation results of model 1.5. Similar to model 1.4, the experience in endoprosthetic interventions has a negative, OR time-reducing, influence on procedures in the same OR areas as for model 1.4 and additionally thorax.[36] In summary, H_3 is partly supported, and therefore H_3 cannot be rejected.

Regarding the patient’s risk classification, the results confirm the expectations. For model 1.3 to 1.5, most coefficients of ASA_i are positive.[37] The absolute value of the coefficients increases with an increase in the ASA score. Thus, the results confirm that the OR time increases with an increasing patient risk classification. A rising blood loss also increases the OR time, as it is with other operations executed besides the main intervention. The patient’s age has a nonlinear relation with the OR time.[38] Not surprisingly, the waiting time from operating activities also extends the OR time. There are differences between surgeons in OR times of non-endoprosthetic interventions.[39]

1.4.2 Complications

Figure 1.2 illustrates the development of complication rates over time. They considerably decline for endoprosthetic and non-endoprosthetic surgeries. Although the drop is larger

---

[36] VIFs for model 1.3 to 1.5 are considerably below 10, indicating that there are no severe problems arising from a problematically high level of multicollinearity. VIFs are larger for AGE and AGE^2 and some surgeon dummies. The correlation matrix of the variables is also provided in table 1.5.

[37] Exceptions: ASA_3 and $\text{ASA}_{4/5}$ in ORT_HEAD (model 1.4 and 1.5).

[38] The two regressors representing age have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors. P-value < 0.05 in all cases except ORT_THO (model 1.4 and 1.5, p-value > 0.1).

[39] Although there are time differences, they should not be overrated, but interpreted cautiously. For example, a surgeon with higher OR times than surgeon 1 must not be regarded as “slow”. There are most likely differences between individual surgery portfolios with regard to complexity and difficulty which cannot be controlled for with the regressors. Thus, a direct comparison of OR times is virtually impossible.

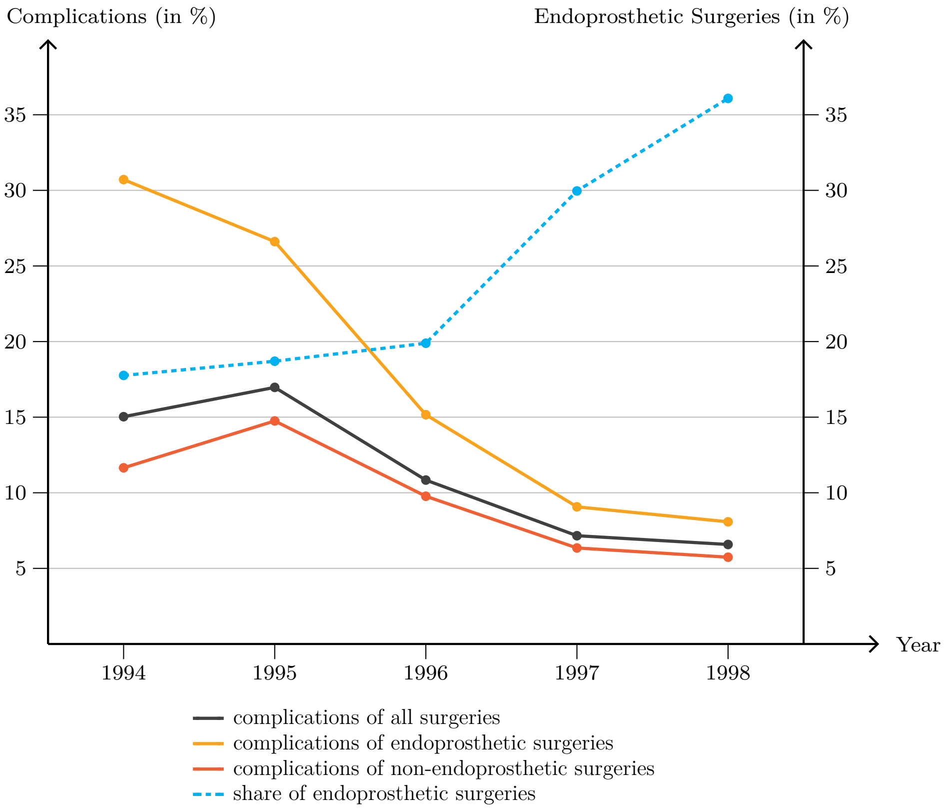

Figure 1.2: Complication Rates of Surgeries

for endoprosthetic interventions, there is also a drop in non-specialized ones, which are seemingly disregarded by specializing in endoprosthetics. However, this does not seem to happen. As with the OR times, the complication rates already decline when the share of endoprosthetics has not risen considerably yet.[40]

Table 1.12 depicts the estimation results of the logit models regarding the probability of experiencing complications in the PACU. Model 1.6 and 1.8 explore whether an increase in the share of specialized procedures has an effect on the complications of all interventions. The negative coefficient of ENDOSHARE (model 1.8) indicates that an increase in the share of specialized procedures leads to an overall decrease in the probability of experiencing complications, i.e. increased specialization leads to a decrease in complications for specialized as well as non-specialized cases. The negative coefficient of ENDOSHARE (model 1.9) has the same indications as in model 1.8. The negative coefficient of ENDO_ENDOSHARE (model 1.9) indicates that the effects of the specialization strategy are stronger for specialized procedures compared to non-specialized ones, i.e.

---

[40] Possible explanations have been outlined in section 1.4.1.

the complication risk is reduced more for specialized cases.[41] The results regarding non-specialized cases indicate the presence of spillover effects. In summary, H_2 and H_4 are supported, and therefore H_2 and H_4 cannot be rejected. This is considered an important result because in contrast to Clark and Huckman (2012), who fail to find spillover effects, there are such effects in these data.

In model 1.6 and 1.7, the coefficients of the dummy for 1996, 1997 and 1998 are negative, as it is for the coefficient of POST95 in model 1.8 and 1.9. This finding suggests that in spite of public distrust, the increasing cost pressure caused by the introduction of rigidly capped budgets in 1996 did not lead to an increase in complications and thus to compromises on quality, at least if a specialization strategy had been implemented as a reaction. Consequently, sophisticated specialization-based cost-reduction strategies and the previously linked learning and spillover effects with regard to short-term quality appear to be strategic complements rather than substitutes.

For model 1.6 to 1.9, all coefficients of ASA_i are positive. The absolute value of the coefficients increases markedly with an increase in the ASA score. Thus, the results confirm the expectations concerning the patient's risk classification. Regarding the other control variables, there is a changing probability of experiencing complications with a change in the patient's age.[42] Positive coefficients of BLOSS, ORW and ORT indicate that an increase in the amount of lost blood, the waiting time from operating activities as well as the OR time itself lead to a higher probability of complications in the PACU. There are no differences in the probability of complications between surgeons.

Summa summarum: It can be stated that the specialization has a positive effect not only for specialized cases, but also for the others due to spillover effects. Besides the OR times, the complication rates in the PACU decrease for all interventions, indicating lower costs in the short, medium, and the long term.

---

[41] VIFs for model 1.6 to 1.9 are considerably below 10, indicating that there are no severe problems arising from a problematically high level of multicollinearity. The only exceptions are ENDO and ENDO_ENDOSHARE in model 1.7 and 1.9 with values slightly higher than 10, but they are still significant, so multicollinearity is not a major problem, as well as AGE and AGE^2 and some surgeon dummies in model 1.6 to 1.9. The correlation matrix of the variables is also provided in table 1.5.

[42] The two regressors representing age have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors. P-value < 0.001 in all cases.

1.4.3 Limitations

There are some limitations to the results. One must bear in mind that the complexity of the study further increases by the fact that, in reality, hospitals are likely to respond to an increase in economic incentives with a mixture of strategies and have, in fact, been shown to do so in empirical research (e.g. Ernst and Szczesny, 2005). The problem is that effects of the respective strategies may cancel out one another, which obviously makes it even harder to predict the resulting effect on quality and costs over the entire patient's case.

Speaking of effects interfering with one another, one must also bear in mind the omission of the non-endoprosthetic experience in model 1.3 to 1.5 due to multicollinearity with regard to $\Sigma\,\text{ORT\_ENDO}$ and $\Sigma\,\text{ORT\_}\overline{\text{ENDO}}/\Sigma\,\text{ORT\_AREA}$. As the endoprosthetic experience is correlated with the non-endoprosthetic one, the remaining regressor $\Sigma\,\text{ORT\_ENDO}$ might also carry parts of the explanatory power of the omitted regressor $\Sigma\,\text{ORT\_}\overline{\text{ENDO}}/\Sigma\,\text{ORT\_AREA}$. However, no alternative measurement of experience has been possible in the present case.

Concerning the definition of the patient's case in this study: It would be reasonable that a patient's case comprises the intervention, the PACU, the intensive care unit, the ward up to the patient's discharge. If there were a readmission (e.g. due to possible complications) associated with the primary intervention, it would need to be part of the patient's case, too. However, it is not possible with the given data set to have a patient's case comprise any treatment after the PACU and it is not possible to connect cases such that a case might be a readmission of a patient from a former case. Therefore, it is not clear whether a reduction of OR times (resulting from learning and spillover effects associated with the specialization strategy) leads to higher complication rates at a later point beyond the patient's case as defined in this context. Complications might, for example, show up in the ward, potentially leading to a longer LOS and therefore higher costs (e.g. Kalish et al., 1995; Collins et al., 1999; Zhan and Miller, 2003). Unfortunately, these data have not been available.

1.5 Robustness of Estimation Models

Several robustness tests have been conducted on the results. Model 1.3 to 1.5 have been estimated with a different measurement of experience: via the sum of interventions instead of cumulated OR times. The results do not relevantly differ.

The same results are obtained when using the one-month, six-month, or nine-month moving average of the percentage of endoprosthetic cases for ENDOSHARE instead of the three-month moving average in model 1.8 and 1.9.[43] Furthermore, the estimation results of model 1.6 to 1.9 virtually do not change even when the observation period is enlarged to 2002, i.e. the year prior to the introduction of the DRG system.

To control for potential differences between statutorily and privately insured patients, a dummy variable if the patient was privately insured has been included in model 1.6 to 1.9. Regarding the complications, the short-term quality measure, no significantly divergent results for privately and statutorily insured patients are found.

All regression models have been controlled for heteroskedasticity by using the Breusch-Pagan test (level of significance $\alpha = 0.05$). Robust standard errors (White) have been used in case of heteroskedasticity in order to receive valid statistical inferences.

VIFs have been computed for all models in order to detect problematically high levels of multicollinearity. Apart from the exceptions explicitly mentioned (dummies for the individual years, AGE and AGE^2 as well as some surgeon dummies), VIFs are below 10 and do not indicate problematically high levels of multicollinearity.

1.6 Conclusions

Since there has been the need to draw a more comprehensive view of isolated cost savings on the entire patient's case, this paper contributes to the literature by providing an important link between the cost savings related to learning and spillover effects in the OR triggered by a specialization strategy and the potentially higher follow-up costs which might overcompensate the savings in the OR.

It has been hypothesized that isolated cost-reducing activities in the OR (through a specialization strategy and resulting learning and spillover effects) shorten OR times and simultaneously increase quality by lowering the probability of experiencing complications and therefore reduce the costs over the entire patient's case for both specialized (learning effects) and non-specialized cases (spillover effects). To make statements regarding how follow-up costs are affected by the specialization strategy, a quality proxy has been used.

---

[43] In contrast to the other ENDOSHARE variables, ENDOSHARE_1 (one-month moving average) does not use the month of the observation, but the previous month. Further, it is a single value and only titled "average" for consistency reasons.

The available data have enabled the measurement of short-term quality by estimating the effects of different factors on the probability of experiencing complications in the PACU. It has been found that with an increased degree of specialization, the hospital can reduce OR times and simultaneously improve short-term quality. This result applies for specialized cases (endoprosthetic surgeries) as well as for non-specialized cases. Therefore, it has been concluded that increased specialization efforts and associated considerations regarding the (re-)organization of processes and procedures within the specialized area imply positive spillovers on other surgeries. The effects of the endoprosthetic process optimization have been named a form of learning. These effects also show that efficiency efforts in the OR do not necessarily lead to higher follow-up costs after the surgical intervention. On the contrary, the hypothesis of a reduction in complications and therefore a cost reduction over the entire patient's case can be confirmed.

The findings reveal a hint of association between learning effects in the OR triggered by the specialization strategy, resulting in efficiency gains, and higher patient volumes using economies of scale. Improved patient outcomes (i.e. lower complication rates) link these developments with overall costs of a patient's case. A strong reduction in the complication appearance for endoprosthetic as well as non-endoprosthetic cases increases the specialization effects described above. Thus, this paper contributes to the literature by showing that isolated cost-saving activities have a cost-saving effect not only in the isolated area, but also in the rest of the patient's case in the form of lower follow-up costs due to a lower complication risk.

Future research might obtain deeper insights in individual vs. organizational learning and spillover effects as well as the consequences on costs. With a larger data set, it would be possible to conduct estimations on surgeon level, which have not been done here due to mostly a low number of cases by a single surgeon within a specific OR area and partially also a low number of interventions in general. With an enhanced database, future research might investigate the specialization, learning, and spillover effects in a larger perspective, defining the patient's case the way it has been proposed in the introduction, i.e. a patient's case comprises the intervention, the PACU, the intensive care unit, the ward up to the patient's discharge. If there were a readmission (e.g. due to possible complications) associated with the primary intervention, it would be supposed to be part of the initial patient's case, too. With an even further enhanced database, possibly from the health insurers, future research might investigate the effects even in a larger perspective, not restricting the patient's case to the hospital, but regarding the case from the hospital up to the after-hospital care, i.e. to the point, in which the case is ultimately over from the patient's perspective, e.g. after the hospital stay and a subsequent physiotherapy. Then, research would be able to evaluate the specialization strategy from the overall perspective. Linked studies should also think about a different, more concise measurement of experience in order to ease the problem with multicollinearity, to be able to better separate between experience in specialization and OR area, and to not have to omit the experience in the respective OR area from the model. However, this would obviously require much medical expertise and very detailed data.

Appendix 1.A

Table 1.1: Description of Variables

Variable	Description
AGE	Patient's age
ASA_i	Dummy for ASA class i $\quad \forall\ i = \{1, 2, ..., 5\}$. Classification of patients with regard to their physical condition. The smaller, the better the condition is. For a description of the ASA classes, see table 1.3
BLOSS	Blood loss (measured in ml)
COMPL	Dummy for complication in the PACU
ENDO	Dummy for endoprosthetic procedure
ENDOSHARE	Three-month moving average of the share of endoprosthetic procedures as a proportion of all cases
ENDO_ENDOSHARE	Interaction term of ENDO and ENDOSHARE. 0 if ENDO is 0 and equal to ENDOSHARE if ENDO is 1
OOR	Number of other operations which are done simultaneously with the main intervention
ORT	Operation time (time from incision to suture; in minutes)
ORW	Waiting time resulting from operating activities
POST95	Dummy for observation between 1996 and 1998
SUR_i	Dummy for the i-th surgeon $\quad \forall\ i = \{1, 2, ..., 11\}$
URG	Urgency of an operation (1 = elective surgery, 2 = urgent surgery, 3 = emergency surgery)
$\Sigma\,\text{ORT\_ENDO}$	Experience in endoprosthetic area (measured as cumulated operation time)
$\Sigma\,\text{ORT\_}\overline{\text{ENDO}}$	Experience apart from endoprostethic area (measured as cumulated operation time)
$\Sigma\,\text{ORT\_AREA}$	Experience in OR area (measured as cumulated operation time). For the OR areas, see table 1.2

Continued on the next page

Variable	Description
ε	Error term
1995/1996/1997/1998	Dummy for the year 1995/1996/1997/1998

Table 1.2: OR Areas

Abbr.	OR Area
ABD	Abdomen
EXT	Extremities (endoprosthetic interventions excluded)
NECK	Neck
HEAD	Head
PRO	Proctology
THO	Thorax
URO	Urology

Table 1.3: ASA Classes (American Society of Anesthesiologists, 2019)

Class	Description
ASA 1	A normal healthy patient
ASA 2	A patient with mild systemic disease
ASA 3	A patient with severe systemic disease
ASA 4	A patient with severe systemic disease that is a constant threat to life
ASA 5	A moribund patient who is not expected to survive without the operation
ASA 6$^{1)}$	A declared brain-dead patient whose organs are being removed for donor purposes

$^{1)}$ Data set does not contain patients classified as this ASA class.

Table 1.4: Summary Statistics of Relevant Variables

		1994	1995	1996	1997	1998	Σ
AGE	$\bar{x}$	46.75	51.21	50.99	55.08	57.68	52.15
	σ	25.6888	23.3509	23.4360	22.6586	21.4096	23.7038
ASA$_1$	n	300	195	296	186	152	1,129
	p	22.11 %	14.71 %	21.25 %	15.31 %	12.67 %	17.39 %
ASA$_2$	n	560	617	681	614	591	3,063
	p	41.27 %	46.53 %	48.89 %	50.53 %	49.25 %	47.19 %
ASA$_3$	n	397	452	374	383	391	1,997
	p	29.26 %	34.09 %	26.85 %	31.52 %	32.58 %	30.77 %
ASA$_4$	n	98	61	41	32	65	297
	p	7.22 %	4.60 %	2.94 %	2.63 %	5.42 %	4.58 %
ASA$_5$	n	2	1	1	0	1	5
	p	0.15 %	0.08 %	0.07 %	0.00 %	0.08 %	0.08 %
$\sum_{i=1}^{5} \text{ASA}_i$	n	1,357	1,326	1,393	1,215	1,200	6,491
	$\bar{x}$	2.2203	2.2881	2.1170	2.2148	2.3100	2.2275
$\sum_{i=1}^{5} \text{ASA}_i$ (ENDO)1)	$\bar{x}$	2.7469	2.6734	2.4657	2.4533	2.5219	2.5547
$\sum_{i=1}^{5} \text{ASA}_i$ ($\overline{\text{ENDO}}$)2)	$\bar{x}$	2.1066	2.1994	2.0305	2.1128	2.1904	2.1238
COMPL	n	204	225	151	87	79	746
	p	15.03 %	16.97 %	10.84 %	7.16 %	6.58 %	11.49 %
COMPL (ENDO)1)	n	74	66	42	33	35	250
	p	30.71 %	26.61 %	15.16 %	9.07 %	8.08 %	15.99 %
COMPL ($\overline{\text{ENDO}}$)2)	n	130	159	109	54	44	496
	p	11.65 %	14.75 %	9.77 %	6.35 %	5.74 %	10.06 %
ENDO	n	241	248	277	364	433	1,563
	p	17.76 %	18.70 %	19.89 %	29.96 %	36.08 %	24.08 %

Continued on the next page

		1994	1995	1996	1997	1998	Σ
$\overline{\text{ENDO}}$	n	1,116	1,078	1,116	851	767	4,928
	p	82.24 %	81.30 %	80.11 %	70.04 %	63.92 %	75.92 %
ORT	$\bar{x}$	47.4134	46.8741	43.9734	45.8848	46.5042	46.1108
	σ	42.3310	38.7100	37.6850	48.5631	38.4952	41.2601
ORT (ENDO)1)	$\bar{x}$	61.6805	55.9677	52.0036	56.4973	50.9238	54.8720
	σ	28.1110	23.9849	27.0621	51.4731	20.3367	32.9885
ORT ($\overline{\text{ENDO}}$)2)	$\bar{x}$	44.3324	44.7820	41.9803	41.3455	44.0091	43.3320
	σ	44.2128	41.0786	39.6346	46.5304	45.4723	43.1870

1) Only endoprosthetic surgeries considered.

2) Only non-endoprosthetic surgeries considered.

Table 1.5: Correlation Coefficients

	AGE	AGE^2	BLOSS	ENDO	ENDO-SHARE	ENDO_ ENDO-SHARE	OOR	ORT	ORW	POST95	URG	Σ ORT_ ENDO
AGE	1.0000											
AGE^2	0.9696 (0.0000)	1.0000										
BLOSS	0.1417 (0.0000)	0.1457 (0.0000)	1.0000									
ENDO	0.4096 (0.0000)	0.4077 (0.0000)	0.1484 (0.0000)	1.0000								
ENDO-SHARE	0.1423 (0.0000)	0.1346 (0.0000)	0.0790 (0.0000)	0.1818 (0.0000)	1.0000							
ENDO_ ENDOSH.	0.3861 (0.0000)	0.3844 (0.0000)	0.1691 (0.0000)	0.9394 (0.0000)	0.3464 (0.0000)	1.0000						
OOR	0.0368 (0.0030)	0.0165 (0.2000)	0.1381 (0.0000)	−0.1029 (0.0000)	0.0216 (0.0818)	−0.0930 (0.0000)	1.0000					
ORT	0.3090 (0.0000)	0.2802 (0.0000)	0.3742 (0.0000)	0.1196 (0.0000)	−0.0025 (0.8422)	0.1000 (0.0000)	0.2694 (0.0000)	1.0000				
ORW	0.2412 (0.0000)	0.2274 (0.0000)	0.1751 (0.0000)	0.2588 (0.0000)	0.3911 (0.0000)	0.3325 (0.0000)	0.0502 (0.0001)	0.2009 (0.0000)	1.0000			
POST95	0.1132 (0.0000)	0.1031 (0.0000)	0.0683 (0.0000)	0.1149 (0.0000)	0.6488 (0.0000)	0.2127 (0.0000)	0.0115 (0.3528)	−0.0211 (0.0895)	0.3655 (0.0000)	1.0000		
URG	0.0462 (0.0002)	0.0689 (0.0000)	0.0013 (0.9166)	−0.1922 (0.0000)	0.1025 (0.0000)	−0.1829 (0.0000)	0.0322 (0.0096)	0.0043 (0.7282)	0.0396 (0.0014)	0.1310 (0.0000)	1.0000	
Σ ORT_ ENDO	0.1476 (0.0000)	0.1374 (0.0000)	0.0864 (0.0000)	0.1533 (0.0000)	0.8491 (0.0000)	0.2914 (0.0000)	0.0295 (0.0174)	−0.0112 (0.3661)	0.4450 (0.0000)	0.8163 (0.0000)	0.1359 (0.0000)	1.0000

P-values in parentheses.

Appendix 1.B

Table 1.6: Estimation Results Model 1.1 and 1.2

$\hat{\beta}_i$	Model 1.1 (ORT_ENDO, years, OLS regression)		Model 1.2 (ORT_ENDO, POST95, OLS regression)	
(Intercept)	62.50508 **	(23.82396)	63.07661 **	(23.83174)
Σ ORT_ENDO	−0.00012	(0.00015)	−0.00012 *	(0.00006)
ASA_2	4.31085	(7.15033)	4.39384	(7.15164)
ASA_3	10.49217	(7.19876)	10.54168	(7.19910)
$ASA_{4/5}$	8.69071	(7.96947)	8.71348	(7.95751)
BLOSS	0.02366 ***	(0.00256)	0.02371 ***	(0.00256)
OOR	4.27956 .	(2.38421)	4.25616 .	(2.38160)
ORW	0.07564	(0.09804)	0.10556	(0.09622)
AGE	−0.39718	(0.63563)	−0.42271	(0.63598)
AGE^2	0.00377	(0.00474)	0.00399	(0.00475)
SUR_2	6.47802	(12.46254)	5.52145	(12.41763)
SUR_3	6.28951	(11.43868)	5.73926	(11.36265)
SUR_4	−5.96298	(10.92831)	−7.25021	(10.83764)
SUR_5	—	—	—	—
SUR_6	—	—	—	—
SUR_7	18.31565	(11.20279)	16.82042	(11.12097)
SUR_8	11.46042	(12.38259)	11.18241	(12.30038)
SUR_9	17.74206	(11.44548)	16.57724	(11.34916)
SUR_{10}	7.00818	(14.14672)	6.15889	(14.11276)
SUR_{11}	8.36716	(12.10661)	6.28302	(11.99921)
1995	−2.41056	(3.50819)	—	—
1996	−7.85422	(4.96002)	—	—
1997	−2.70985	(7.23471)	—	—
1998	−5.71742	(10.10500)	—	—
POST95	—	—	−4.62474	(2.84637)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses.

n = 1,563

R^2 (model 1.1) = 0.17054, R^2 (model 1.2) = 0.16773

Table 1.7: Estimation Results Model 1.3

$\hat{\beta}_i$	Model 1.3 (ORT_$\overline{\text{ENDO}}$, OLS regression)		
(Intercept)	26.89226	***	(4.05161)
Σ ORT_ENDO	−0.00024	*	(0.00010)
ASA_2	3.69651	***	(1.01032)
ASA_3	5.95151	***	(1.58051)
ASA_45	8.28044	**	(3.08991)
BLOSS	0.06539	***	(0.00663)
OOR	20.52799	***	(1.26370)
ORW	0.62537	***	(0.09299)
AGE	0.50176	***	(0.09230)
AGE^2	−0.00383	***	(0.00102)
SUR$_2$	−13.05248	**	(5.05485)
SUR$_3$	14.16087	**	(4.65559)
SUR$_4$	−10.69805	**	(3.70955)
SUR$_5$	−22.02584	***	(3.92281)
SUR$_6$	−17.19298	***	(3.88379)
SUR$_7$	16.60418	***	(3.82666)
SUR$_8$	−7.48901		(4.67298)
SUR$_9$	8.57913	*	(4.32699)
SUR$_{10}$	−17.49577	***	(3.83551)
SUR$_{11}$	7.04908	.	(4.27590)
1995	−2.86974		(1.90323)
1996	−1.27006		(3.24131)
1997	1.82202		(4.55936)
1998	5.06048		(6.96916)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. Robust standard errors (White) due to heteroskedasticity.

n = 4,928

$R^2 = 0.43270$

Table 1.8: Estimation Results Model 1.4 for ABD, EXT, HEAD, NECK

Model 1.4 (ORT_AREA, years, OLS regression)

$\hat{\beta}_i$	ABD		EXT		HEAD		NECK	
(Intercept)	30.17111 ***	(8.15904)	27.74989 ***	(4.75470)	−14.46656 *	(6.08221)	139.91780 **	(45.99854)
Σ ORT_ENDO	0.00008	(0.00022)	−0.00035 *	(0.00014)	0.00025	(0.00020)	0.00006	(0.00017)
ASA_2	3.14043	(2.22513)	3.82097 .	(2.08119)	0.73703	(2.18481)	−0.66145	(0.84218)
ASA_3	7.74887 *	(3.55649)	3.91413	(2.40603)	−11.99969 *	(4.09537)	5.84439	(4.85582)
$ASA_{4/5}$	8.68524	(6.44637)	9.56199 **	(3.63557)	−30.88612 **	(7.61392)	−10.63684	(14.25533)
BLOSS	0.05212 ***	(0.00663)	0.06553 ***	(0.00732)	0.30064 **	(0.07993)	0.16319 ***	(0.03394)
OOR	15.38061 ***	(2.12625)	21.25076 ***	(2.59740)	−0.04219	(2.78181)	13.06435 ***	(3.81296)
ORW	0.72411 ***	(0.19976)	0.41000 ***	(0.11160)	0.21195	(0.91750)	0.04874	(0.51134)
AGE	0.73101 **	(0.25257)	0.40731 **	(0.12805)	−0.35834 *	(0.16525)	0.97774 *	(0.39731)
AGE^2	−0.00311	(0.00292)	−0.00383 **	(0.00128)	0.00853 ***	(0.00198)	−0.01414 *	(0.00628)
SUR_2	−28.08120 ***	(7.39471)	−4.32223	(5.86105)	—	—	−145.62029 **	(48.39232)
SUR_3	9.62702	(7.99934)	10.68565 *	(5.21650)	34.22977 **	(8.52746)	−107.05370 *	(49.43001)
SUR_4	−22.60949 **	(6.96846)	−4.60400	(4.26313)	24.76911 *	(8.73984)	−116.96934 *	(48.00314)
SUR_5	—	—	—	—	18.92644 **	(5.98716)	−143.88231 **	(46.06584)
SUR_6	−19.82159 **	(7.06514)	−22.19581 ***	(4.64318)	—	—	—	—
SUR_7	8.16257	(6.95611)	18.45161 ***	(4.47790)	32.46839 ***	(4.84764)	−56.13828	(47.99230)
SUR_8	−28.10958 **	(10.86487)	0.76462	(5.35967)	—	—	—	—
SUR_9	8.65967	(8.16539)	6.98531	(4.65089)	20.55813 **	(5.93634)	−112.32905 *	(50.12487)
SUR_{10}	−14.57138 *	(7.17102)	−3.19145	(4.67315)	—	—	−111.63634 *	(48.31958)
SUR_{11}	−5.89138	(7.41176)	11.67157 *	(5.10455)	37.43395 ***	(6.21090)	—	—
1995	−10.16944 *	(4.33996)	0.97923	(3.02088)	0.56467	(4.74853)	2.98877	(2.11640)
1996	−13.27850 .	(7.24939)	1.17496	(4.83514)	−4.35502	(7.14510)	−0.34396	(4.73735)
1997	−16.05105	(10.33887)	5.83311	(6.91379)	−13.19681	(10.26005)	−3.97335	(7.42205)
1998	−14.98600	(15.34627)	12.47275	(10.28727)	−21.52230	(15.16340)	−4.90314	(11.62089)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. For ABD, EXT and NECK, robust standard errors (White) due to heteroskedasticity.

n (ABD) = 1,367, n (EXT) = 1,939, n (HEAD) = 34, n (NECK) = 474

R^2 (ABD) = 0.42167, R^2 (EXT) = 0.29709, R^2 (HEAD) = 0.97017, R^2 (NECK) = 0.87301

Table 1.9: Estimation Results Model 1.4 for PRO, THO, URO

Model 1.4 (ORT_AREA, years, OLS regression)

$\hat{\beta}_i$	PRO		THO		URO	
(Intercept)	13.34864	(14.73185)	−0.79744	(32.80808)	36.69582 ***	(4.91395)
Σ ORT_ENDO	−0.00059 **	(0.00019)	−0.00025	(0.00083)	−0.00032 **	(0.00012)
ASA_2	2.40258	(2.50351)	8.50078	(12.17366)	0.71210	(1.42562)
ASA_3	−2.39401	(3.44435)	13.54180	(16.02671)	4.27589	(2.65480)
$ASA_{4/5}$	−3.12263	(5.88275)	−0.96873	(23.07768)	−0.99066	(3.02070)
BLOSS	0.07201	(0.08804)	0.08208	(0.05181)	0.11832 ***	(0.01208)
OOR	6.59168 *	(2.77409)	13.84663	(9.46869)	18.41840 ***	(1.89408)
ORW	0.18652	(0.19001)	−0.49856	(0.56976)	0.93668 **	(0.29706)
AGE	0.15980	(0.28435)	1.38804	(0.93065)	0.19725	(0.16857)
AGE^2	0.00039	(0.00287)	−0.01264	(0.00873)	−0.00062	(0.00185)
SUR_2	—	—	−11.84336	(42.07789)	—	—
SUR_3	4.89708	(13.68407)	35.11700	(29.00227)	−1.59887	(5.80873)
SUR_4	0.02490	(13.10518)	4.34564	(25.86571)	−18.91645 ***	(4.38983)
SUR_5	—	—	—	—	—	—
SUR_6	6.19558	(13.92509)	−5.31233	(31.53623)	−27.42315 ***	(3.81155)
SUR_7	11.20083	(12.96891)	19.18777	(24.18541)	14.25696	(15.42437)
SUR_8	5.33692	(18.17062)	—	—	—	—
SUR_9	19.03945	(13.29472)	44.09732	(28.70539)	−29.83865 ***	(5.02701)
SUR_{10}	2.76276	(13.34488)	−4.81763	(29.05730)	−30.46395 ***	(3.77868)
SUR_{11}	18.16574	(13.22207)	23.24289	(28.68812)	—	—
1995	−1.44672	(3.90051)	−9.82391	(16.00775)	1.04701	(2.69468)
1996	9.83790	(6.22971)	−1.27363	(25.87655)	6.05727	(4.46386)
1997	22.14191 *	(9.41904)	25.04016	(40.07725)	9.53161	(6.56738)
1998	29.34345 *	(13.26388)	−4.38456	(56.58075)	11.43131	(8.85171)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. For URO, robust standard errors (White) due to heteroskedasticity.

n (PRO) = 206, n (THO) = 242, n (URO) = 666

R^2 (PRO) = 0.37509, R^2 (THO) = 0.16235, R^2 (URO) = 0.71002

Table 1.10: Estimation Results Model 1.5 for ABD, EXT, HEAD, NECK

Model 1.5 (ORT_AREA, POST95, OLS regression)

$\hat{\beta}_i$	ABD		EXT		HEAD		NECK	
(Intercept)	30.76233 ***	(8.03935)	26.91032 ***	(4.74835)	−7.50915	(6.18400)	140.41489 **	(47.08956)
Σ ORT_ENDO	−0.00005	(0.00009)	−0.00013 **	(0.00005)	−0.00002	(0.00009)	0.00002	(0.00005)
ASA_2	3.08499	(2.20182)	3.77650 .	(2.08854)	−0.60720	(2.38572)	0.04174	(0.77331)
ASA_3	7.47185 *	(3.51423)	3.96692 .	(2.40967)	−13.43314 **	(4.57836)	6.25889	(4.97485)
$ASA_{4/5}$	9.13818	(6.49350)	10.19279 **	(3.64443)	−26.33284 **	(7.82263)	−9.30807	(14.13418)
BLOSS	0.05253 ***	(0.00662)	0.06569 ***	(0.00728)	0.25554 **	(0.07111)	0.16492 ***	(0.03507)
OOR	15.21526 ***	(2.14897)	21.50182 ***	(2.57543)	0.87773	(2.92074)	12.35213 **	(3.86764)
ORW	0.68158 ***	(0.18703)	0.41180 ***	(0.11183)	−0.16459	(0.99550)	0.06293	(0.52515)
AGE	0.72878 **	(0.25068)	0.40830 **	(0.12833)	−0.17659	(0.15261)	0.95538 *	(0.39228)
AGE^2	−0.00303	(0.00292)	−0.00386 **	(0.00128)	0.00602 **	(0.00173)	−0.01385 *	(0.00623)
SUR_2	−30.36149 ***	(7.40440)	−5.41552	(5.81655)	—	—	−146.46618 **	(49.16950)
SUR_3	7.91678	(8.04220)	9.58125 .	(5.16000)	25.50515 **	(8.71987)	−108.11105 *	(50.55478)
SUR_4	−25.84814 ***	(6.91520)	−6.28493	(4.15539)	20.67776 *	(9.49019)	−116.38939 *	(48.79288)
SUR_5	—	—	—	—	13.37536 .	(6.38264)	−143.16462 **	(46.82480)
SUR_6	−23.10689 ***	(6.94585)	−22.04427 ***	(4.58213)	—	—	—	—
SUR_7	4.14946	(6.74944)	16.37403 ***	(4.30824)	31.56286 ***	(5.34806)	−55.39074	(48.77722)
SUR_8	−32.35612 **	(10.69618)	−0.96600	(5.21700)	—	—	—	—
SUR_9	5.90540	(8.00123)	5.29723	(4.54120)	19.18714 *	(6.65622)	−111.73285 *	(51.08548)
SUR_10	−17.94881 *	(7.03961)	−4.48640	(4.57776)	—	—	−112.39844 *	(49.32117)
SUR_11	−9.85366	(7.31143)	9.32928 .	(4.96651)	33.38953 ***	(6.84005)	—	—
POST95	−4.82010	(3.72685)	−3.20732	(2.34266)	0.70235	(4.08376)	−1.81152	(2.31039)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. For ABD, EXT and NECK, robust standard errors (White) due to heteroskedasticity.

n (ABD) = 1,367, n (EXT) = 1,939, n (HEAD) = 34, n (NECK) = 474

R^2 (ABD) = 0.41822, R^2 (EXT) = 0.29519, R^2 (HEAD) = 0.95298, R^2 (NECK) = 0.87128

Table 1.11: Estimation Results Model 1.5 for PRO, THO, URO

Model 1.5 (ORT_AREA, POST95, OLS regression)

$\hat{\beta}_i$	PRO		THO		URO	
(Intercept)	13.19380	(15.10368)	−3.03177	(32.73071)	35.28071 ***	(4.63233)
Σ ORT_ENDO	−0.00025 **	(0.00007)	−0.00062 .	(0.00032)	−0.00021 ***	(0.00005)
ASA_2	3.08428	(2.53191)	8.45641	(12.21893)	0.72884	(1.42446)
ASA_3	−1.11136	(3.47180)	11.87763	(15.98880)	4.22564	(2.68393)
$ASA_{4/5}$	−4.19086	(5.97385)	−1.21254	(22.76273)	−0.84502	(2.97231)
BLOSS	0.09273	(0.08984)	0.07907	(0.05200)	0.11839 ***	(0.01196)
OOR	6.13208 *	(2.82140)	13.73323	(9.50965)	18.39674 ***	(1.88939)
ORW	0.27591	(0.19085)	−0.00999	(0.53122)	0.94437 **	(0.29008)
AGE	0.09653	(0.29111)	1.50598	(0.93580)	0.19241	(0.16868)
AGE^2	0.00095	(0.00294)	−0.01358	(0.00879)	−0.00058	(0.00185)
SUR_2	—	—	−3.14109	(41.64207)	—	—
SUR_3	2.55440	(13.93898)	46.90587 .	(28.23832)	−0.28614	(5.87736)
SUR_4	−2.10482	(13.40518)	7.48248	(25.09485)	−18.21405 ***	(4.29314)
SUR_5	—	—	—	—	—	—
SUR_6	3.71969	(14.27210)	3.60618	(30.87695)	−27.08267 ***	(3.78800)
SUR_7	5.91392	(13.21846)	16.29327	(23.62985)	14.60806	(15.57779)
SUR_8	−3.24941	(18.47852)	—	—	—	—
SUR_9	15.45296	(13.59783)	46.83314	(28.47221)	−29.87318 ***	(5.06899)
SUR_{10}	−0.42128	(13.66043)	−3.11047	(28.04538)	−30.01440 ***	(3.79097)
SUR_{11}	13.43061	(13.49423)	16.36057	(27.58398)	—	—
POST95	5.34578	(3.43889)	19.20090	(13.39528)	3.63099	(2.38936)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. For URO, robust standard errors (White) due to heteroskedasticity.

n (PRO) = 206, n (THO) = 242, n (URO) = 666

R^2 (PRO) = 0.33104, R^2 (THO) = 0.13743, R^2 (URO) = 0.70937

Table 1.12: Estimation Results Model 1.6 to 1.9

$\hat{\beta}_i$	Model 1.6 (COMPL, no interaction term, years, logit)		Model 1.7 (COMPL, interaction term, years, logit)		Model 1.8 (COMPL, no interaction term, POST95, logit)		Model 1.9 (COMPL, interaction term, POST95, logit)	
(Intercept)	−4.13607 ***	(0.60656)	−4.28158 ***	(0.60634)	−3.67233 ***	(0.58798)	−3.86590 ***	(0.58830)
ENDO	0.11576	(0.11307)	0.67211 *	(0.30173)	0.11111	(0.11250)	0.75173 *	(0.29330)
ENDOSHARE	−0.00991	(0.01195)	−0.00082	(0.01256)	−0.03778 ***	(0.00809)	−0.02544 **	(0.00937)
ENDO_ENDOSHARE	—	—	−0.02476 *	(0.01247)	—	—	−0.02842 *	(0.01201)
ASA_2	0.54859 *	(0.22445)	0.55372 *	(0.22385)	0.51867 *	(0.22624)	0.52527 *	(0.22537)
ASA_3	1.17876 ***	(0.23780)	1.18215 ***	(0.23728)	1.14711 ***	(0.23943)	1.15208 ***	(0.23861)
$ASA_{4/5}$	1.65698 ***	(0.27287)	1.66624 ***	(0.27254)	1.61103 ***	(0.27387)	1.62511 ***	(0.27321)
BLOSS	0.00083 ***	(0.00014)	0.00084 ***	(0.00014)	0.00083 ***	(0.00014)	0.00086 ***	(0.00015)
OOR	0.02897	(0.08505)	0.02757	(0.08482)	0.03337	(0.08535)	0.03167	(0.08502)
ORT	0.00306 **	(0.00098)	0.00292 **	(0.00096)	0.00297 **	(0.00099)	0.00280 **	(0.00097)
ORW	0.01918 ***	(0.00555)	0.01974 ***	(0.00555)	0.01554 **	(0.00514)	0.01631 **	(0.00516)
URG	0.09213	(0.10326)	0.07134	(0.10379)	0.07106	(0.10323)	0.04671	(0.10366)
AGE	0.01957	(0.01416)	0.01938	(0.01416)	0.02039	(0.01422)	0.02007	(0.01420)
AGE^2	0.00000	(0.00012)	0.00000	(0.00012)	−0.00001	(0.00012)	−0.00001	(0.00012)
SUR_2	0.49752	(0.69065)	0.53329	(0.69246)	0.39655	(0.69033)	0.43822	(0.69294)
SUR_3	0.47997	(0.46788)	0.40953	(0.46924)	0.36493	(0.46220)	0.29001	(0.46313)
SUR_4	0.30817	(0.42270)	0.28856	(0.42017)	0.37809	(0.41875)	0.35045	(0.41614)
SUR_5	−0.51188	(0.66981)	−0.52213	(0.66795)	−0.42786	(0.67027)	−0.44632	(0.66815)
SUR_6	−0.00560	(0.45506)	−0.02240	(0.45362)	0.06031	(0.45298)	0.03357	(0.45178)
SUR_7	0.41477	(0.41950)	0.42287	(0.41707)	0.53880	(0.41510)	0.54236	(0.41250)
SUR_8	0.28596	(0.58371)	0.28376	(0.58241)	0.22840	(0.58550)	0.22341	(0.58470)
SUR_9	0.18299	(0.45710)	0.14186	(0.45539)	0.23211	(0.45282)	0.18024	(0.45142)
SUR_{10}	0.13133	(0.44432)	0.11676	(0.44268)	0.18544	(0.44231)	0.16235	(0.44073)
SUR_{11}	0.24355	(0.44804)	0.25471	(0.44556)	0.36856	(0.44360)	0.37778	(0.44069)
1995	0.05075	(0.11923)	0.04873	(0.11950)	—	—	—	—
1996	−0.45367 ***	(0.13676)	−0.45900 ***	(0.13680)	—	—	—	—
1997	−1.16876 ***	(0.22603)	−1.15136 ***	(0.22480)	—	—	—	—
1998	−1.31995 ***	(0.29856)	−1.27507 ***	(0.29850)	—	—	—	—
POST95	—	—	—	—	−0.59378 ***	(0.11787)	−0.59807 ***	(0.11738)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. Robust standard errors (White) due to heteroskedasticity.

n = 6,491

BIC (model 1.6) = 4,146.8, BIC (model 1.7) = 4,151.6, BIC (model 1.8) = 4,139.7, BIC (model 1.9) = 4,143.0

McFadden's R^2 (model 1.6) = 0.15566, McFadden's R^2 (model 1.7) = 0.15653, McFadden's R^2 (model 1.8) = 0.15151, McFadden's R^2 (model 1.9) = 0.15270

2 Team vs. Individual Learning in Hospitals

Carsten Bauer, Oliver Unger and Martin Holderried[44]

2.1 Introduction

In most developed countries, the economic pressure on hospitals to improve their efficiency has increased considerably over the past decades. Thus, it is as important for hospitals as it is for every company to obtain detailed information about the economic implications of learning curve effects.[45] Another issue increasingly discussed in recent literature is teamwork in health care.[46] In the focus of interest are surgeons' learning curves effects in a teamwork vs. an individual work setting in tonsillectomies (as a very common surgical procedure in the field of otorhinolaryngology),[47] this means a combination of both aspects discussed in the literature. In order to do so, the effects of their learning and experience in both settings are analyzed and evaluated on a broad scale that considers the short, medium, and long term. This means this paper contributes in combining learning curve effects with team considerations by comparing team and individual learning curves.

The time from incision to suture (in the following "OR time", i.e. operating room time, for short) of tonsillectomies is used as an indicator of learning curve effects which influence the short-term costs related to the OR. The proxy for the medium- to long-term follow-up costs of a tonsillectomy is the probability of a patient experiencing a complication after the intervention. Quality is used to act as an indicator for costs incurred at later points in the treatment process, i.e. the short- to medium-term quality measure serves as a proxy for medium- to long-term follow-up costs. Actual cost data are virtually unavailable for research; however, the data set from a German tertiary care hospital contains costs per

---

[44] We thank Jun.-Prof. Dr. Daniel Schaupp for his support.

[45] For an overview of how learning curve effects have been studied, see the review by Ramsay et al. (2000).

[46] For an overview of how teamwork processes in health care have been studied, see the review by Dinh et al. (2006).

[47] Surgical excision of the palatine tonsils. With about 80,000 cases per year, tonsillectomies are "standard interventions" and one of the most conducted elective surgeries in Germany (Papaspyrou et al., 2012).

case. This enables assuring that the assumed cost indicators are valid and substantiates the results with regard to the effects of learning and medical experience on cost indicators. Since there is no data listing specific cost details,[48] the paper does not try to evaluate the effects of experience on the costs per case directly.

The remainder of this paper is organized as follows: In section 2.2, the hypotheses are developed based on the literature. The data set as well as the research models are described in section 2.3. A discussion of the results and their limitations is presented in section 2.4, followed by some considerations about the robustness of estimations in section 2.5. The paper closes with a short summary and outlines some suggestions for further research in section 2.6.

2.2 Background and Hypotheses

2.2.1 OR Times as a Proxy for Costs

OR times are of special importance, since ORs require the highest labor utilization within the hospital (Pernerstorfer and Huemer, 2008), and therefore, OR times have a large impact on hospitals' costs. OR times being reduced, more interventions are feasible within a given period of time, leading to a fixed costs degression, which contains labor cost savings related to learning effects (Ernst and Szczesny, 2006).

Based on the reasoning, and supported by the literature, the hypothesis regarding OR times reads:

H_1: *Ceteris paribus, surgeons' experience leads to a decrease in the OR time for tonsillectomies.*

Besides, surgeons' experience might also affect follow-up costs. Therefore, complications as an indicator for medium- to long-term follow-up costs are also investigated. Taking into consideration short- as well as medium- to long-term cost indicators, the economic implications of learning in tonsillectomies can be evaluated comprehensively.

2.2.2 Complications as a Proxy for Costs

Whereas OR times are used as an indicator for costs in the short term, complications, being a short- to medium-term quality measure, serve as a medium- to long-term cost

---

[48] E.g. details with regard to costs incurred by OR time or complications.

indicator.[49] In this analysis, the probability of a patient suffering from a complication after the tonsillectomy is employed to measure the quality consequences of surgeons' experience in the OR. This quality proxy thus accomplishes two objectives. First, it can answer the question how experience influences short- to medium-term outcomes with regard to quality. Second, it can link these results to the follow-up costs of the intervention.

Quality is linked to treatment costs because negative outcomes (such as complications or adverse events) are typically associated with higher resource use. There is distinct proof in the literature that complications are a proxy for higher costs. Kalish et al. (1995) depict an impressive increase in costs of $ 16,023 if patients experienced complications. They also show that complications extend the length of stay (LOS), which is consistent with the result of higher costs.[50] Khan et al. (2006) find a cost increase of 78 % when the patient suffers from a postoperative complication after non-cardiac surgery. Dimick et al. (2004) find a cost increase of $ 9,607 when a minor complication occurred and $ 23,869 in case of a major one.[51] Complications and longer LOSs are described as strong markers of resource consumption and therefore as cost drivers (Hoonhout et al., 2009). These results also have important implications for hospital costs because there is a clear positive correlation between poor quality and costs per case (e.g. Baldwin et al., 2003; Chung et al., 2006). Dimick et al. (2006) also offer support for the highly positive correlation between complications and costs: They find that reimbursement for patient care without complications exceeds average hospital costs, resulting in a profit margin of 23 % for the hospital, which collapses to a mere 3.4 % if complications occur. In addition, the literature also claims the inverse view, i.e. that lower health care system costs and the quality of health care provided to patients are positively correlated (Veit et al., 2012; Fleming, 1991; Flood et al., 1994). For example, Jha et al. (2009) hypothesize that hospitals with lower costs may be more efficient and thus may provide higher quality than hospitals with higher costs. Deily and McKay (2006) provide evidence that more cost-efficient hospitals in Florida have lower mortality rates than less efficient ones. The aforementioned literature shows that outcome quality and treatment costs can be understood as complements. Even if it is assumed that the immediate cost effects of these complications are unlikely to be large, it has already been shown above that even minor complications may have a large

---

[49] They are designated short- to medium-term quality measure due to the fact that by far most complications occur several days (on average 4.5 days, day of surgery excluded) after the tonsillectomy.

[50] Kalish et al. (1995) investigate cases of "major surgery patients".

[51] Major complications are defined as those that are considered significant enough to result in prolonged LOS or the need for additional interventions. This shows that even minor complications that do not result in a prolonged LOS lead to remarkable additional costs.

impact on medium- and long-term costs. Further impact of complications such as the patient's readmission and the need for additional interventions (Kim and Soeken, 2005), support the argument that short-term quality measured by complications can indeed be linked to medium- and long-term quality and therefore affects costs per case in a meaningful way.

Regarding complications, the following hypothesis results:

H_2: *Ceteris paribus, surgeons' experience leads to a decrease in the probability of complications for tonsillectomies.*

2.3 Data and Estimation Models

2.3.1 Information on the Data

In this analysis, an anonymous data set from the otorhinolaryngology department of a German tertiary care hospital is deployed. The data set covers interventions from June 2014 until December 2016. Nine hundred observations are used to examine and compare the learning curve effects in tonsillectomies with regard to OR time and complications in a teamwork and an individual work setting.[52] Most of the tonsillectomies are conducted by one of the 28 surgeons individually. In 51 cases, an OR team of two surgeons operates. In one case, there are three surgeons involved. In 44 out of 52 team interventions, a resident physician is involved. The most frequent team consists of a senior physician documented as the first surgeon and a resident physician as the second one (n = 26). This indicates that team interventions are employed for supervised learning.[53] Twenty surgeons take part in team interventions during the observation period, conducting 788 tonsillectomies (in teams or individually). In addition, the data set contains the costs per case.[54] With these, the cost indicators influenced by surgeons' experience can be checked for validity.

2.3.2 Estimation Models – OR Times

The following regression model tries to explain the OR times of tonsillectomies:[55]

---

[52] Table 2.3 to 2.6 show the descriptive statistics for all relevant variables.

[53] See table 2.4 for the team compositions.

[54] The costs per case are based on hospital-internal cost calculations.

[55] See table 2.1 for a short description of all the variables. Models are estimated by Ordinary Least Squares (OLS) regression unless otherwise specified.

$$\begin{aligned}\text{ORT} = \beta_0 &+ \beta_1 \cdot \text{FA} + \beta_2 \cdot \text{OA} + \beta_3 \cdot \text{PROF} + \beta_4 \cdot \text{TEAM} \\ &+ \beta_5 \cdot \text{ASA}_2 + \beta_6 \cdot \text{ASA}_{3/4} \\ &+ \beta_7 \cdot \text{NUM\_DIAG} + \beta_8 \cdot \text{AGE} + \beta_9 \cdot \text{AGE}^2 + \beta_{10} \cdot \text{WEEKEND} \\ &+ \beta_{11} \cdot 2015 + \beta_{12} \cdot 2016 + \varepsilon\end{aligned} \tag{2.1}$$

In model 2.1, the surgeons' experience is measured via the level of qualification which in turn is done by the dummy variables FA, OA, and PROF. The first level "resident physician" serves as the reference category.[56] The usage of qualification levels as indicators of surgeons' experience is legitimated by the requirements for achieving a certain qualification level.[57] As the data set contains interventions by a single surgeon as well as team interventions, the dummy variable TEAM is used to denote team interventions. It seems plausible that these interventions take longer, as the surgeons interact with each other (especially because team interventions are probably meant for supervised learning). Since the patient's physical status might also influence the OR time, it is controlled for the risk classification respectively the medical condition of the patient. To do so, the American Society of Anesthesiologists (ASA) classification of patients is employed. The literature confirms this score as a high-quality and appropriate predictor for complications and postoperative outcomes (e.g. Arvidsson et al., 1996; Hall and Hall, 1996; Wolters et al., 1996). The observation's ordinal ASA score is translated into dummy variables ASA_i $(i = 1, 2, 3, 4)$, which respectively represent the classification of the ASA scores (1 for a normal healthy patient and 4 for one with severe systemic disease). $\text{ASA}_{3/4}$ is the dummy for the patient classified as ASA 3 or 4.[58] NUM_DIAG reprensents the number of diagnoses. It seems likely that the more diagnoses the patient gets the longer will be the OR time due to more processes and/or a higher complexity. A differentiation of variably severe diagnoses is not necessary, since all cases in the data set are classified as Diagnosis-Related Group (DRG) D30B, which reads the case is without complex diagnosis (Institut für das Entgeltsystem im Krankenhaus, 2014a,b, 2015). In addition, the patient's age (AGE) is assumed to be a signifier of the complexity of a case. Higher age might correlate with a longer OR time. Although AGE does not signify complexity itself, it can be used as a proxy for various unobservable signifiers of complexity. As there might be a disproportionately high in-

---

[56] Therefore, the first level "resident physician" is not included in the regression model.

[57] E.g. to become an otorhinolaryngologist (FA), resident physicians need to pass a 24-month basic training and subsequently a 36-month otorhinolaryngologist training.

[58] For a description of the ASA scores, see table 2.2. ASA 1 is the reference category, and therefore, ASA_1 is not included in the regression model. Due to the small number of observations with ASA 3 or 4, there is a single dummy for ASA 3 and 4.

crease in complexity with increasing age, i.e. the relation between age and OR time might be nonlinear, the squared age (AGE^2) is also contained. The regressor WEEKEND is a dummy for weekend surgeries (on Sat. / Sun.). Such surgeries might hint at their urgency, which can possibly also be used as an indicator of complexity of the case. 2015 and 2016 are dummy variables for the individual years to control for potential unobserved changes over time. The term ε represents the error term.

In order to obtain more specific results with regard to the surgeons, the regressand ORT is replaced by ORT_TEAM_SUR in model 2.2 and $\text{ORT\_}\overline{\text{TEAM}}\text{\_SUR}$ in model 2.3. In model 2.2, only tonsillectomies conducted by a surgeon who is part of a surgery team during the observation period are considered. In model 2.3, the OR times of the cases by surgeons who are not part of a surgery team during the observation period are explained.

$$\begin{aligned}\text{ORT\_TEAM\_SUR} = {} & \beta_0 + \beta_1 \cdot \text{FA} + \beta_2 \cdot \text{OA} + \beta_3 \cdot \text{PROF} + \beta_4 \cdot \text{TEAM} \\ & + \beta_5 \cdot \text{ASA}_2 + \beta_6 \cdot \text{ASA}_{3/4} \\ & + \beta_7 \cdot \text{NUM\_DIAG} + \beta_8 \cdot \text{AGE} + \beta_9 \cdot \text{AGE}^2 + \beta_{10} \cdot \text{WEEKEND} \\ & + \beta_{11} \cdot 2015 + \beta_{12} \cdot 2016 + \varepsilon\end{aligned} \tag{2.2}$$

$$\begin{aligned}\text{ORT\_}\overline{\text{TEAM}}\text{\_SUR} = {} & \beta_0 + \beta_1 \cdot \text{FA} + \beta_2 \cdot \text{OA} + \beta_3 \cdot \text{PROF} \\ & + \beta_4 \cdot \text{ASA}_2 + \beta_5 \cdot \text{ASA}_{3/4} \\ & + \beta_6 \cdot \text{NUM\_DIAG} + \beta_7 \cdot \text{AGE} + \beta_8 \cdot \text{AGE}^2 + \beta_9 \cdot \text{WEEKEND} \\ & + \beta_{10} \cdot 2015 + \beta_{11} \cdot 2016 + \varepsilon\end{aligned} \tag{2.3}$$

2.3.3 Estimation Models – Complications

To test the hypothesis regarding the probability of a patient experiencing a complication connected to a tonsillectomy, a maximum likelihood model for dichotomous dependent variables (binary logit model) is estimated. The model examines the influence of a set of independent variables on the probability of a complication $P(\text{COMPL})$.

Based on the assumptions and the clinical information about further independent variables, the following logit model is established:[59]

---

[59] See table 2.1 for a short description of all the variables.

$$
\begin{aligned}
&\mathrm{E}(\mathrm{COMPL} \mid \boldsymbol{x}) = P(\mathrm{COMPL} = 1 \mid \boldsymbol{x}) = \int_{-\infty}^{z} f(t)dt = \int_{-\infty}^{z} \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \\
&z = \boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \mathrm{FA} + \beta_2 \cdot \mathrm{OA} + \beta_3 \cdot \mathrm{PROF} \\
&\qquad + \beta_4 \cdot \mathrm{ASA}_2 + \beta_5 \cdot \mathrm{ASA}_{3/4} \\
&\qquad + \beta_6 \cdot \mathrm{NUM\_DIAG} + \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{WEEKEND} \\
&\qquad + \beta_{10} \cdot \mathrm{ORT} + \beta_{11} \cdot \mathrm{ABSCESS} \\
&\qquad + \beta_{12} \cdot 2015 + \beta_{13} \cdot 2016
\end{aligned}
\tag{2.4}
$$

WEEKEND is again used as a possible indicator of case complexity. The probability of a complication is assumed to be larger for an emergency operation than for an elective surgery planned in advance. To represent complexity, the OR time is also used, since more complex interventions supposedly take longer and might have a higher probability of a complication. The dummy ABSCESS which represents tonsillectomies conducted due to a peritonsillar abscess, an acute infection, is also deployed.[60] Thus, the dummy is used as an indicator of complication risk. Maybe, there is a higher risk for abscess tonsillectomies. It seems likely that the complication risk also rises with a higher number of diagnoses.[61]

2.3.4 Estimation Models – Costs per Case

Based on the argumentation in section 2.2.1 to 2.2.2, the regression models in section 2.3.2 to 2.3.3 which use OR time and complications as cost indicators have been set up. A regression of the costs per case is conducted on these two cost indicators as well as the LOS[62] as another one (with control variables) to assure that they are valid and to substantiate the results:[63]

---

[60] An acute infection of the tissue between the palatine tonsils and the musculus constrictor pharyngis next to the palatine tonsils.

[61] Due to the small number of complications in the given data set (n = 35), it does not make sense to conduct regressions of complications in tonsillectomies by surgeons who attend team interventions (n = 32) and by surgeons who do not (n = 3).

[62] Table 2.7 and 2.8 show descriptive statistics of LOS.

[63] 899 observations (one missing value of costs per case). Initially, all control variables from model 2.1 to 2.4 have been inserted. The omitted control variables have been tested for joint significance using an F-test with heteroskedasticity-robust standard errors. P-value = 0.2165. The paper does not try to evaluate the effect of experience (FA, OA, PROF) on the costs per case directly, since there is no detailed cost data precisely stating which amount is incurred by a certain cost driver (e.g. OR time, LOS or complications) available. See table 2.1 for a short description of all the variables.

$$\begin{aligned}\text{COSTS} = {} & \beta_0 + \beta_1 \cdot \widehat{\text{ORT}} + \beta_2 \cdot \text{COMPL} + \beta_3 \cdot \text{LOS} \\ & + \beta_4 \cdot \text{TEAM} + \beta_5 \cdot \text{NUM\_DIAG} + \beta_6 \cdot \text{AGE} + \beta_7 \cdot \text{AGE} \\ & + \beta_8 \cdot 2015 + \beta_9 \cdot 2016 + \varepsilon\end{aligned} \tag{2.5}$$

Model 2.5 might suffer from endogeneity (OLS model assumption of the error term being uncorrelated with the regressors does not hold), as its error term might not be uncorrelated with the error term of model 2.1. In order to resolve the endogeneity problem, the estimated values of ORT ($\widehat{\text{ORT}}$) from regression (2.1) are used instead of the actual values of ORT in model 2.5 (2SLS regression[64]).

The regressand is replaced according to the concept seen in section 2.3.2 here, too:

$$\begin{aligned}\text{COSTS\_TEAM\_SUR} = {} & \beta_0 + \beta_1 \cdot \widehat{\text{ORT}} + \beta_2 \cdot \text{COMPL} + \beta_3 \cdot \text{LOS} \\ & + \beta_4 \cdot \text{TEAM} + \beta_5 \cdot \text{NUM\_DIAG} + \beta_6 \cdot \text{AGE} + \beta_7 \cdot \text{AGE}^2 \\ & + \beta_8 \cdot 2015 + \beta_9 \cdot 2016 + \varepsilon\end{aligned} \tag{2.6}$$

$$\begin{aligned}\text{COSTS\_}\overline{\text{TEAM}}\text{\_SUR} = {} & \beta_0 + \beta_1 \cdot \widehat{\text{ORT}} + \beta_2 \cdot \text{COMPL} + \beta_3 \cdot \text{LOS} \\ & + \beta_4 \cdot \text{NUM\_DIAG} + \beta_5 \cdot \text{AGE} + \beta_6 \cdot \text{AGE}^2 \\ & + \beta_7 \cdot 2015 + \beta_8 \cdot 2016 + \varepsilon\end{aligned} \tag{2.7}$$

Model 2.6 and 2.7 will help to quantify the cost effects of surgeons sometimes operating in teams and the ones always operating on their own respectively.

2.4 Results

2.4.1 OR Times

Table 2.9 depicts the estimation results of model 2.1 to 2.3. Implied by the negative[65] coefficients of FA, OA, and PROF, the surgeon's experience, represented by qualification levels, reduces OR times in tonsillectomies. With rising experience, this time reduction in comparison to resident physicians grows.[66] Before juxtaposing the qualification level

---

[64] Two-Stage Least Squares regression. In the first stage, the values of ORT are estimated via OLS in regression (2.1). In the second stage, the estimated values are used as a regressor in OLS regression (2.5).

[65] All interpreted coefficients are significantly different from zero (in the following "significant" for short).

[66] Exception: The absolute value of the coefficient of OA is slightly smaller than the one of FA in model 2.6.

coefficients in model 2.2 and 2.3, it is reasonable to inspect average OR times. For resident physicians, the OR time is shorter for surgeons who take part in team surgeries during the observation period (28.22 min.) than for their peers who do not (33.76 min.).[67] This holds when comparing the OR times of tonsillectomies conducted by a single surgeon: Surgeons who take part in team interventions have a shorter average OR time (28.05 min. for residents, 22.57 min. for all) than the remaining ones (33.76 min. for residents, 25.93 min. for all).[68] Figure 2.1 and 2.2 illustrate differences in average OR times. Turning to

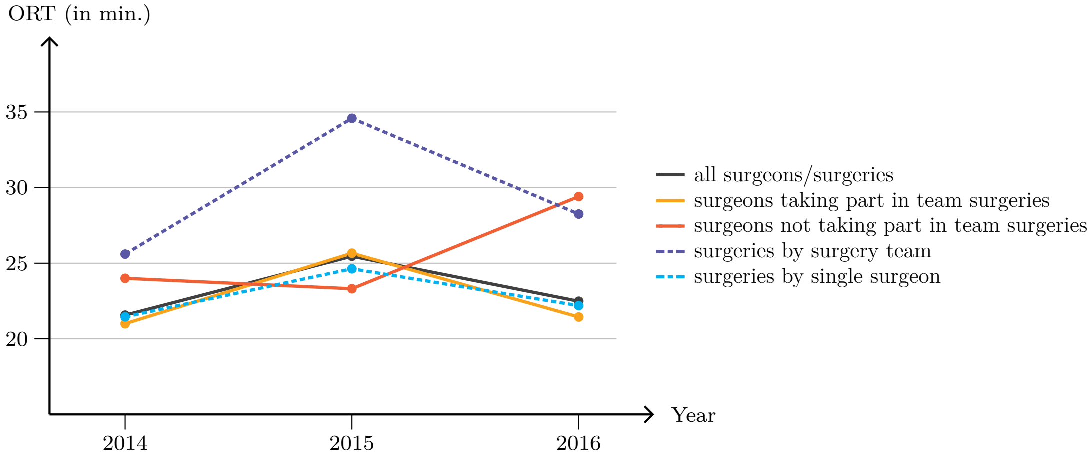

Figure 2.1: Average OR Times of Surgeries

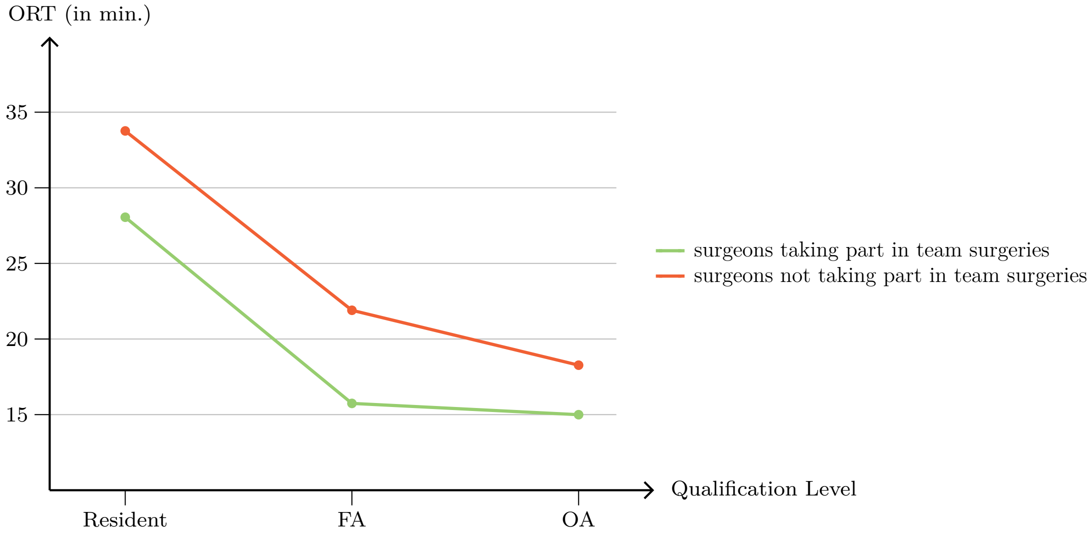

Figure 2.2: Average OR Times of Surgeries by Single Surgeon

---

[67] See table 2.3.

[68] See table 2.5.

the coefficients in model 2.2 and 2.3, the reduction of OR times in comparison to resident physicians is about the same for otorhinolaryngologists taking part in team surgeries (model 2.2) and the ones not taking part in team surgeries (model 2.3). For senior physicians who take part the reduction of OR times compared to resident physicians is smaller than for the ones who do not, as the former ones virtually have the same reduction as otorhinolaryngologists.[69] The combination of average OR times and qualification level coefficients suggests resident physicians learning faster when sometimes operating in teams, and reaching the next qualification level with a higher level of knowledge and expertise. As a consequence, the learning effect in the following qualification level is smaller. In summary, H_1 is supported, and therefore H_1 cannot be rejected.

The coefficients of TEAM (model 2.1, 2.2) meet the expectations. Team interventions take longer than interventions by a single surgeon. Regarding the patient's risk classification, the results do not confirm the expectations, since there are no significant coefficients for ASA_i. This might be traced back to the fact that tonsillectomies are rather "small" interventions. In "larger" interventions including higher blood loss, the situation can be completely different, e.g. in endoprosthetic interventions[70] (Bauer et al., 2020). With regard to NUM_DIAG, the coefficient is positive (model 2.1, 2.2) as it has been assumed. The number of diagnoses within the tonsillectomy enlarges the OR time. The remaining regressors do not have an influence on the OR time.[71]

To end this section, it is worthwhile having a look at the coefficients of determination of model 2.2 and 2.3. The latter model has a considerably higher one and thus explains a higher share of the variation in the regressand.

2.4.2 Complications

The estimation results of the logit model regarding the probability of experiencing complications are given in table 2.10. The insignificant coefficients of FA, OA, and PROF indicate that the complication risk is unchanged with rising experience. This does not support H_2, and therefore H_2 is rejected.

---

[69] Considering the OR time reductions compared to resident physicians, one has to keep in mind that the OR times also differ between the resident physicians who take part in team interventions and the ones who do not. See figure 2.2 and table 2.5.

[70] Total hip or knee replacements.

[71] The two regressors representing age have also been tested for joint significance using an F-test. P-value > 0.1 / no rejection of "no joint significance" in all three cases.

Patients classified as ASA 3 or 4 face a higher probability of experiencing complications compared to patients classified as ASA 1. In contrast, ASA 2 patients have a lower probability of experiencing a complication. Hence, the assumption only holds partially. It might be the case that in tonsillectomies, rather “small” interventions, the complication risk is not that much represented by the ASA score as it is in “larger” interventions. The positive coefficient of NUM_DIAG supports the argumentation of rising complication risk with rising number of diagnoses. Age has a nonlinear relation with the complication risk.[72] Interventions on weekends have a lower complication risk; hence, the expectations are not met. The negative coefficient of ORT implies: Patients with a shorter OR time have a lower complication risk. This is an indication that OR time can represent complexity, since more complex interventions seem to take longer and have a higher probability of a complication. Surprisingly, the coefficient of ABSCESS is negative which implies a lower complication risk for tonsillectomies due to a peritonsillar abscess. The probability of a complication sinks over time. The negative coefficient of 2016 can be regarded as a sign of learning which is not covered by the individual experience, i.e. the qualification levels. It might be traced back to improved processes or innovations in operation methods or after-surgery care.

2.4.3 Costs per Case

Table 2.11 depicts the estimation results of model 2.5 to 2.7. OR time and complications positively influence the costs per case, and therefore they can be regarded as valid cost indicators.[73]

Summa summarum: It can be stated that surgeons’ experience has a favorable, negative effect on the OR time and the probability of complications. Since OR time and complications positively affect the costs per case, experience indicates lower costs in the short, medium, and long term. This finding implies hospitals should be obliged to create an environment which accelerates learning. As already discussed in section 2.4.1, team interventions are a favorable learning environment to acquire knowledge fast, especially

---

[72] The two regressors representing age have also been tested for joint significance using an F-test. P-value = 0.0760.

[73] The direction of influence is of course irrelevant for being a cost indicator. The estimated OR times have been used instead of the actual ones in model 2.5 to 2.7, see section 2.3.4. If the costs per case are regressed on estimated OR time and complications only, the coefficient of determination $R^2 = 0.1077$, which means that about 11 % of the variation in the costs per case can be explained by these two cost indicators. If the costs per case are regressed on estimated OR time, LOS, and complications, $R^2 = 0.4929$. For model 2.5, $R^2 = 0.5528$.

for inexperienced surgeons.

To illustrate the effects numerically: A reduction of the OR time by one minute on average reduces the costs per case by 8.50 €.[74] Surgeons who engage in team interventions take about 3.36 minutes shorter in individual surgeries,[75] which results in a cost reduction by 28.56 €. A complication increases costs by 975.72 €.

2.4.4 Limitations

There are some limitations to the results. First of all, the paper is not able to precisely measure surgeons' experience with the given data set, but has to resort to qualification levels which are not that accurate. For example, a resident physician doing first interventions is attributed the same experience as a resident having conducted many tonsillectomies and reaching the next qualification level (otorhinolaryngologist) in the near future.

Next, with 900 observations, the data set is a relatively small sample. This gets obvious in model 2.4, which explores the effect of experience on the probability of complications: There are only 35 tonsillectomies with complications.[76]

One must also bear in mind that the complexity of the individual cases has more dimensions than the ones the present study has been able to use. However, it is virtually impossible to exactly represent them in a regression model.

The findings recommend team surgeries especially for resident physicians due to faster learning and a connected lower resource consumption. But one has to keep in mind that team interventions themselves are very costly, since more than one surgeon is bound.[77] So there is a need for finding the optimal frequency which trades off the extra costs of team interventions against the resulting cost savings. However, it is not possible to achieve this with the relatively small data set at hand.[78]

---

[74] See table 2.11.

[75] See table 2.5.

[76] This is the reason no separate regressions for surgeons operating and not operating in teams have been conducted. See also footnote 61. Nevertheless, the low complication rate (3.89 %) might be an indicator of the overall high quality of this tertiary care hospital in this "standard intervention".

[77] However, there is no significant difference in costs per case between interventions by a single surgeon and team interventions which does not rely on different OR times (as the OR times influence the costs per case and differ between interventions by a single surgeon and team interventions). See table 2.11.

[78] It can be noticed that the surgeons who take part in team interventions altogether conduct 788 interventions whereof only 52 are done in team, which equals about 6.6 %. Anyhow, the study finds differences between the two groups of surgeons. This means the optimal frequency most likely is not large.

2.5 Robustness of Estimation Models

The separate estimations in model 2.2 and 2.3 have been merged into a single model like model 2.1 which additionally contains a dummy variable to distinguish surgeons who take part in team interventions from the ones not taking part. The results do not relevantly differ: With rising experience, the time reduction in comparison to resident physicians grows. Team interventions take longer. If a surgeon who takes part in team interventions operates, the surgery by a single surgeon is shorter. For model 2.6 and 2.7, it is the same: The results of the combined model are in line with the previous ones. There is no significant difference in costs per case between team interventions and interventions by a single surgeon represented by TEAM.[79] However, there is a difference in costs per case between the two groups of surgeons represented by the dummy variable (about –103 € if a surgeon who takes part in team interventions operates).

Model 2.1 to 2.4 have been estimated with a different measurement of experience: via the cumulated number of tonsillectomies by a surgeon at the ENT department of the focal university hospital. This method disregards the experience the surgeons might have gathered before. Therefore, the regression has only been conducted for tonsillectomies by surgeons being resident physicians when entering the department in order to not disregard "too much" experience. The results do not relevantly differ.

All regression models have been controlled for heteroskedasticity by using the Breusch-Pagan test (level of significance $\alpha = 0.05$). Robust standard errors (White) have been used in case of heteroskedasticity in order to receive valid statistical inferences.

The variance inflation factors (VIFs) have been computed for all models. VIFs are close to 1, indicating that there are no severe problems arising from a problematically high level of multicollinearity.

2.6 Conclusions

This paper contributes to the literature by not only providing insights into learning curve effects in the OR, but simultaneously also into differences of these learning effects between surgeons who engage in team interventions and surgeons who do not. Furthermore, it

---

[79] However, there are differences in costs relying on different OR times (as the OR times influence the costs per case and differ between the two groups of surgeries).

can validate hospital's cost indicators by being able to resort to the costs per case in tonsillectomies.

It has been hypothesized that surgeons' experience in tonsillectomies shortens OR times as well as lowers the probability of complications and therefore reduces the costs in the short, medium, and long term. To make statements regarding how short-term costs are affected, OR times have been used. In order to investigate how medium- to long-term follow-up costs are affected by surgeons' experience, complications, a quality measure, have been used.

Increased surgeons' experience has a cost-saving effect not only in the OR itself via reduced OR times, but also with regard to medium- to long-term follow-up costs due to a lower complication risk. This implies a cost reduction and an enhancement in quality are not mutually exclusive. It has also been found that team interventions accelerate surgeons' learning, especially for inexperienced surgeons. Summa summarum, the findings suggest especially resident physicians engaging in team surgeries to acquire a higher level of knowledge already in the first qualification level.

Future research might obtain deeper insights into the effect of team interventions. With a larger data set, it may be possible to investigate the team learning effects for surgeons graded by the frequency they operate (and learn) in teams. With this knowledge, one might concretize the recommendation for team surgeries with regard to an optimal frequency. One might also study the differences between team and individual learning curve effects regarding complications, which has not been possible with the data set at hand. Linked studies should also think about a more concise measurement of experience, to be able to conduct regressions with enhanced data quality. However, this would obviously require much medical expertise. Future research might also obtain more precise economic implications of surgeons' learning and experience with detailed cost data precisely stating costs incurred by different cost drivers.

Appendix 2.A

Table 2.1: Description of Variables

Variable	Description
ABSCESS	Dummy for tonsillectomy conducted due to peritonsillar abscess (acute infection)
AGE	Patient's age
ASA_i	Dummy for ASA class $i \quad \forall\, i = \{1, 2, 3, 4\}$. Classification of patients with regard to their physical condition. The smaller, the better the condition is. $\text{ASA}_{3/4}$ as dummy for ASA class 3 or 4. For a description of the ASA classes, see table 2.2.
COMPL	Dummy for complication (secondary bleeding)
FA	Dummy for surgeon is otorhinolaryngologist ("Facharzt")
LOS	Length of stay (in days)
NUM_DIAG	Number of diagnoses
OA	Dummy for surgeon is senior physician ("Oberarzt")
ORT	Operation time (time from incision to suture; in minutes)
PROF	Dummy for surgeon is university professor
TEAM	Dummy for team intervention
WEEKEND	Dummy for weekend surgery (Saturday/Sunday)
ε	Error term
2015/2016	Dummy for the year 2015/2016

Table 2.2: ASA Classes (American Society of Anesthesiologists, 2019)

Class	Description
ASA 1	A normal healthy patient
ASA 2	A patient with mild systemic disease
ASA 3	A patient with severe systemic disease
ASA 4	A patient with severe systemic disease that is a constant threat to life
ASA 5(1)	A moribund patient who is not expected to survive without the operation
ASA 6(1)	A declared brain-dead patient whose organs are being removed for donor purposes

(1) Data set does not contain patients classified as this ASA class.

Table 2.3: Surgeons – Team Interventions

Surgeons		doing team interventions	not doing team interventions	Σ
Resident	n	13	4	17
	$\overline{\text{ORT}}$[1)]	28.2165	33.7609	28.7328
FA	n	2	3	5
	$\overline{\text{ORT}}$	16.7971	21.9000	18.6697
OA	n	5	2	7
	$\overline{\text{ORT}}$	17.0171	18.2692	17.1423
PROF	n	1	0	1
	$\overline{\text{ORT}}$	13.0000	—	13.0000
All	n	20[2)]	8[2)]	28[2)]
	$\overline{\text{ORT}}$	23.1764	25.9286	23.5189

[1)] Average OR time for interventions with a resident physician documented as the first surgeon. Other qualification levels analog.

[2)] Sum of surgeons sorted by qualification level does not match due to one resident physician and one otorhinolaryngologist (FA) reaching the next level within the observation period.

Table 2.4: Surgeons – Team Composition

Surgeries of		alone	with Resident	with FA	with OA	with PROF	Σ
Resident[1)]	n	475	14	0	0	5	494
FA	n	102	4	0	0	3	109
OA	n	234	26	0	0	0	260
PROF	n	37	0	0	0	0	37
All	n	848	44	0	0	8	900

[1)] Resident physician documented as the first surgeon. Other qualification levels analog.

Table 2.5: Summary Statistics of Surgeries

		2014[1]	2015	2016	Σ
Surgeries	n	191	372	337	900
ORT	$\bar{x}$	21.5654	25.4597	22.4837	23.5189
COMPL	n	9	19	7	35
	p	4.71 %	5.11 %	2.08 %	3.89 %
Surgeries (TEAM_SUR)[2]	n	155	340	293	788
ORT_TEAM_SUR	$\bar{x}$	21.0000	25.6618	21.4437	23.1764
Surgeries ($\overline{\text{TEAM}}$_SUR)[3]	n	36	32	44	112
ORT_$\overline{\text{TEAM}}$_SUR	$\bar{x}$	24.0000	23.3125	29.4091	25.9286
Surgeries (TEAM)[4]	n	5	31	16	52
ORT_TEAM	$\bar{x}$	25.6000	34.5806	28.2500	31.7692
ORT_TEAM (TEAM_Res.)[5]	$\bar{x}$				31.8947
(TEAM_FA)	$\bar{x}$				26.1429
(TEAM_OA)	$\bar{x}$				33.1923
Surgeries ($\overline{\text{TEAM}}$)	n	186	341	321	848
ORT_$\overline{\text{TEAM}}$	$\bar{x}$	21.4570	24.6305	22.1963	23.0130
ORT_$\overline{\text{TEAM}}$ (TEAM_SUR)[6]	$\bar{x}$				22.5693
($\overline{\text{TEAM}}$_SUR)[7]	$\bar{x}$				25.9286
(TEAM_Res.)	$\bar{x}$				28.0536
($\overline{\text{TEAM}}$_Res.)	$\bar{x}$				33.7609
(TEAM_FA)	$\bar{x}$				15.7419
($\overline{\text{TEAM}}$_FA)	$\bar{x}$				21.9000
(TEAM_OA)	$\bar{x}$				14.9952
($\overline{\text{TEAM}}$_OA)	$\bar{x}$				18.2692
(TEAM_PROF)	$\bar{x}$				13.0000

[1] June to December.

[2] Surgeries conducted by a surgeon who is part of a surgery team during the observation period.

[3] Surgeries conducted by a surgeon who is not part of a surgery team during the observation period.

[4] Surgeries conducted by a surgery team.

[5] Surgeries conducted by a surgery team with a resident physician documented as the first surgeon. Other qualification levels analog.

[6] Surgeries conducted by a single surgeon. Surgeon is part of a surgery team during the observation period. Individual qualification levels analog.

[7] Surgeries conducted by a single surgeon. Surgeon is not part of a surgery team during the observation period. Individual qualification levels analog.

Table 2.6: Summary Statistics of Further Variables

		2014¹⁾	2015	2016	Σ
AGE	x̄	26.0524	26.7392	26.0059	26.3189
	σ	13.3447	14.8857	13.6903	14.1378
ASA_1	n	114	246	223	583
	p	59.69 %	66.13 %	66.17 %	64.78 %
ASA_2	n	74	120	107	301
	p	38.74 %	32.26 %	31.75 %	33.44 %
$ASA_{3/4}$	n	3	6	7	16
	p	1.57 %	1.61 %	2.08 %	1.78 %
$\sum_{i=1}^{4} ASA_i$	n	191	372	337	900
	x̄	1.4188	1.3548	1.3620	1.3711
NUM_DIAG	x̄	8.7853	8.1935	9.1840	8.6900
	σ	2.3671	2.4659	2.4316	2.4722
WEEKEND	x̄	0.0628	0.0645	0.0475	0.0578

1) June to December.

Table 2.7: Length of Stay – Year of Surgery

		2014¹⁾	2015	2016	Σ
Surgeries	n	191	372	337	900
LOS²⁾	x̄	5.8429	5.2473	4.6736	5.1589
DRG-planned LOS³⁾	x̄	4.8000	4.6000	4.5000	

1) June to December.

2) The day of discharge is not considered.

3) Planned LOS by the DRG system. The day of discharge is not considered.

Table 2.8: Length of Stay – Patient's Age

AGE		≤ 10	11-17	≥ 18	Σ
Surgeries	n	91	120	689	900
LOS	x̄	5.7912	5.2750	5.0552	5.1589

Appendix 2.B

Table 2.9: Estimation Results Model 2.1 to 2.3

$\hat{\beta}_i$	Model 2.1 (ORT, OLS regression)		Model 2.2 (ORT_TEAM_SUR, OLS regression)		Model 2.3 (ORT_$\overline{\text{TEAM}}$_SUR, OLS regression)	
(Intercept)	22.80878 ***	(4.13855)	22.28831 ***	(4.67899)	28.35809 ***	(6.51712)
FA	−10.48071 ***	(2.45816)	−11.93426 ***	(3.18527)	−11.52996 ***	(2.98575)
OA	−12.18048 ***	(1.68347)	−11.74696 ***	(1.85232)	−15.10348 ***	(3.30308)
PROF	−15.40540 ***	(3.69463)	−14.72200 ***	(3.87589)	—	—
TEAM	10.22789 **	(3.13714)	10.79791 **	(3.30679)	—	—
ASA$_2$	−0.47837	(1.60324)	0.07082	(1.77360)	−3.55175	(2.94118)
ASA$_{3/4}$	0.38005	(5.69067)	0.45825	(7.07185)	−0.95469	(6.65238)
NUM_DIAG	0.76927 *	(0.30626)	0.79537 *	(0.34037)	0.60558	(0.54260)
AGE	−0.06902	(0.16330)	−0.10018	(0.18193)	0.05452	(0.27918)
AGE^2	0.00076	(0.00229)	0.00108	(0.00255)	−0.00038	(0.00399)
WEEKEND	3.35888	(3.11594)	2.79724	(3.48840)	8.43624	(5.30374)
2015	1.34271	(2.05307)	1.75491	(2.37571)	−2.31533	(3.16836)
2016	−1.60650	(2.09816)	−2.15369	(2.42147)	0.13093	(3.03383)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses.

n (model 2.1) = 900, n (model 2.2) = 788, n (model 2.3) = 112

R^2 (model 2.1) = 0.09070, R^2 (model 2.2) = 0.08877, R^2 (model 2.3) = 0.28305

Table 2.10: Estimation Results Model 2.4

$\hat{\beta}_i$	Model 2.4 (COMPL, logit)	
(Intercept)	−7.18607 ***	(1.27501)
FA	−0.30232	(0.59269)
OA	−0.77845	(0.57339)
PROF	1.01849	(0.68319)
ASA_2	−1.13403 .	(0.57938)
$ASA_{3/4}$	2.14550 **	(0.74040)
NUM_DIAG	0.44623 ***	(0.07409)
AGE	0.10870 *	(0.04708)
AGE^2	−0.00184 **	(0.00063)
WEEKEND	−16.21837 ***	(0.42042)
ORT	−0.05428 *	(0.02707)
ABSCESS	−16.44712 ***	(0.94332)
2015	0.43335	(0.47270)
2016	−1.01297 .	(0.56486)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. Robust standard errors (White) due to heteroskedasticity.

n = 900

McFadden's $R^2 = 0.24288$

Table 2.11: Estimation Results Model 2.5 to 2.7

$\hat{\beta}_i$	Model 2.5 (COSTS, OLS regression)		Model 2.6 (COSTS_TEAM_SUR, OLS regression)		Model 2.7 (COSTS_$\overline{\text{TEAM}}$_SUR, OLS regression)	
(Intercept)	−353.34838	(265.21039)	−352.44016	(307.23635)	−421.99508	(285.63183)
$\widehat{\text{ORT}}$	8.50117 *	(3.60184)	7.78751 .	(4.09213)	16.58385 **	(6.18008)
LOS	407.33114 ***	(43.31282)	410.78529 ***	(49.20092)	389.21243 ***	(35.99960)
COMPL	975.72232 ***	(182.82468)	1,014.51442 ***	(199.37726)	557.68068	(354.49339)
TEAM	−40.32713	(62.92768)	−30.23657	(64.58310)	—	—
NUM_DIAG	60.78477 ***	(15.85173)	61.13511 ***	(18.27534)	53.55093 **	(17.27782)
AGE	−16.30832 **	(5.94887)	−16.88860 **	(6.12623)	−15.65028	(16.97341)
AGE^2	0.18469 *	(0.07762)	0.17700 *	(0.07168)	0.29374	(0.29570)
2015	185.18694 *	(78.16076)	205.77573 *	(93.23107)	39.29453	(100.80506)
2016	335.10928 ***	(66.24225)	331.33407 ***	(74.52828)	333.44606 **	(110.70945)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. Robust standard errors (White) due to heteroskedasticity.

n (model 2.5) = 899, n (model 2.6) = 787, n (model 2.7) = 112

R^2 (model 2.5) = 0.55283, R^2 (model 2.6) = 0.55421, R^2 (model 2.7) = 0.62015

3 Learning Curve Effects in Stapes Surgery

Carsten Bauer, Johannes Taeger and Kristen Rak[80]

3.1 Introduction

Stapes surgery as a therapy for conductive hearing loss caused by otosclerosis[81] is conceptually simple, but technically difficult (Hughes, 1991; Sergi and Paludetti, 2016). Besides, its outcome can easily be measured.[82] Therefore, it is regarded as the optimum to study learning curve effects in surgery (Sergi and Paludetti, 2016).

This study contributes in providing insights into learning curve effects in stapes surgery. In order to properly evaluate how surgeons' learning affects the outcome quality, it takes other possibly relevant variables for the success of stapes surgery into account. Thus, it provides a comprehensive study of learning curve effects and possibly predictive factors. Previous research has often investigated either the learning curve in stapes surgery or possibly relevant factors for predicting surgical success. Evaluations of learning curves are often investigations of the authors' own ones. They virtually do not control for other possibly relevant variables for success. Other studies engage in variables for the prediction of success; however, they do not control for surgeons' experience. The regression model approach to learning curve effects is virtually nonexistent in stapes surgery literature.[83]

The time from incision to suture (in the following "OR time", i.e. operating room time, for short) of stapes surgeries is used as an indicator of learning curve effects affecting the short-term costs. Besides, the postoperative air-bone gap (ABG),[84] a parameter widely used for measuring success in stapes surgery, is deployed to measure learning curve effects

---

[80] We thank Prof. Dr. Rüdiger Pryss and especially Ralph Keim for their support.

[81] Otosclerosis causes overflow bone growth (overflow osteogenesis). It often results in a fixation of the stapes footplate (stapesankylosis) which causes a conductive hearing loss. Stapes surgery reconstructs the stapes function by means of a prosthesis, which takes over the sound conduction to the inner ear.

[82] Sergi and Paludetti (2016) palpably refer to the postoperative air-bone gap measured by audiometry.

[83] A common approach are descriptive statistics, e.g. the comparison of mean values with rising experience.

[84] The difference between air conduction and bone conduction threshold is referred to as air-bone gap.

with regard to quality, which influences average medium- to long-term follow-up costs (incurred after surgery). "Quality" is used to act as an indicator for costs incurred at later points in the treatment process, i.e. the quality measure serves as a proxy for follow-up costs. The surgeons' experience as the result of their learning process is measured by their respective sum of stapes surgeries. With 1,066 observations, the data set is of considerable size. This enables the verification of the results of previous research which has often used distinctly smaller samples. Moreover, there are 24 surgeons in the data set. This improves the comparison of individual learning curves, since all surgeons worked under widely identical conditions. So far, studies of learning curves in stapes surgery mostly investigated only one surgeon from one hospital. Comparisons of different learning curves therefore were only possible across organizations. With data from an 18-year period, the study is also able to investigate learning on organizational (otorhinolaryngology department) level.

The remainder of this paper is organized as follows: In section 3.2, the hypotheses are developed based on the literature. The data set as well as the research models are described in section 3.3. A discussion of the results and their limitations is presented in section 3.4, followed by some considerations about the robustness of estimations in section 3.5. The paper closes with a short summary and outlines some suggestions for further research in section 3.6.

3.2 Background and Hypotheses

3.2.1 OR Times as a Proxy for Costs

In the literature concerned with stapes surgery, OR times are rarely considered.[85] While considering quality, this study also takes the OR times into account; they serve as an additional short-term cost indicator. Taking short- as well as medium- to long-term cost indicators into consideration, this study can comprehensively evaluate the economic implications of learning in stapes surgery.

OR times are relevant cost indicators, since ORs are among the highest resource-utilizing facilities within the hospital, accounting for 25 % of total costs for inpatient cases (Berry et al., 2008). Bellini et al. (2020) even report a proportion of 35 to 40 % of hospital

---

[85] OR times are considered sporadically in the recent literature, which is concerned with endoscopic stapes surgery (first studied by Poe in 2000). Most of the time, only mean OR times of the studied endoscopic and microscopic interventions are reported and there is no investigation of the development over time.

costs and thus designate the OR as the financial center. Childers and Maggard-Gibbons (2018) report direct OR costs of 20 to $ 21 per minute.[86] These examples illustrate that OR times have a considerable impact on costs per case. Decreased OR times therefore can ultimately result in a fixed costs degression due to a more efficient OR use and lower costs per case.[87]

It seems likely that with rising experience in a certain type of surgery, routine emerges which improves surgeons' movements and consequently, surgeries get shorter. Among the studies considering OR times in stapes surgery, there can be found a tendency of shorter OR times over time. Lucidi et al. (2020) declare a mean OR time of 49 min. for the first ten surgeries and 42.2 min. as the average of surgery no. 91 to 100.[88] Iannella and Magliulo (2016) calculate a mean OR time of 52.5 min. in the first four months of endoscopic stapes surgery and 35.9 min. in the third four-month period.[89]

Derived from research, the hypothesis regarding OR times reads:

H_1: *Ceteris paribus, surgeons' experience leads to a decrease in the OR time in stapes surgery.*

3.2.2 Quality as a Proxy for Costs

There is some evidence from previous research that surgeons' learning and experience positively influences outcomes in stapes surgery. Hughes (1991) describes that approximatively 50 surgeries are needed to reach a postoperative ABG $\leq$ 10 dB in 90 % of the cases. Sargent (2002) comes to the the same conclusion. Watson et al. (2015) quote 43 surgeries to reach a postoperative ABG $\leq$ 15 dB in 90 % of the cases. Yung and Oates (2007) define the completion of the learning process by the curve reaching its plateau. They state 70 and 80 interventions respectively.[90]

---

[86] They investigate Californian acute care hospitals in fiscal year 2014. Labor costs are 13 to $ 14 per minute.

[87] Roberts et al. (1999) find 84 % of hospital costs to be fixed costs. 20.9 % of hospital costs are fixed costs incurred for physicians, which is about a quarter of all fixed costs. This emphasizes the relevance of labor costs.

[88] They study the first 100 endoscopic stapes surgeries of a single surgeon. The differences are however not significantly different from zero. A possible explanation might be the surgeon having gained experience in stapes surgery as well as endoscopic ear surgery before.

[89] They examine stapes surgeries conducted by the second author. Endoscopic stapes surgery has been introduced in their department 18 months before the start of the observation period.

[90] The authors study their own first 100 primary stapes surgeries.

In this analysis, the success of stapes surgery is employed by measuring the postoperative ABG in order to measure the quality consequences of surgeons' learning and experience.[91] For the calculation of the postoperative ABG, the first audiogram after surgery is used in this paper. This will yield conservatively measured, comparatively poor results as hearing tends to improve (i.e. the ABG is reduced) in the time following surgery. For example, Nash et al. (2021) report significant improvements between six weeks and six months after surgery. This tendency can also be detected in the first postoperative weeks in the data set at hand: The mean ABG of audiograms made on day 7 after surgery is 17.51 vs. 12.45 dB on day 28. On average, the ABGs on day 28 are significantly smaller than the ones on day 7.[92] The mean ABG of audiograms up to day 28 is in turn significantly larger than the one of audiograms from day 29 on (15.72 vs. 10.56 dB).[93] On average, postoperative audiograms are made 61 days after surgery. Other studies use postoperative audiograms with a longer time lag to surgery, for example Yung and Oates (2007) use audiograms from minimum six months after surgery, Sergi and Paludetti (2016) measure the ABG after 24 months or later. Therefore, the results in this paper seem to be worse at first sight. Besides solely relying on the postoperative ABG, this paper employs a success criterion which combines the postoperative ABG with the absence of revisions. Quality in turn serves as a medium- to long-term cost indicator, thus it accomplishes two objectives. First, it can answer the question how learning and experience influences outcomes with regard to quality. Second, it can link these results to the follow-up costs of the intervention.

Quality in stapes surgery is linked to treatment costs because negative outcomes (no considerable improvement or deterioration of conductive hearing loss) as well as complications (e.g. sensorineural hearing loss) are typically associated with higher resource use due to higher follow-up costs. Cases with negative outcomes and complications in primary surgery likely require a revision surgery, which considerably increases costs per case. To make matters worse, the success rate in revision has been reported lower than in primary surgery for decades with success rates between 24 % and 80 % (Glasscock et al., 1995;

---

[91] There are mainly two methods of measuring success in stapes surgery, postoperative hearing gain and postoperative ABG. However, defining success via the postoperative ABG predominates. Therefore, this method is used to evaluate success and quality. On the basis of Hughes (1991) who defines competence, i.e. the accomplishment of the learning curve, as a 10 dB postoperative ABG in 90 % of the patients, a successful intervention is defined by a postoperative ABG $\leq$ 10 dB, corresponding to the most widely accepted criterion of success (Lovato et al., 2019). Besides, there must not be a revision.

[92] P-value = 0.0004.

[93] P-value = 0.0000.

Blijleven et al., 2019) compared to a range from 72 % to 94 % in primary surgery (Wegner, 2018). This means if primary surgery is unsuccessful and revision costs incur, there is a higher probability of further follow-up costs compared to the primary intervention. A higher risk of hearing loss, inner ear damage and vertigo in revision even intensifies the problem (Durko et al., 2007).

There is distinct proof that complications are a proxy for higher costs. Klask and Schmelzer (2003) calculate the costs of a stapes surgery at a German university hospital, which amount to 964 € for a case without complications. For a case with inner ear hearing loss and vertigo (no revision surgery), they calculate 1,754 € and thus a cost increase by 82 %.[94] For a case with perilymphatic fistula, which was revised by surgery, the costs account for 3,504 €, meaning a cost increase by 262 %.

Most complications occur with cases simultaneously failing to reach a postoperative ABG $\leq$ 10 dB. However, there are cases regarded successful based on the ABG in which complications occur. Vincent et al. (2006) report postoperative ABG $\leq$ 10 dB in 94.2 %, suggesting a failure rate of 5.8 %. However, they report a rate of 6.6 % for failures.[95] This means 0.8 % of all cases have a postoperative ABG $\leq$ 10 dB, though they suffer from complications. To put it differently, 12 % of complications occur in seemingly successful cases (postoperative ABG $\leq$ 10 dB). By using the success criterion, cases are regarded as unsuccessful if they have to be revised, irrespective of the ABG.

Even if it is assumed that there are no immediate cost effects of negative outcomes and complications in stapes surgery, i.e. primary surgery costs are not affected, it has been argued that unsuccessful surgeries and complications, even the ones which do not require a revision surgery, may have a large impact on follow-up and thus overall costs. To conclude, surgeons' learning and experience are supposed to positively affect quality in stapes surgery, which in turn negatively affects costs. Thus, surgeons' learning is supposed to negatively influence costs.

As a result, the hypothesis regarding quality follows:

H_2: *Ceteris paribus, surgeons' experience leads to an increase in quality in stapes surgery.*

---

[94] Though, one patient still suffered from inner ear hearing loss on the day of discharge. This means this case probably had further follow-up costs not taken into account in the calculation.

[95] They subsume postoperative ABGs $>$ 10 dB and complications under the term "failure".

3.3 Data and Estimation Models

3.3.1 Information on the Data

In this analysis, an anonymous data set from the otorhinolaryngology department of a German tertiary care hospital is used. The data set contains 1,066 interventions by 24 surgeons from 2003 until 2020, which are employed to examine and compare the learning curve effects in stapes surgery. Nine out of these 24 surgeons are contained in an individual investigation as each of them has at least 25 surgeries in the data set.[96] Five out of these nine are experienced surgeons having conducted many stapes surgeries before the observation period, so they are called experts in the following. The other four surgeons are rather inexperienced in stapes surgery, they are subsequently called beginners. However, as beginners gather experience in stapes surgery, there comes the time when they cannot be called beginners anymore. Thus, beginners can have at most an experience of 99 surgeries; with an experience of 100 or more, they are regarded as experts in this study.[97] One surgeon who is initially a beginner therefore becomes an expert within the observation period.[98] The number of stapes surgeries in the department decreased over the observation period from 113 in 2003 to 36 in 2020. The indication for surgery was usually given by an ABG $\geq 25\,\text{dB}$.[99]

The data set was collected with the help of Würzburg Ear Forms, a system to precisely document ear surgery and follow-up evaluation within the otorhinolaryngology department of the university hospital Würzburg. The system is implemented in ENTstatistics from INNOFORCE Est.[100]

3.3.2 Estimation Models – OR Times

The following regression model is supposed to explain the OR times for surgeon i:[101]

---

[96] Altogether, they account for 916 surgeries. Table 3.2 to 3.8 show the descriptive statistics for all relevant variables.

[97] Yung and Oates (2007) report having completed the learning curve (with regard to the postoperative ABG) after 70 and 80 surgeries respectively. This appears to be an appropriate reference as they consider the completion to be the learning curve reaching its plateau and not by reaching a predefined success rate. As a heterogeneous case mix possibly influences the learning curve (Lovato et al., 2019), the cutoff value is set higher to 99 surgeries.

[98] For an overview of the surgeons regarded as beginners and experts respectively in this study, see table 3.2.

[99] ABGs in this study are four-frequency (0.5, 1, 2 and 4 kHz) pure-tone average values.

[100] See Schön and Müller (2002) as well as Keim et al. (2014) for more details on Würzburg Ear Forms.

[101] See table 3.1 for a short description of all the variables. Models are estimated by Ordinary Least Squares (OLS) regression unless otherwise specified.

$$\begin{aligned}\text{ORT}_i = \beta_0 + \beta_1 \cdot \text{EXP}_i + \beta_2 \cdot \text{EXP}_i^2 \\ + \beta_3 \cdot \text{ECTOMY} + \beta_4 \cdot \text{REV} + \beta_5 \cdot \text{OOR} + \varepsilon\end{aligned} \tag{3.1}$$

In model 3.1, the surgeon's experience is measured via the sum of stapes surgeries surgeon i has conducted. The squared experience is used in order to represent a possibly nonlinear relation between OR times and experience. Learning curves are usually nonlinear, with the initial effect of learning being largest and decreasing with the level of experience. To control for possible differences in the OR time between the methods of stapes surgery, stapedotomy and stapedectomy, ECTOMY is employed as a dummy variable for a stapedectomy instead of a stapedotomy.[102] As already broached in section 3.2, there are considerable differences between primary and revision stapes surgery with regard to success. More challenging interventions may also be reflected in larger OR times. Therefore, it is controlled for revisions using the dummy variable REV. OOR is a count variable of other operations conducted simultaneously in addition to the stapes surgery.[103] The term ε represents the error term.

To distinguish between beginners and experts, separate estimation models are set up:[104]

$$\begin{aligned}\text{ORT}_b = \beta_0 &+ \beta_1 \cdot \text{EXP}_i + \beta_2 \cdot \text{EXP}_i^2 \\ &+ \beta_3 \cdot \text{ECTOMY} + \beta_4 \cdot \text{REV} + \beta_5 \cdot \text{OOR} \\ &+ \beta_6 \cdot \text{SUR}_4 + \beta_7 \cdot \text{SUR}_5 + \beta_8 \cdot \text{SUR}_9 \\ &+ \beta_9 \cdot \text{2007-10} + \beta_{10} \cdot \text{2011-15} + \beta_{11} \cdot \text{2016-20} + \varepsilon\end{aligned} \tag{3.2}$$

$$\begin{aligned}\text{ORT}_e = \beta_0 &+ \beta_1 \cdot \text{EXP}_i + \beta_2 \cdot \text{EXP}_i^2 \\ &+ \beta_3 \cdot \text{ECTOMY} + \beta_4 \cdot \text{REV} + \beta_5 \cdot \text{OOR} \\ &+ \beta_6 \cdot \text{SUR}_2 + \beta_7 \cdot \text{SUR}_6 + \beta_8 \cdot \text{SUR}_7 + \beta_9 \cdot \text{SUR}_8 + \beta_{10} \cdot \text{SUR}_9 \\ &+ \beta_{11} \cdot \text{2007-10} + \beta_{12} \cdot \text{2011-15} + \beta_{13} \cdot \text{2016-20} + \varepsilon\end{aligned} \tag{3.3}$$

Model 3.2 and 3.3 try to explain the OR times of interventions by beginners and by

---

[102] Stapedotomy has become the standard method.

[103] For example, ablation of hyperostosis (overflow bone) in the external auditory canal.

[104] See section 3.3.1 for the definition of beginners and experts in this study. Note that only the interventions by the nine surgeons investigated individually are considered. As the initial beginner surgeon 9 becomes an expert, SUR_9 is contained in both models. However, in model 3.2 and 3.3, only the interventions of surgeon 9 as beginner and expert respectively are used. In model 3.2, surgeon 3 serves as the reference category; in model 3.3, it is surgeon 1.

experts respectively. They use the experience of the respective surgeon i.[105] SUR_i is a dummy variable for the i-th surgeon in the data set.[106] Using these dummies, it can be controlled for differences between surgeons. It might be the case that a certain surgeon is generally faster or slower than others. Besides, the time period dummy variables 2007-10, 2011-15, 2016-20 are engaged to control for unobserved changes over time, e.g. due to improvements in equipment or technique.[107]

For learning curve effects on organizational (otorhinolaryngology department) level, the following model is estimated:[108]

$$\begin{aligned}\text{ORT} = {} & \beta_0 + \beta_1 \cdot \text{EXP} + \beta_2 \cdot \text{EXP}^2 \\ & + \beta_3 \cdot \text{ECTOMY} + \beta_4 \cdot \text{REV} + \beta_5 \cdot \text{OOR} \\ & + \beta_6 \cdot \text{SUR}_2 + \cdots + \beta_{28} \cdot \text{SUR}_{24} \\ & + \beta_{29} \cdot \text{2007-10} + \beta_{30} \cdot \text{2011-15} + \beta_{31} \cdot \text{2016-20} + \varepsilon\end{aligned} \tag{3.4}$$

The experience is measured via the sum of stapes surgeries conducted in the department, irrespective of the surgeon involved. Since it is not possible to trace the very first stapes surgery in the hospital, counting starts in the year 2003.[109] On organizational level, learning might take place by exchanging experiences among colleagues as the product of their individual learning processes.

3.3.3 Estimation Models – Quality

The following model tries to explain the ABG after stapes surgery[110] by surgeon i:[111]

---

[105] In contrast to model 3.1, which is estimated for each surgeon i and in each case, i is held constant, i varies as different surgeons are contained in model 3.2 and 3.3, which are each only estimated once. For example, if an intervention is conducted by surgeon j, EXP_j is used as EXP_i. If surgeon k conducted a surgery, then EXP_k is used.

[106] In model 3.2, surgeon 3 serves as the reference category; in model 3.3, it is surgeon 1. The reference category is not included in the regression model.

[107] The period 2003-2006 serves as the reference category.

[108] Surgeon 1 serves as the reference category.

[109] From 2003 on, cases could be accessed digitally.

[110] ABGs in this study are four-frequency (0.5, 1, 2 and 4 kHz) pure-tone average values. First postoperative audiograms are used for calculation. On average, postoperative audiograms used in this paper are made 61 days after surgery.

[111] See table 3.1 for a short description of all the variables.

$$\begin{aligned} \mathrm{ABG}_i = {} & \beta_0 + \beta_1 \cdot \mathrm{EXP}_i + \beta_2 \cdot \mathrm{EXP}_i^2 \\ & + \beta_3 \cdot \mathrm{L\_ABG} + \beta_4 \cdot \mathrm{ECTOMY} + \beta_5 \cdot \mathrm{L\_DIA} + \beta_6 \cdot \mathrm{REV} \\ & + \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{FEMALE} + \beta_{10} \cdot \mathrm{BI} + \varepsilon \end{aligned} \tag{3.5}$$

In regression model 3.5, it is controlled for the size of the preoperative ABG by employing the dummy L_ABG, which represents a large ABG $\geq$ 30 dB. This threshold is chosen, since various studies report mean preoperative ABGs of about 30 dB.[112] Research results rather indicate that large preoperative ABGs provide improved results (e.g. Marchese et al., 2009; Khorsandi A. et al., 2018).[113] There is some controversy about the influence of the method of stapes surgery: Khorsandi A. et al. (2018) find that the success rate does not differ between stapedotomies and stapedectomies,[114] whereas the results of Sargent (2002) imply different learning curves for these two alternatives: He finds that in comparison to stapedotomies, four times the number of cases is needed in stapedectomies to reach the same success rate.[115] With regard to the diameter of the piston of the prosthesis, research result also do not depict a uniform picture. There can be found a tendency toward better outcomes with larger-diameter pistons (e.g. Rosowski and Merchant, 1995; Sim et al., 2012). However, Wegner et al. (2016) fail to find such tendency in their reviewed studies. It is controlled for possible differences in ABGs by the technique used via ECTOMY as a dummy variable for a stapedectomy instead of a stapedotomy, and L_DIA, which is a dummy variable for the usage of a large-diameter piston (0.6mm). As already mentioned in section 3.2, the success rate has been found to be smaller in revision compared to primary stapes surgery. Therefore, this paper controls for revisions using the dummy variable REV. Marchese et al. (2009) as well as Khorsandi A. et al. (2018) find that the postoperative ABG is smaller in female patients. Marchese et al. (2009) also report a higher success rate in patients younger than 50 years. Iurato et al. (2007) support this result with 70 years as their threshold. To control for potential differences in the success of stapes surgery caused by the patient's characteristics, the dummy variables AGE and FEMALE are deployed. To represent a possibly disproportionately high decline in the success of stapes surgery with increasing age, i.e. a possibly nonlinear relation between age and success, the squared age (AGE^2) is also employed. BI is employed as a dummy variable for bilateral otosclerosis, i.e. both ears are affected by otosclerosis, since Khorsandi A. et al. (2018) find higher

---

[112] E.g. Vincent et al. (2006): 26 dB, Iurato et al. (2007): 32 dB, Marchese et al. (2009): 28 dB, Sergi and Paludetti (2016): 32 dB, Khorsandi A. et al. (2018): 36 dB.

[113] Marchese et al. (2009) identify success by a postoperative hearing gain $\geq$ 10 dB.

[114] Stapedectomy is riskier than stapedotomy, which has become the standard method (Wegner, 2018).

[115] The success rate for Sargent's (2002) comparison is ABG $\leq$ 15 dB in 90 % of cases.

success rates in such cases.

Analog to the OR times, separate estimation models for beginners and experts are set up for the postoperative ABG:[116]

$$\begin{aligned}\mathrm{ABG}_b = {} & \beta_0 + \beta_1 \cdot \mathrm{EXP}_i + \beta_2 \cdot \mathrm{EXP}_i^2 \\ & + \beta_3 \cdot \mathrm{L\_ABG} + \beta_4 \cdot \mathrm{ECTOMY} + \beta_5 \cdot \mathrm{L\_DIA} + \beta_6 \cdot \mathrm{REV} \\ & + \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{FEMALE} + \beta_{10} \cdot \mathrm{BI} \\ & + \beta_{11} \cdot \mathrm{SUR}_4 + \beta_{12} \cdot \mathrm{SUR}_5 + \beta_{13} \cdot \mathrm{SUR}_9 \\ & + \beta_{14} \cdot \text{2007-10} + \beta_{15} \cdot \text{2011-15} + \beta_{16} \cdot \text{2016-20} + \varepsilon\end{aligned} \tag{3.6}$$

$$\begin{aligned}\mathrm{ABG}_e = {} & \beta_0 + \beta_1 \cdot \mathrm{EXP}_i + \beta_2 \cdot \mathrm{EXP}_i^2 \\ & + \beta_3 \cdot \mathrm{L\_ABG} + \beta_4 \cdot \mathrm{ECTOMY} + \beta_5 \cdot \mathrm{L\_DIA} + \beta_6 \cdot \mathrm{REV} \\ & + \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{FEMALE} + \beta_{10} \cdot \mathrm{BI} \\ & + \beta_{11} \cdot \mathrm{SUR}_2 + \beta_{12} \cdot \mathrm{SUR}_6 + \beta_{13} \cdot \mathrm{SUR}_7 + \beta_{14} \cdot \mathrm{SUR}_8 + \beta_{15} \cdot \mathrm{SUR}_9 \\ & + \beta_{16} \cdot \text{2007-10} + \beta_{17} \cdot \text{2011-15} + \beta_{18} \cdot \text{2016-20} + \varepsilon\end{aligned} \tag{3.7}$$

In model 3.8, the study tries to explain the expected success of a stapes surgery by means of a maximum likelihood model for dichotomous dependent variables (binary logit model):[117]

$$\begin{aligned}\mathrm{E}(\mathrm{SUCC}_i \,|\, \boldsymbol{x}) &= P(\mathrm{SUCC}_i = 1 \,|\, \boldsymbol{x}) = \int_{-\infty}^{z} f(t)dt = \int_{-\infty}^{z} \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \\ z = {} & \boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \mathrm{EXP}_i + \beta_2 \cdot \mathrm{EXP}_i^2 \\ & + \beta_3 \cdot \mathrm{L\_ABG} + \beta_4 \cdot \mathrm{ECTOMY} + \beta_5 \cdot \mathrm{L\_DIA} + \beta_6 \cdot \mathrm{REV} \\ & + \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{FEMALE} + \beta_{10} \cdot \mathrm{BI}\end{aligned} \tag{3.8}$$

As discussed in section 3.2, there are seemingly successful interventions providing a postoperative ABG $\leq 10\,\mathrm{dB}$, but they are attended by complications. These cases must be regarded unsuccessful. Therefore, the dummy variable SUCC only takes the value 1 if the patient has a postoperative ABG $\leq 10\,\mathrm{dB}$ and does not undergo a revision.

Analog to the OR times and the postoperative ABG, separate estimation models for

---

[116] See section 3.3.1 for the definition of beginners and experts in this study. Note that only the interventions by the nine surgeons investigated individually are considered. As the initial beginner surgeon 9 becomes an expert, SUR_9 is contained in both models. However, in model 3.6 and 3.7, only the interventions of surgeon 9 as beginner and expert respectively are used. In model 3.6, surgeon 3 serves as the reference category; in model 3.7, it is surgeon 1.

[117] See table 3.1 for a short description of all the variables.

beginners and experts are set up for the expected success:[118]

$$\begin{aligned}
\mathrm{E}(\mathrm{SUCC}_b \,|\, \boldsymbol{x}) = P(\mathrm{SUCC}_b = 1 \,|\, \boldsymbol{x}) &= \frac{1}{1+e^{-z}} \\
z = \ \boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \mathrm{EXP}_i &+ \beta_2 \cdot \mathrm{EXP}_i^2 \\
&+ \beta_3 \cdot \mathrm{L\_ABG} + \beta_4 \cdot \mathrm{ECTOMY} + \beta_5 \cdot \mathrm{L\_DIA} + \beta_6 \cdot \mathrm{REV} \\
&+ \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{FEMALE} + \beta_{10} \cdot \mathrm{BI} \\
&+ \beta_{11} \cdot \mathrm{SUR}_4 + \beta_{12} \cdot \mathrm{SUR}_5 + \beta_{13} \cdot \mathrm{SUR}_9 \\
&+ \beta_{14} \cdot \text{2007-10} + \beta_{15} \cdot \text{2011-15} + \beta_{16} \cdot \text{2016-20}
\end{aligned} \tag{3.9}$$

$$\begin{aligned}
\mathrm{E}(\mathrm{SUCC}_e \,|\, \boldsymbol{x}) = P(\mathrm{SUCC}_e = 1 \,|\, \boldsymbol{x}) &= \frac{1}{1+e^{-z}} \\
z = \ \boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \mathrm{EXP}_i &+ \beta_2 \cdot \mathrm{EXP}_i^2 \\
&+ \beta_3 \cdot \mathrm{L\_ABG} + \beta_4 \cdot \mathrm{ECTOMY} + \beta_5 \cdot \mathrm{L\_DIA} + \beta_6 \cdot \mathrm{REV} \\
&+ \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{FEMALE} + \beta_{10} \cdot \mathrm{BI} \\
&+ \beta_{11} \cdot \mathrm{SUR}_2 + \beta_{12} \cdot \mathrm{SUR}_6 + \beta_{13} \cdot \mathrm{SUR}_7 + \beta_{14} \cdot \mathrm{SUR}_8 + \beta_{15} \cdot \mathrm{SUR}_9 \\
&+ \beta_{16} \cdot \text{2007-10} + \beta_{17} \cdot \text{2011-15} + \beta_{18} \cdot \text{2016-20}
\end{aligned} \tag{3.10}$$

In order to evaluate the learning curve effects regarding quality on organizational level, the following models are employed:

$$\begin{aligned}
\mathrm{ABG} = \beta_0 + \beta_1 \cdot \mathrm{EXP} &+ \beta_2 \cdot \mathrm{EXP}^2 \\
&+ \beta_3 \cdot \mathrm{L\_ABG} + \beta_4 \cdot \mathrm{ECTOMY} + \beta_5 \cdot \mathrm{L\_DIA} + \beta_6 \cdot \mathrm{REV} \\
&+ \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{FEMALE} + \beta_{10} \cdot \mathrm{BI} \\
&+ \beta_{11} \cdot \mathrm{SUR}_2 + \cdots + \beta_{33} \cdot \mathrm{SUR}_{24} \\
&+ \beta_{34} \cdot \text{2007-10} + \beta_{35} \cdot \text{2011-15} + \beta_{36} \cdot \text{2016-20} + \varepsilon
\end{aligned} \tag{3.11}$$

---

[118] See section 3.3.1 for the definition of beginners and experts in this study. Note that only the interventions by the nine surgeons investigated individually are considered. As the initial beginner surgeon 9 becomes an expert, SUR_9 is contained in both models. However, in model 3.9 and 3.10, only the interventions of surgeon 9 as beginner and expert respectively are used. In model 3.9, surgeon 3 serves as the reference category; in model 3.10, it is surgeon 1.

$$
\begin{aligned}
&\mathrm{E}(\mathrm{SUCC}\,|\,\boldsymbol{x}) = P(\mathrm{SUCC}=1\,|\,\boldsymbol{x}) = \frac{1}{1+e^{-z}} \\
&z = \ \boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 \cdot \mathrm{EXP} + \beta_2 \cdot \mathrm{EXP}^2 \\
&\qquad + \beta_3 \cdot \mathrm{L\_ABG} + \beta_4 \cdot \mathrm{ECTOMY} + \beta_5 \cdot \mathrm{L\_DIA} + \beta_6 \cdot \mathrm{REV} \\
&\qquad + \beta_7 \cdot \mathrm{AGE} + \beta_8 \cdot \mathrm{AGE}^2 + \beta_9 \cdot \mathrm{FEMALE} + \beta_{10} \cdot \mathrm{BI} \\
&\qquad + \beta_{11} \cdot \mathrm{SUR}_2 + \cdots + \beta_{33} \cdot \mathrm{SUR}_{24} \\
&\qquad + \beta_{34} \cdot \text{2007-10} + \beta_{35} \cdot \text{2011-15} + \beta_{36} \cdot \text{2016-20}
\end{aligned}
\tag{3.12}
$$

3.4 Results

3.4.1 OR Times

Before the estimation results regarding the OR times are examined, the OR times of consecutive stapes surgeries by individual surgeons are beheld. The plotted OR times in figure 3.1 depict a tendency of shorter OR times with a rising number of cases operated before, i.e. there seems to exist a learning curve. The effect of learning on OR times seems

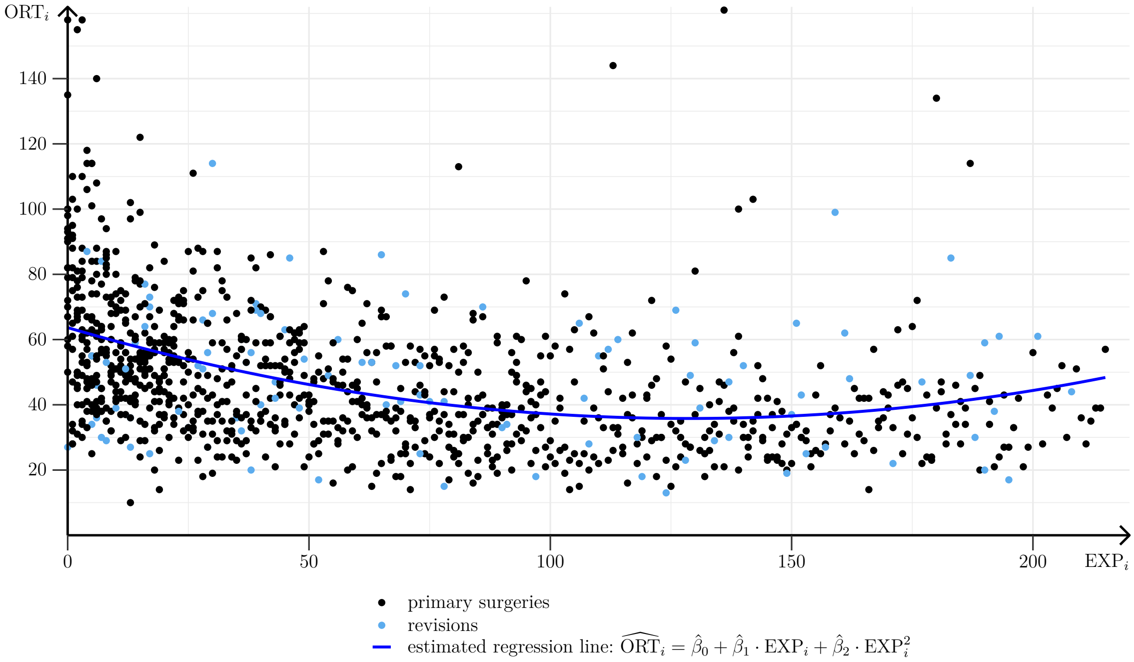

Figure 3.1: OR Time and Experience of Respective Surgeon

to feature the already mentioned declining development at decreasing rates. Figure 3.1 contains the regression line from a regression with ORT being regressed on EXP as well as EXP^2.[119] It illustrates the seemingly nonlinear relation. The estimated regression line has a negative slope, the marginal effect of an additional stapes surgery decreases with increasing experience until the marginal effect is zero. After this turning point, the OR times increase with increasing experience at increasing marginal effects. Longer OR times with increasing experience—at an already high level of experience—might be a result of cases known to be more difficult in advance being allocated to the most experienced surgeons. For the revisions displayed in figure 3.1, the OR times scatter without a trend.[120] The share of revisions is higher with higher levels of experience, which indicates that the more difficult revisions (compared to primary surgeries) are allocated to very experienced surgeons.[121]

As a next step, it is differentiated between beginners and experts.[122] Figure 3.2 and 3.3 contain a scatterplot for beginners and experts respectively. The learning curve is

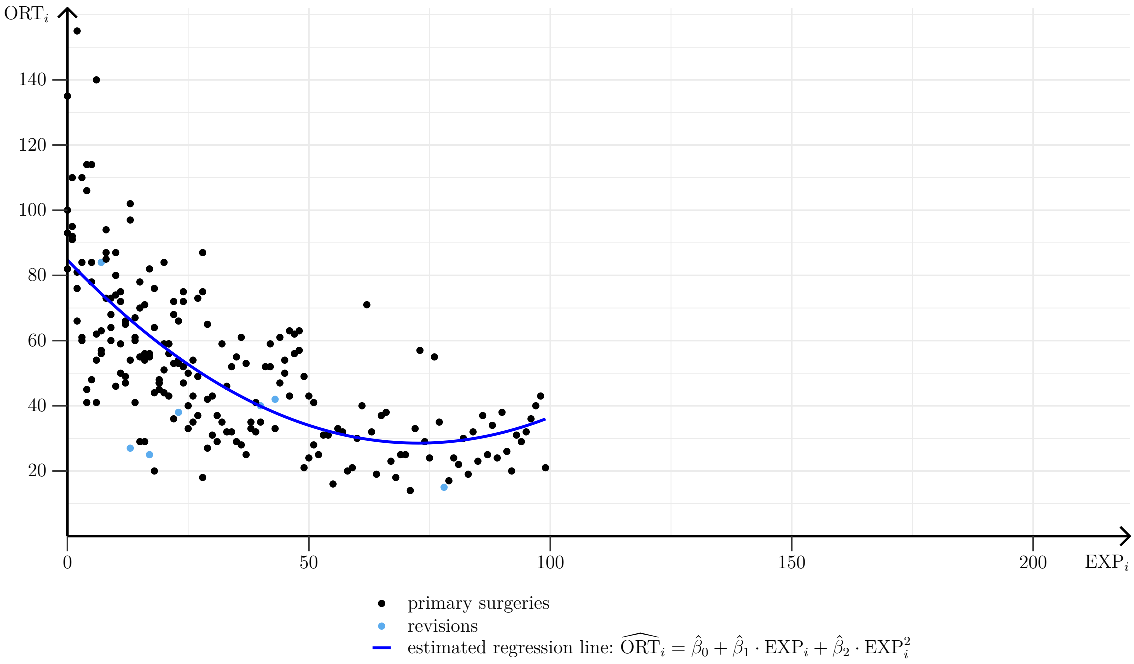

Figure 3.2: OR Time and Experience of Respective Surgeon – Beginners

---

[119] See figure 3.9 and 3.10 for a separate scatterplot for primary surgeries and revisions respectively.

[120] See figure 3.10 for a separate scatterplot for revisions.

[121] Lovato et al. (2019) discuss the same phenomenon.

[122] See section 3.3.1 for the definition of beginners and experts in this study. Note that only the interventions by the nine surgeons investigated individually are considered.

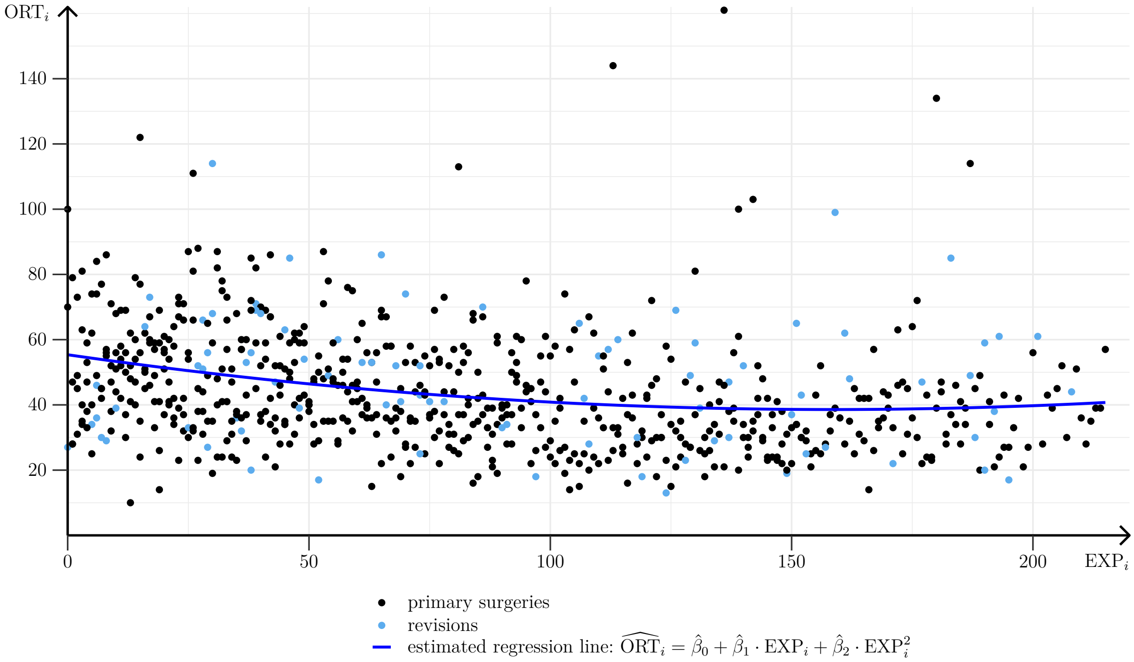

Figure 3.3: OR Time and Experience of Respective Surgeon – Experts

strikingly more distinct for the beginners than for the experts. The estimated regression line for the beginners has a highly negative slope compared to the estimated regression line in the overall scatterplot in figure 3.1. The marginal effect of an additional stapes surgery is initially relatively high. For the experts, there is virtually no learning curve anymore.

Table 3.9 depicts the estimation results of model 3.1. The surgeon's experience has a negative effect on the OR time with six out of nine surgeons.[123] In three of these six regressions, the squared experience has a positive influence on the OR time, which means the algebraic sign as well as the height of the marginal effect of experience depend on the level of experience here.[124] Initially, the absolute value of the negative marginal effect gets smaller with experience and after having become zero, the marginal effect is positive and gets larger with experience. For example, the turning point for surgeon 1 is at an experience of about 66 surgeries. In summary, H_1 is supported, and therefore H_1 cannot be rejected on individual level.

---

[123] All coefficients which are interpreted are significantly different from zero (in the following "significant" for short).

[124] The two regressors representing experience have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors if necessary ($i = \{1, 4, 6\}$). P-value < 0.01 / rejection of "no joint significance" in all cases except $i = \{7, 8\}$ (p-value > 0.1).

The insignificant coefficients of ECTOMY imply that, regarding the method of stapes surgery, there are no differences in OR times between stapedotomy and stapedectomy. Somewhat surprisingly, OR times do not significantly differ between primary surgeries and revisions in all but three regressions, as the coefficients of REV are insignificant in these cases.[125] In case of surgeon 2, who has the highest percentage of revisions, a stapes revision on average and ceteris paribus takes about 9.4 minutes longer than a primary stapes surgery.[126] Astonishingly, revisions by surgeon 4 and 9 are on average about 7.5 and 8.1 minutes respectively shorter than primary surgeries. In four out of nine regressions, the coefficient of OOR is positive, which means the OR time gets longer when another operation is conducted simultaneously besides the stapes surgery. The increase for an additional intervention varies from about 3.9 to about 18.1 minutes. The large range might also be a product of the allocation of surgeries, which is also reflected in quantitative differences in OOR between surgeons.[127]

Table 3.10 depicts the estimation results of model 3.2 and 3.3. Both beginners' and experts' experience has a negative effect on the OR time. In both regressions, the squared experience has a positive influence on the OR time, which means the algebraic sign as well as the height of the marginal effect of experience depend on the level of experience.[128] The absolute values of the marginal effect as well as the changes of marginal effects with rising experience are considerably larger with beginners.[129] This matches previous explanations as well as learning curve theory: With rising experience, the absolute reduction caused by a further unit gets smaller.[130] The turning point (from negative to positive marginal effect) for beginners is at an experience of about 52 surgeries, the one for experts at about

---

[125] It has to be considered that the percentage of revisions is very low for some surgeons. See table 3.2.

[126] OLS regression coefficients have to be interpreted as the average marginal effect of the respective regressor with all other regressors being equal ("ceteris paribus"). In order to avoid repetition and to keep explanations short, this is not always repeated in the following.

[127] See table 3.3 for the quantitative differences in OOR.

[128] The two regressors representing experience have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors if necessary (model 3.2). P-value < 0.001 in both cases.

[129] Note that an experience of zero denotes different points on the overall learning curve of an individual surgeon. In case of a beginner, it implies that the surgeon has not gathered experience in stapes surgery. In case of an expert, the surgeon has just gathered "sufficient" experience to not be classified as beginner anymore. The marginal effect with an experience of one surgery is -1.40 for a beginner and -0.18 for an expert (i.e. the absolute value is about eight times larger with beginners). With an experience of 20 surgeries, it is -0.87 and -0.16 respectively. Note that the change in the marginal effect with a change in experience from 10 to 20 surgeries reduces the absolute value of the marginal effect by 0.53 and 0.02 respectively.

[130] In business management, learning curve theory is often used in production context, hence the unit is frequently the piece of a commodity. In this context, a unit corresponds to a stapes surgery.

221.[131]

Table 3.11 depicts the estimation results of model 3.4. Experience within the otorhinolaryngology department has a negative effect on the OR time. The squared experience has a positive influence on the OR time.[132] For the number of observations in the data set, the negative marginal effect gets smaller with experience.[133] This means there can be detected nonlinear effects even after decades of stapes surgery. H_1 is supported and therefore can be rejected on organizational level neither.

There cannot be found differences between stapedotomy and stapedectomy as well as primary surgeries and revisions. With regard to the surgeon dummies, 12 coefficients are observed to be positive and 8 to be negative (ranging from –30.8 to 53.0) which implies differences in OR times between the surgeons. For example, the coefficient of the dummy variable for surgeon 2 is –12.9 which means if ceteris paribus surgeon 2 conducts the intervention instead of surgeon 1, the OR time reduces by 12.9 min.[134] Although there are time differences between surgeons, they should not be overrated, but interpreted cautiously. For example, a surgeon with higher OR times than surgeon 1 must not be regarded as “slow”. As already discussed, interventions are palpably not allocated at random. This indicates unobserved differences between individual surgery portfolios with regard to complexity and difficulty which cannot be controlled for with the regressors. Thus, a direct comparison of OR times is virtually impossible.

To conclude this section, the approximate economic implications of the learning curve effects are considered.[135] Using the DRG system for 2021, a stapes surgery without highly

---

[131] These estimated turning points represent the different parts of the learning curve beginners and experts respectively are in. Beginners are in the “early part” where marginal effects are relatively large and the learning curve is considerably bent, whereas experts are in the “late part” where marginal effects are relatively small and the curve is only slighty bent. With relatively large marginal effects, the turning point is reached faster than with relatively small ones. Note that these estimated turning points are different from the ones illustrated in the figures, since those are derived from a regression on experience (EXP and EXP^2) only.

[132] The two regressors representing experience have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors. P-value < 0.001.

[133] The algebraic sign as well as the height of the effect depend on the level of experience in the department. With the number of observations, the turning point (negative to positive marginal effect) is not reached.

[134] See table 3.4 for the average OR times of the individually investigated surgeons.

[135] The following calculations are based on values of the Diagnosis-Related Groups (DRG) system for 2021 (Institut für das Entgeltsystem im Krankenhaus, 2020, 2021; GKV-Spitzenverband, 2021; § 15 KHEntG in version from July 11, 2021), not on data from the focal hospital. These values are average values of the respective DRG from all hospitals that provided data for the calculation of the DRG system 2021.

severe complications is probably classified as DRG D30A.[136] The resulting DRG lump-sum compensation is 2,964.65 €.[137] Besides, there is a compensation for nursing staff. It amounts to 408.39 € for a three-day stay.[138] The overall compensation then would amount to 3,373.04 €. Since the DRG system is based on the remuneration of actual costs, this amount can also be considered the average costs of D30A.[139] If 31 % of the costs can be traced back to the OR and are sensitive to OR time,[140] these would equal 919.04 €. Given that the DRG compensation is suited to average surgeries, it is assumed that the mean OR time in the data set approximately equals the mean OR time of D30A surgeries considered for the DRG system 2021. Average OR time for the data set is 48.05 min., which means revenues/costs of 19.13 €/min. To illustrate what ignoring the learning curve effects might cause, surgeon 7 is considered.[141] The surgeries by surgeon 7 have a mean OR time of 35.21 min.[142] Calculating the 196 surgeries by surgeon 7 with 48.05 min. (mean OR time of data set) instead of 35.21 min. means additional 2,516.64 min. Using the OR for other interventions during the time it is unused would yield 48,143.32 € when calculating evenly with 19.13 €/min. Another approach is to consider a single stapes surgery with the average OR time by surgeon 7. As computed above, the DRG compensation for D30A regarding costs affected by OR time with mean OR time adds up to 919.04 €. The mean OR time of surgeries by surgeon 7 is 35.21 min. Using the OR for other interventions during the time it is unused would mean that only 673.57 €[143] are incurred by the stapes surgery. This would result in a profit of 245.47 €.[144] Given the overall compensation (in

---

[136] It has to be considered that several surgeries can result in the same DRG. However, DRGs are supposed to comprise cases with comparable effort. Stapes surgeries account for 10.51 % of DRG D30A cases.

[137] The evaluation relation of DRG D30A is 0.791 in 2021. To determine the actual compensation, the federal base rate, which is 3,747.98 € in 2021, is multiplied by the evaluation relation. There is a specific base rate for each federal state; however, in order to keep the computations as universal as possible, the federal amount is used.

[138] The evaluation relation for nursing staff of DRG D30A is 0.8347 in 2021. To determine the actual nursing staff compensation per day, the nursing staff base rate is multiplied by the evaluation relation. The nursing staff base rate is determined individually for each hospital. Here, the pre-determination default, which is 163.09 € per day, is used. The nursing staff compensation per day is multiplied by the days of stay (the day of discharge is not considered). Here, the rounded average length of stay for D30A according to the DRG system 2021, three days, is used.

[139] The hospitals which provided data for the calculation of the DRG system 2021 on average have these costs.

[140] These costs comprise labor and infrastructure costs. They account for about 31 % of D30A according to the DRG system 2021. Material costs, e.g. the costs of the stapes prosthesis, are not changed by shorter or longer OR times.

[141] Surgeon 7 seems appropriate because the share of revisions is about the same as for the whole data set. See table 3.2 and 3.7.

[142] See table 3.4.

[143] The per-minute rate 19.13 €/min. is multiplied by 35.21 min.

[144] The assumption is that apart from the OR time profit, DRG reimbursement matches actual costs.

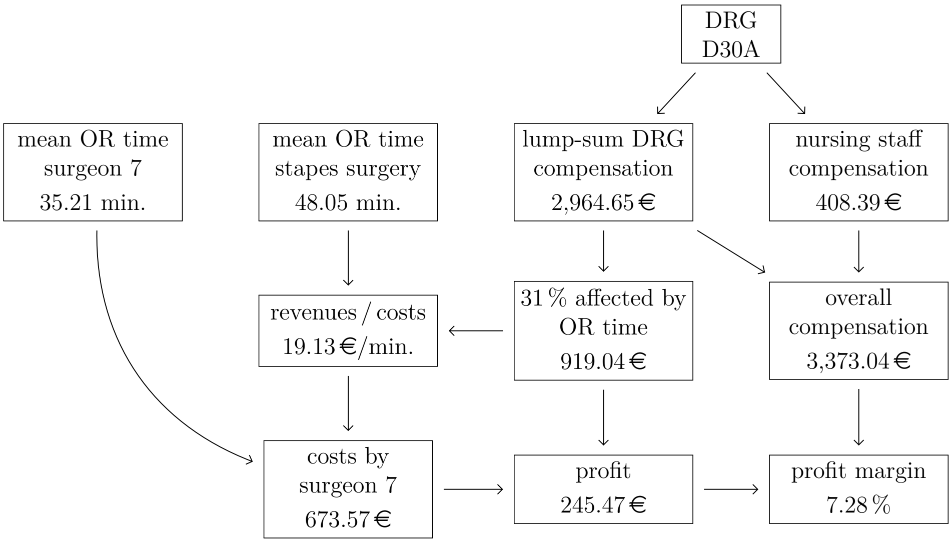

Figure 3.4: Economic Implications of Improved OR Usage

case of three days of stay), the profit margin is 7.28 %. Figure 3.4 illustrates this per-surgery consideration. The unoccupied time, e.g. for emergencies, would not change, since the overall planned OR usage would do so neither. The economic implications might seem moderate. But taking into account that the DRG system is basically designed for hospitals to reach the break-even point and comes along with a high cost pressure,[145] the possible profits by considering learning curve effects and improving OR usage are respectable. They can determine surplus or shortfall of the hospital. It is blatantly obvious that (mean) OR times cannot be estimated that exactly in reality and scarce scheduling might result in an overload of the OR (no time left for emergencies or even no time left to conduct all planned interventions). However, even if a risk-averse planning is made, utilization of the OR can be improved and idle time costs can be reduced by considering learning curve effects. A first, rough implementation might just differ between surgeons with low-, medium-, or high-level experience and scheduling the stapes surgery accordingly with large, medium, or short OR time, adjusted by the complexity of the case.[146] Decision support systems can help refine OR scheduling.[147] In order to be able to improve planning, it is necessary to quantify learning curve effects. The paper at hand contributes to this.

---

[145] In 2019, 44 % of hospitals had an annual shortfall and only 46 % had a surplus (Blum et al., 2020).

[146] Surgeries are palpably not allocated at random. This has been discussed in this section before. Thus, complexity is supposed to be considered.

[147] For a decision support system in OR scheduling, see for example Naderi et al. (2021). For a literature review of OR scheduling, see for example Rahimi and Gandomi (2021).

3.4.2 Quality

Before starting the discussion about the estimation results concerning quality, it is reasonable to have a look at the postoperative ABG of consecutive stapes surgeries by individual surgeons. In contrast to the OR times, the ABGs in figure 3.5 do not depict a tendency with increasing experience, i.e. there does not seem to exist a learning curve with regard to postoperative ABGs.[148] The estimated regression line from a regression with ABG being regressed on EXP as well as EXP^2 in figure 3.5 illustrates this.[149] The dashed line represents the common criterion of success, i.e. a postoperative ABG $\leq 10\,\text{dB}$.[150] Successful interventions do not lie above the dashed line.[151] As already stated in section 3.2.2, the first audiograms after surgery are used for calculation of postoperative ABGs. As hearing tends to improve (i.e. the ABG is reduced) in the time following surgery, this yields conservatively measured, comparatively poor results.[152]

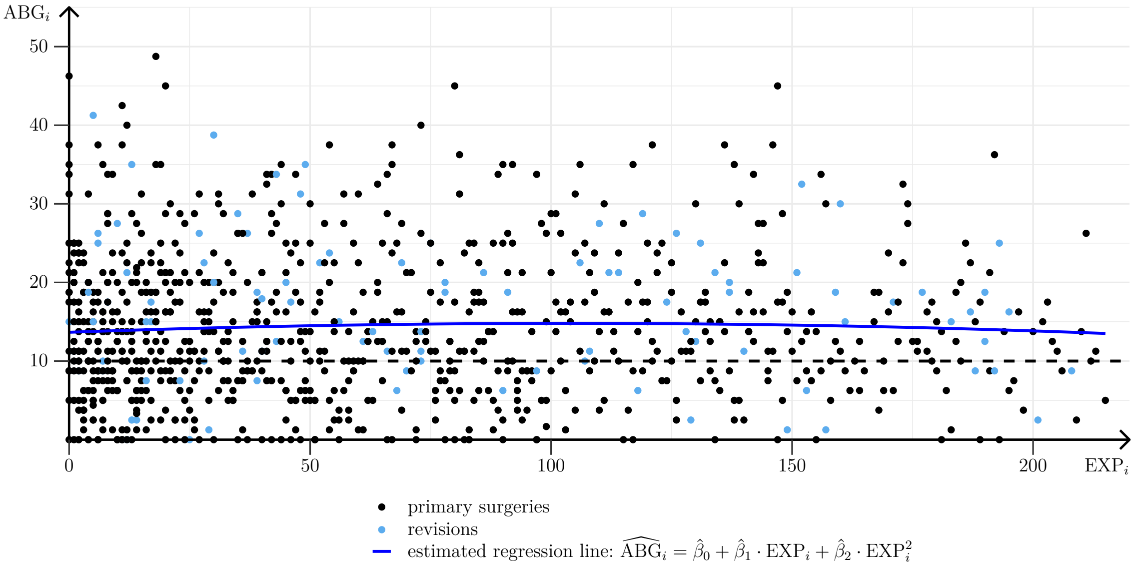

Figure 3.5: Postoperative ABG and Experience of Respective Surgeon

Figure 3.6 and 3.7 depict scatterplots for the beginners and the experts respectively.[153]

---

[148] There was one deaf ear. This observation has been excluded from the figures as well as the estimations due to a missing valid postoperative ABG.

[149] See figure 3.11 and 3.12 for a separate scatterplot for primary surgeries and revisions respectively.

[150] This aspect is elaborated on in section 3.2.2.

[151] Complications are not considered a factor determining success or failure here.

[152] See figure 3.15 and 3.16 for a separate scatterplot for primary surgeries and revisions respectively with color-coded day after surgery the audiograms are made.

[153] See section 3.3.1 for the definition of beginners and experts in this study. Note that only the surgeries by the nine surgeons investigated individually are considered.

For both figures, there is no tendency as it is neither in the overall scatterplot in figure 3.5.

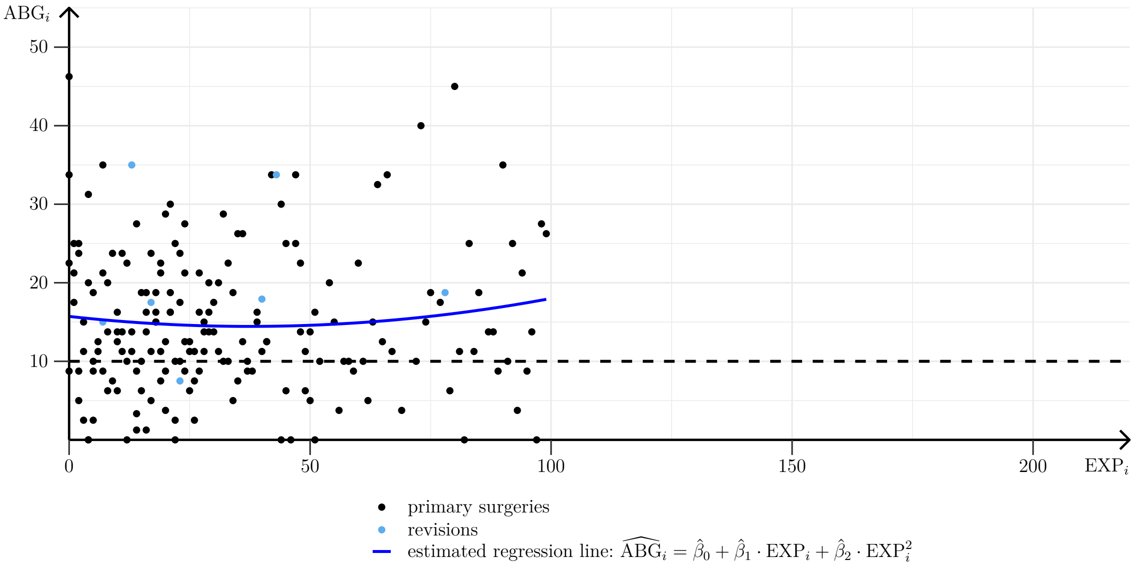

Figure 3.6: Postoperative ABG and Experience of Respective Surgeon – Beginners

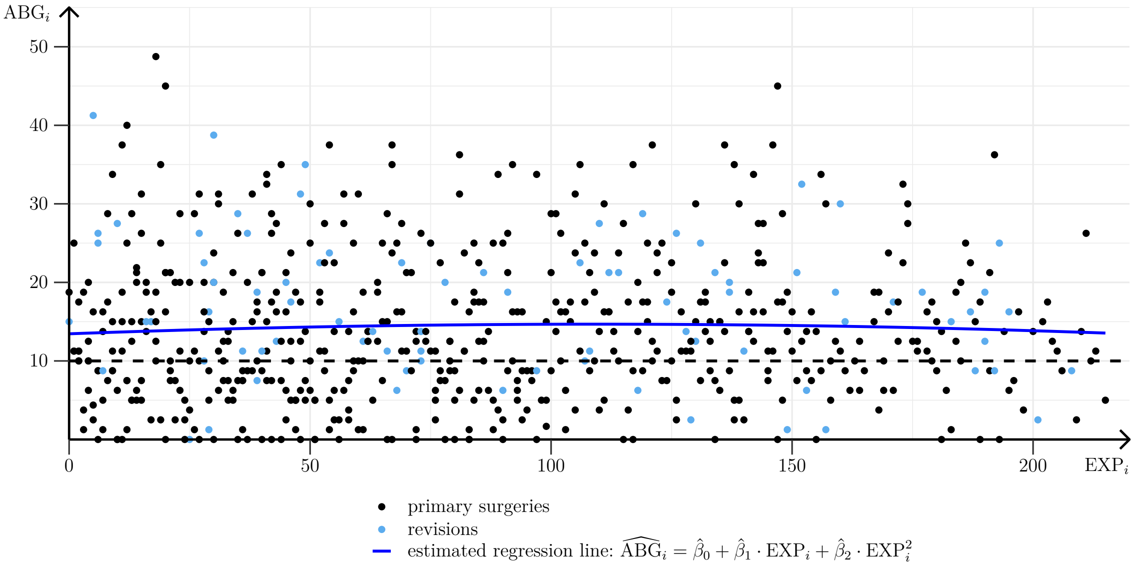

Figure 3.7: Postoperative ABG and Experience of Respective Surgeon – Experts

The estimation results of the model regarding quality on individual level measured by postoperative ABGs (model 3.5) are given in table 3.12. Experience does not have an influence on the ABG with all but one surgeon.[154] From the patient's perspective, this result is desirable, since it implies that surgery outcome measured by the postoperative ABG does not depend on the surgeon's experience.[155] This in turn supports the assumption that the surgeons in the investigated otorhinolaryngology department have a reasonable level of knowledge when doing stapes surgeries on their own, for example by having assisted an experienced surgeon before.[156] In summary, H_2 is not supported, and therefore H_2 is rejected on individual level.

The preoperative ABG has an effect on the postoperative ABG only in two cases. There, a preoperative ABG of at least 30 dB is accompanied by a 7.86 and 2.55 dB respectively larger postoperative ABG. The tendency of research results indicating a negative correlation between pre- and postoperative ABGs, i.e. larger ABGs yield smaller postoperative ABGs, cannot be supported by the data. As with OR times, there cannot be found differences regarding the method of stapes surgery between stapedotomy and stapedectomy as well as regarding the diameter of the stapes piston. The tendency of better outcomes with larger-diameter pistons in research is not substantiated. Somewhat surprisingly, there are differences between primary surgeries and revisions only in two cases. There, revisions yield poorer results as the postoperative ABG is larger by 6.63 and 14.36 dB respectively.[157] Age has a significant influence on the postoperative ABG only in two cases.[158] The coefficents of AGE and AGE^2 imply an increasing postoperative ABG with increasing age up to a certain age. Beyond this age, the postoperative ABG decreases with increasing age. The turning points are at an age of about 39 and 45 years for surgeon 5 and 6 respectively. Thus, the finding only partly matches the ones by Iurato et al. (2007) and Marchese et al. (2009) who report increasing ABGs with increasing age. The results with regard to FEMALE and BI are ambiguous. In two cases, the negative coefficients of FEMALE imply a 5.23 and 3.26 dB respectively smaller ABG for female patients. This finding matches the ones by Marchese et al. (2009) and Khorsandi A. et al.

---

[154] The two regressors representing experience have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors if necessary ($i = \{6, 7\}$). P-value > 0.1 in all cases except $i = 1$ (p-value < 0.05).

[155] The single case in which the experience has an effect on the postoperative ABG seems to be an exception, especially as surgeon 1 is even an expert. It depends perhaps on the individual surgery portfolio and its differences which cannot be controlled for with the regressors.

[156] "Knowledge" is used here to clearly distinguish between experience in own surgeries ("experience") and experience apart from own surgeries. It is not restricted to theoretical knowledge.

[157] It has to be considered that the percentage of revisions is very low for some surgeons. See table 3.2.

[158] The two regressors representing age have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors if necessary ($i = \{6, 7\}$). P-value > 0.1 in all cases except $i = \{5, 6\}$.

(2018). In another case, the ABG is 6.42 dB higher for women. In three cases, a bilateral otosclerosis improves the outcome by 7.81, 2.38, and 2.58 dB respectively; in another case, it comes along with a 5.81 dB larger postoperative ABG.

The estimation results of model 3.6 and 3.7 are depicted in table 3.13. Experience does not have an influence on the ABG for both beginners and experts.[159]

Table 3.16 depicts the estimation results of model 3.11. Experience within the otorhinolaryngology department does not have an effect on the postoperative ABG either.[160] H_2 is thus not supported, and therefore H_2 is rejected on organizational level, too.

A large preoperative ABG, the method of stapes surgery (stapedotomy and stapedectomy) as well as the diameter of the stapes piston do not have an effect on the postoperative ABG. Revisions however do have such effect on departmental level: The ABG is 2.75 dB larger. Age also has an effect on the postoperative ABG. The ABG decreases with increasing age up to about 45 years. Beyond this age, the ABG increases with increasing age.[161] For women, there cannot be found any effects. On individual level, the finding concerning the effect of bilateral otosclerosis has been ambiguous. On organizational level, it yields slightly better results (1.20 dB). Regarding surgeons, eight significant coefficients (ranging from –9.01 to 7.86) imply differences between surgeons. As already discussed, these differences should be interpreted cautiously. Surgeries are palpably not allocated at random, suggesting unobserved differences between individual surgery portfolios with regard to complexity and difficulty which cannot be controlled for with the regressors. Thus, a direct comparison of postoperative ABGs is virtually impossible. There are no significant changes over time, which would possibly have been caused by improvements in equipment or technique.

The second way of measuring quality uses success, i.e. the combination of a postoperative ABG less or equal to 10 dB and the absence of revisions. Figure 3.8 illustrates the success of surgeries. There are some primary surgeries as well as revisions which lie under the 10 dB line, but which are unsuccessful.[162] The figure does not reveal a tendency

---

[159] The two regressors representing experience have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors if necessary (model 3.6). P-value > 0.1 in both cases.

[160] The two regressors representing experience have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors. P-value > 0.1.

[161] The two regressors representing age have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors. P-value < 0.05.

[162] This aspect is discussed in section 3.2.2.

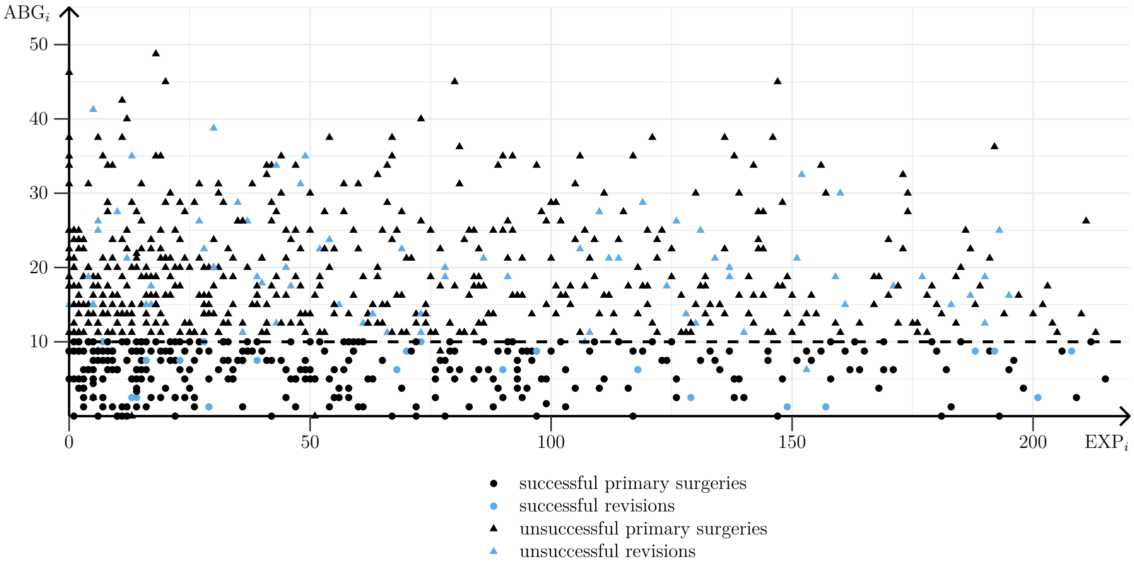

Figure 3.8: Success and Experience of Respective Surgeon

with increasing experience. Though, it depicts that the probability of success in revision is obviously smaller than in primary surgery, which is in line with research.[163]

Table 3.14 contains the estimation results of model 3.8, table 3.15 the ones of model 3.9 and 3.10. Experience also does not have an effect on the quality when being measured by the probability of a successful surgery except with one surgeon in model 3.8.[164] For model 3.12 displayed in table 3.17, it is the same: Experience in the otorhinolaryngology department does not affect the success probability.[165] H_2 is not supported, and therefore H_2 is also rejected on individual as well as organizational level when using SUCC instead of ABG.

3.4.3 Limitations

There are some limitations to the results of this study. To secure anonymity, it has not been controlled for unobserved changes over time on individual level (model 3.1, 3.5 and 3.8).[166]

---

[163] This aspect is elaborated on in section 3.2.2.

[164] The two regressors representing experience have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors if necessary (model $3.8 \,|\, i = \{2, 6, 7, 8\}$, 3.10). P-value > 0.1 in all cases except model $3.8 \,|\, i = 1$ (p-value < 0.05). The exception has already been discussed for the postoperative ABGs.

[165] The two regressors representing experience have also been tested for joint significance using an F-test with heteroskedasticity-robust standard errors. P-value > 0.1.

[166] With the five-year dummy variables, the number of stapes surgeries, and only several surgeons in each five-year period, conclusions to individual surgeons might have been possible.

However, it has been controlled for this with beginners and experts as well as on organizational level where only in some cases differences over time have been found.[167]

When comparing the absolute values of postoperative ABGs with other studies, it is important to be aware of the fact that early postoperative audiograms yield conservatively measured, comparatively poor results as hearing tends to improve (i.e. the ABG is reduced) in the time following surgery. This aspect has been elaborated on in section 3.2.2. Therefore, absolute values are difficult to compare; it is more reasonable to compare their development.

Although it has been possible to classify the nine surgeons investigated individually into beginners and experts, it has not been possible to precisely measure surgeons' initial experience. It has not been possible to control for patients undergoing a revision in another hospital. Therefore, it might be the case that few surgeries classified as successful should have been classified as unsuccessful. As already discussed, there are differences between individual surgery portfolios with regard to complexity and difficulty which cannot be controlled for with the regressors (all models except the ones for individual level, model 3.1, 3.5 and 3.8). However, it is virtually impossible to exactly represent them in a regression model.

3.5 Robustness of Estimation Models

With regard to the threshold of the postoperative ABG in the success criterion, different values have been used: It has been applied 15 dB as well as 20 dB as the threshold in model 3.8 to 3.10 and 3.12. For both values, the estimation results are almost the same as in the models with a 10 dB threshold, and the probability of success is not affected by the surgeon's experience.[168]

As an additional component of success, it has been controlled for the bone conduction in order to control for the inner ear function. For this purpose, the definition of a successful stapes surgery has been augmented with no deterioration of the bone conduction (postoperative BC $\leq$ preoperative BC) in model 3.8 to 3.10 and 3.12. The results do not relevantly differ with regard to experience; no learning effects are observed when controlling for bone conduction deterioration.[169]

---

[167] Significant coefficients for following variables: 2011-15 and 2016-20 in model 3.2; 2007-10, 2011-15 and 2016-20 in model 3.10.

[168] Exception: For surgeon 6, the experience has an influence in model 3.8.

[169] Exception: For surgeon 1, the experience has an influence in model 3.8.

As stated in section 3.2.2, another way of measuring quality is the postoperative hearing gain. Therefore, the change in the ABG after surgery (Δ ABG) has been used instead of the absolute level of the ABG after surgery in model 3.5 to 3.7 and 3.11. The results do not relevantly differ with regard to experience.

Additionally, the observation period has been divided in variously long partial periods: 3-, 4.5- and 9-year period dummies have been used (except for the models for individual level, model 3.1, 3.5 and 3.8, which do not feature time period dummies). Only several period dummies have an influence.[170] The surgeon's experience has an effect on the OR time, but not on the postoperative ABG or the probability of success. All in all, the results do not relevantly differ, so it has been sticked with the original time periods.

As an alternative to the separate models for beginners and experts (model 3.2 and 3.3 for the OR time, model 3.6 and 3.7 for the postoperative ABG, model 3.9 and 3.10 for the probability of success), comprehensive models containing a dummy variable to distinguish between beginners and experts have been estimated. The results do not relevantly differ. The experience affects the OR time, but does not affect the postoperative ABG as well as the probability of success, and there is no difference between beginners and experts indicated by the dummy variable either in OR times or in quality measured by the postoperative ABG and the probability of success.

All regression models have been controlled for heteroskedasticity by using the Breusch-Pagan test (level of significance $\alpha = 0.05$). Robust standard errors (White) have been used in case of heteroskedasticity in order to receive valid statistical inferences.

The variance inflation factors (VIFs) have been computed for all models. Most VIFs are close to 1, indicating that there are no severe problems arising from a problematically high level of multicollinearity. For EXP_i and EXP_i^2 / EXP and EXP^2, VIFs are larger, obviously resulting from a high correlation between these two regressors. Though, $\mathrm{EXP}_i^2/\mathrm{EXP}^2$ has not been dropped from the regression equations due to the theoretical foundation. Besides, both regressors still have significant coefficients (and are jointly significant) in model 3.1 to 3.4. Further exceptions are the time period dummy variables (in parts still significant coefficients). However, performing regressions without these dummies does not relevantly change the estimation results. Thus, these dummy variables have not been removed from

---

[170] Using 3-year periods: 2015-17 in model 3.10, 2015-17 in model 3.12. Using 4.5-year periods: 2012-16 (Jan. 2012 to Jun. 2016) and 2016-20 (July 2016 to Dec. 2020) in model 3.10, 2007-11 (July 2007 to Dec. 2011) and 2016-20 in model 3.12. Using 9-year periods: 2012-20 in model 3.10.

the regression in any case. Additionally, VIFs are larger than 10 for SUR_6 in model 3.3 (still significant coefficient), 3.7 and 3.10 as well as SUR_9 in model 3.2 (still significant coefficient), 3.6 and 3.9.

3.6 Conclusions

This paper contributes to the literature by providing insights into learning curve effects in stapes surgery for nine surgeons from the same otorhinolaryngology department of a German tertiary care hospital. This has improved the comparison of individual learning curves, since all surgeons worked under widely identical conditions. It has also been investigated possible learning effects in the organization comprising 24 surgeons. Moreover, it has been a comprehensive study. While engaging in learning curve effects, it has controlled for various possibly relevant factors. The basis for the systematic investigation of learning curve effects has been a considerably large data set.

It has been hypothesized that surgeons' experience in stapes surgery decreases OR times and improves quality and therefore reduces the overall costs per case. To make statements regarding how quality is affected, the postoperative ABG as well as a success criterion combining the ABG with the absence of revisions have been used. As an additional component of success, it has also been controlled for the bone conduction in order to control for the inner ear function in the robustness tests. OR times have served as the short-term cost indicator, quality as the medium- to long-term one.

Surgeons' experience has been observed to have a cost-saving effect in the short term via reduced OR times. Learning curve effects with regard to quality have not been detected. Summa summarum, experience has been found to have a cost-saving effect via a reduction of short-term costs, i.e. surgery costs. Medium- to long-term costs (costs after surgery) are not affected by experience.

Thus, the findings suggest increased training for otologists in order to flatten the learning curve in vivo and to decrease costs per case. As part of the current training concept, inexperienced otologists watch and assist during operations by experienced colleagues. It would be desirable for this to be done in as many stapes surgeries as possible. However, with a steadily decreasing number of stapes surgeries, this might pose a problem. Besides, having watched and assisted an experienced colleague, surgeons are supposed to do as much as possible on their own to fully train procedures. Simulation-based training may be a viable option. Research has found training results in favor of this option (e.g. Amin and Friedmann, 2013). Recent progress in technology even offers the opportunity to use augmented reality simulation settings. This can be an effective and simultaneously inexpensive way of training, since surgeons do not need an experienced surgeon as instructor while exercising (Nagayo et al., 2021).

Future research might obtain more precise economic implications of surgeons' learning and experience with detailed cost data precisely stating costs incurred by different cost drivers, e.g. costs per minute in the OR. There are also subsequent questions regarding OR-time management not dealt with in this study. For example, there is the aspect of underutilization of ORs. Getting aware of learning effects reducing OR times is only the first step to improved efficiency as further ones are necessary in order to make use of these insights. If the simplified assumption is made that any decrease in OR times was not considered and OR planning used constant OR times instead, the OR would be unused between surgeries and the potential would not be used at all if it is assumed that the OR team cannot sensibly use this time.

Appendix 3.A

Table 3.1: Description of Variables

Variable	Description
ABG	Postoperative air-bone gap (ABGs in this study are four-frequency (0.5, 1, 2 and 4 kHz) pure-tone average values)
ABG_i	Postoperative air-bone gap for interventions by surgeon i
ABG_b	Postoperative air-bone gap for interventions by beginners
ABG_e	Postoperative air-bone gap for interventions by experts
AGE	Patient's age
BI	Dummy for bilateral otosclerosis (both ears are affected)
ECTOMY	Dummy for stapedectomy as method of stapes surgery (instead of stapedotomy)
FEMALE	Dummy for female patient
L_ABG	Dummy for large preoperative ABG (ABG $\geq$ 30 dB)
L_DIA	Dummy for large diameter piston used (0.6 mm)
OOR	Number of other operations which are done simultaneously with the stapes surgery
ORT	Operation time (time from incision to suture; in minutes)
ORT_i	Operation time for interventions by surgeon i
ORT_b	Operation time for interventions by beginners
ORT_e	Operation time for interventions by experts
REV	Dummy for revision stapes surgery
SUCC	Dummy for successful operation (postoperative ABG $\leq$ 10 dB, no revisions)
SUCC_i	Dummy for successful operation for interventions by surgeon i
SUCC_b	Dummy for successful operation for interventions by beginners
SUCC_e	Dummy for successful operation for interventions by experts
SUR_i	Dummy for i-th surgeon $\quad \forall\, i = \{1, 2, ..., 24\}$

Continued on the next page

Variable	Description
EXP	Experience of otorhinolaryngology department in stapes surgery (measured by sum of surgeries)
EXP_i	Experience of surgeon i in stapes surgery (measured by sum of surgeries)
ε	Error term
2007-10/2011-15/ 2016-20	Dummy for the period 2007-2010/2011-2015/2016-2020

Table 3.2: Surgeons – Surgery Volume and Experience

Surgeon		1	2	3	4	5	6	7	8	9
Surgeries	n	100	94	25	52	30	216	196	49	154
REV	n	4	20	1	1	1	31	21	7	10
	p	0.0400	0.2128	0.0400	0.0192	0.0333	0.1435	0.1071	0.1429	0.0649
Status[1)]		E	E	B	B	B	E	E	E	B/E

[1)] B = beginner, E = expert, B/E = initial beginner becomes expert.

Table 3.3: Surgeons – Surgery Scope

Surgeon		1	2	3	4	5	6	7	8	9
Surgeries	n	100	94	25	52	30	216	196	49	154
OOR	$\bar{x}$	0.1000	0.0851	0.0800	0.4423	0.1667	0.0880	0.5663	0.0816	0.0584
$\text{OOR} > 0$	p	0.0800	0.0851	0.0800	0.2308	0.1000	0.0556	0.4439	0.0612	0.0390

Table 3.4: Surgeons – Surgery Duration

Surgeon	1	2	3	4	5	6	7	8	9
n (Surgeries)	100	94	25	52	30	216	196	49	154
$\overline{\text{ORT}}$ (1-10)1)	64.80	70.86	97.60	68.40	93.00	51.50	33.50	39.11	69.20
$\overline{\text{ORT}}$ (11-20)	47.70	63.40	70.30	58.30	68.10	57.70	38.50	53.40	36.30
$\overline{\text{ORT}}$ (21-30)	45.40	64.00	69.80	60.20	52.80	62.60	38.20	44.80	39.00
$\overline{\text{ORT}}$ (31-40)	46.30	63.60	—	48.20	—	62.20	42.20	45.00	30.60
$\overline{\text{ORT}}$ (41-50)	45.60	60.80	—	52.30	—	55.40	38.30	45.56	46.22
$\overline{\text{ORT}}$ (1-n)	46.18	54.45	81.12	56.88	71.30	50.44	35.21	45.71	35.21

1) Mean ORT for stapes surgery no. 1 to 10. Other values analog.

Table 3.5: Surgeons – Surgery Outcome: Postoperative ABG

Surgeon	1	2	3	4	5	6	7	8	9
n (Surgeries)	100	94	25	52	30	216	196	49	154
$\overline{\text{ABG}}$ (1-10)1)	13.33	11.72	11.75	19.22	15.00	15.83	12.71	12.58	19.53
$\overline{\text{ABG}}$ (11-20)	13.89	21.61	11.81	15.56	8.70	15.94	16.12	11.88	19.58
$\overline{\text{ABG}}$ (21-30)	18.19	10.62	13.50	15.50	12.75	13.28	13.00	10.62	13.88
$\overline{\text{ABG}}$ (31-40)	14.17	15.94	—	12.12	—	11.72	16.39	12.08	18.61
$\overline{\text{ABG}}$ (41-50)	12.36	18.75	—	21.67	—	14.12	10.62	15.97	12.55
$\overline{\text{ABG}}$ (1-n)	12.10	15.33	12.14	16.56	11.96	15.26	13.08	12.60	16.89

1) Mean postoperative ABG for stapes surgery no. 1 to 10. Other values analog.

Table 3.6: Surgeons – Surgery Outcome: Success

Surgeon	1	2	3	4	5	6	7	8	9
n (Surgeries)	100	94	25	52	30	216	196	49	154
$P(\text{SUCC}\,\vert\,\text{1-10})$1)	0.3333	0.6250	0.3000	0.2500	0.4286	0.4444	0.1667	0.5000	0.3750
$P(\text{SUCC}\,\vert\,\text{11-20})$	0.4444	0.4286	0.5556	0.1111	0.5556	0.3000	0.2000	0.5000	0.0000
$P(\text{SUCC}\,\vert\,\text{21-30})$	0.3333	0.6250	0.2000	0.2000	0.5000	0.5000	0.4000	0.6000	0.4000
$P(\text{SUCC}\,\vert\,\text{31-40})$	0.4444	0.5000	—	0.5000	—	0.5000	0.1111	0.6667	0.2222
$P(\text{SUCC}\,\vert\,\text{41-50})$	0.4444	0.2222	—	0.1111	—	0.3000	0.6000	0.4444	0.3750
$P(\text{SUCC}\,\vert\,\text{1-n})$	0.5054	0.4267	0.3750	0.2292	0.5000	0.3057	0.3902	0.5455	0.3071

1) Probability of success for stapes surgery no. 1 to 10. Other values analog.

Table 3.7: Summary Statistics of Surgeries

		2003-2006	2007-2010	2011-2015	2016-2020	Σ
Surgeries	n	392	292	221	161	1,066
ECTOMY	x̄	0.8418	0.8116	0.8643	0.9130	0.8490
REV	n	57	31	16	6	110
	p	0.1454	0.1062	0.0724	0.0373	0.1032
L_ABG	x̄	0.5969	0.4760	0.5928	0.5342	0.5535
PREOP. ABG	x̄	32.81	29.35	31.56	30.40	31.24
L_DIA	x̄	0.9515	0.9623	0.8778	0.9627	0.9409
ORT	x̄	52.38	43.64	42.68	53.39	48.05
ABG	x̄	14.54	14.31	14.73	13.08	14.29
ABG (REV)	x̄	18.50	14.51	15.50	13.54	16.60
SUCC	x̄	0.3879	0.3834	0.3430	0.4362	0.3845
SUCC (REV)	x̄	0.1569	0.3667	0.2000	0.5000	0.2451
OOR	x̄	0.0561	0.1370	0.3258	0.4534	0.1942

Table 3.8: Summary Statistics of Patients

		2003-2006	2007-2010	2011-2015	2016-2020	Σ
AGE	x̄	46.47	48.15	47.10	48.81	47.42
	σ	13.97	11.53	13.07	13.57	13.10
BI	x̄	0.3087	0.2363	0.3801	0.3168	0.3049
FEMALE	x̄	0.6111	0.6130	0.5759	0.5460	0.5944

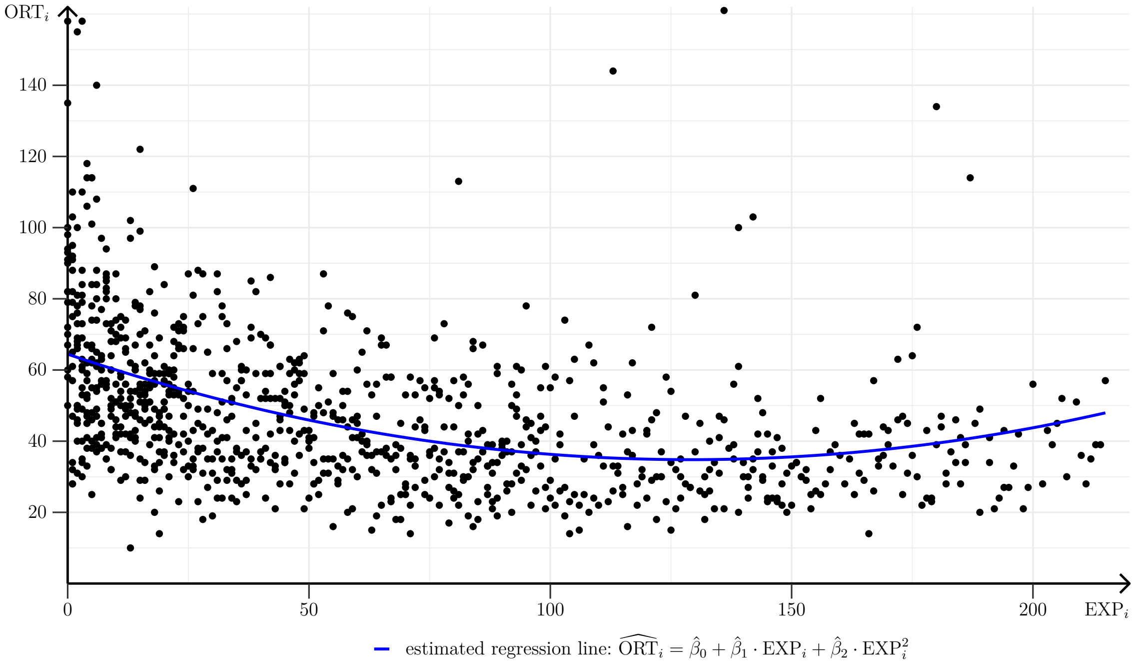

Figure 3.9: OR Time and Experience of Respective Surgeon – Primary Surgeries

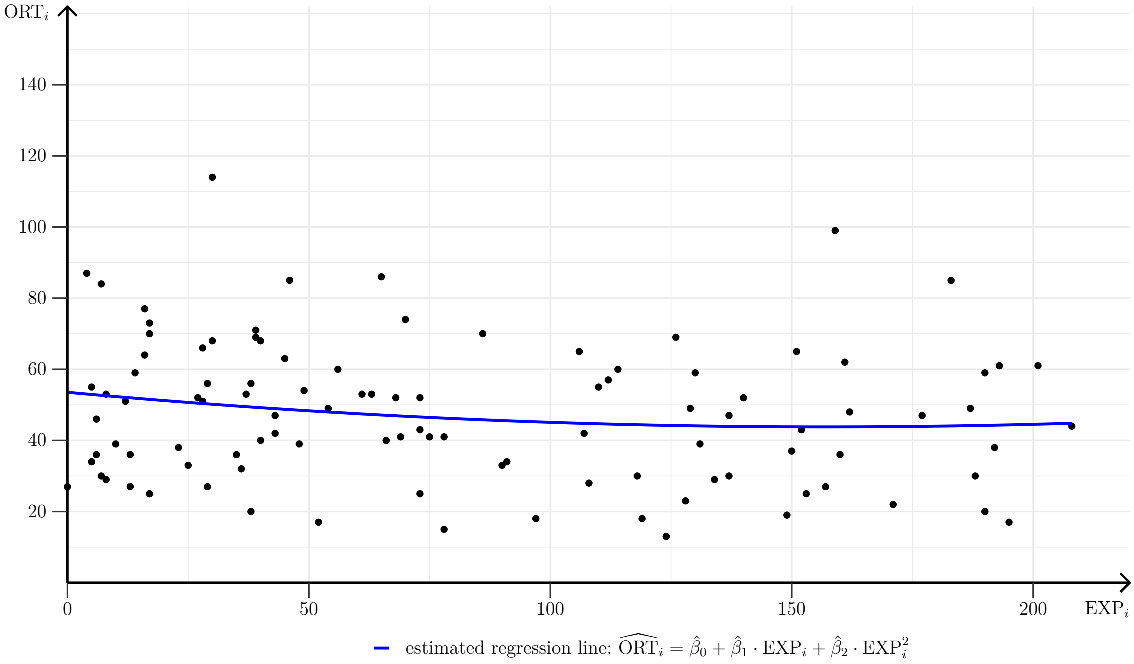

Figure 3.10: OR Time and Experience of Respective Surgeon – Revisions

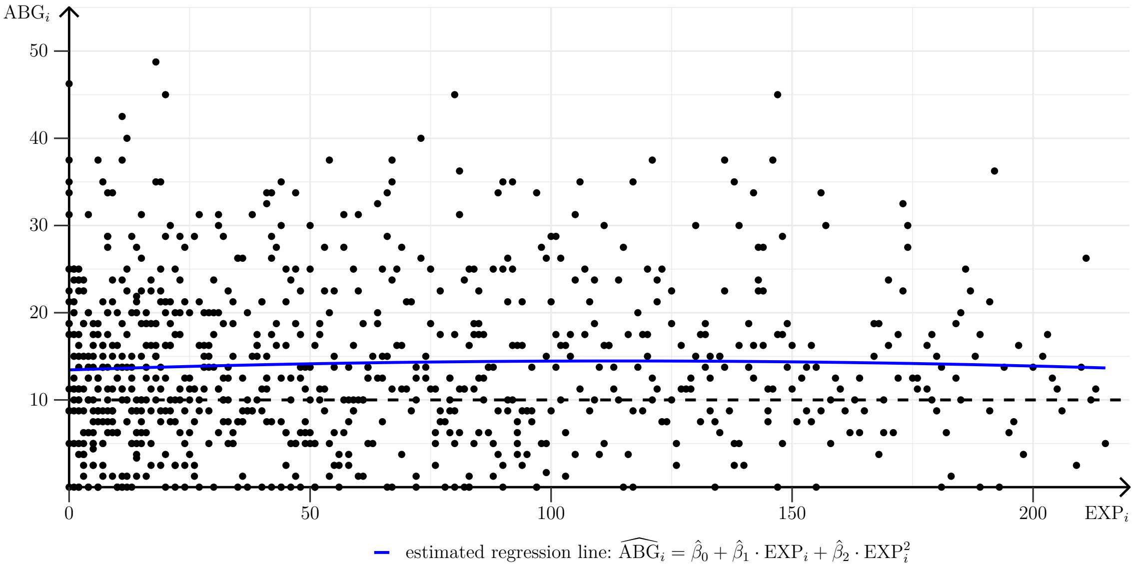

Figure 3.11: Postoperative ABG and Experience of Respective Surgeon – Primary Surgeries

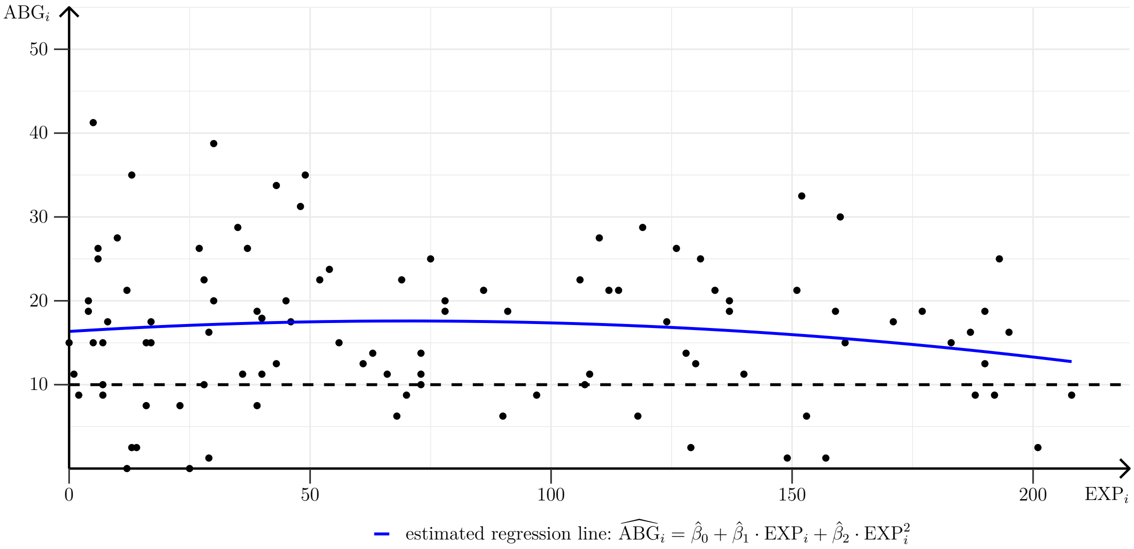

Figure 3.12: Postoperative ABG and Experience of Respective Surgeon – Revisions

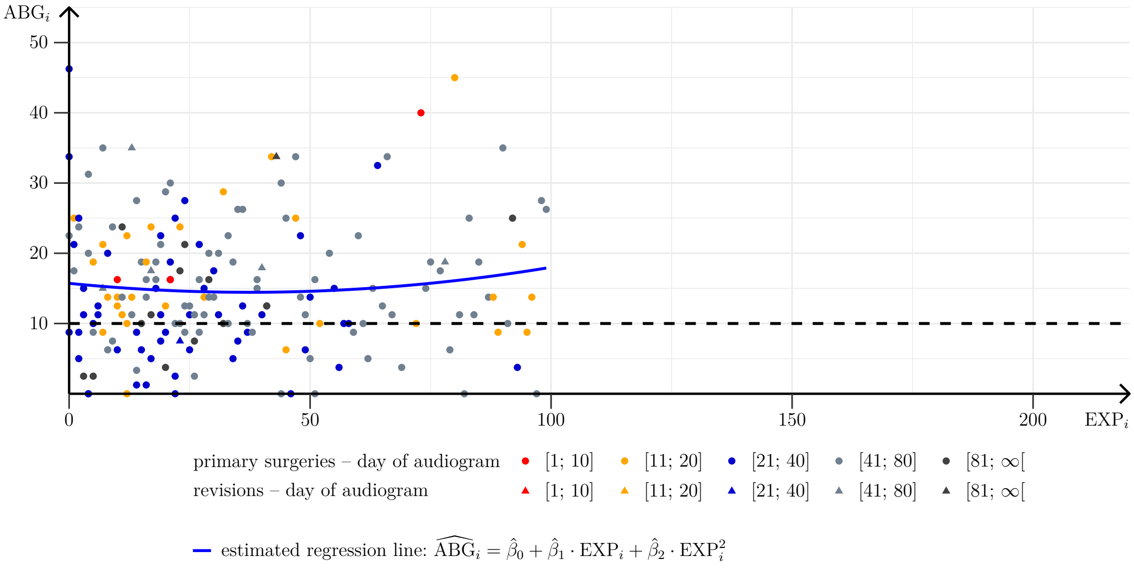

Figure 3.13: Postoperative ABG and Experience of Respective Surgeon – Beginners with Color-Coded Day of Audiogram

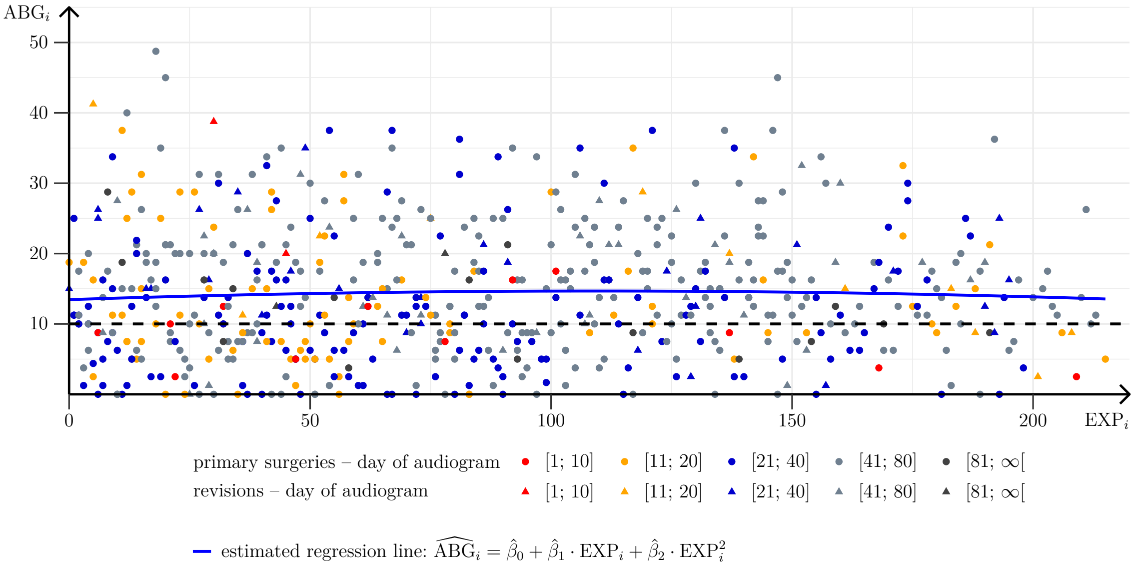

Figure 3.14: Postoperative ABG and Experience of Respective Surgeon – Experts with Color-Coded Day of Audiogram

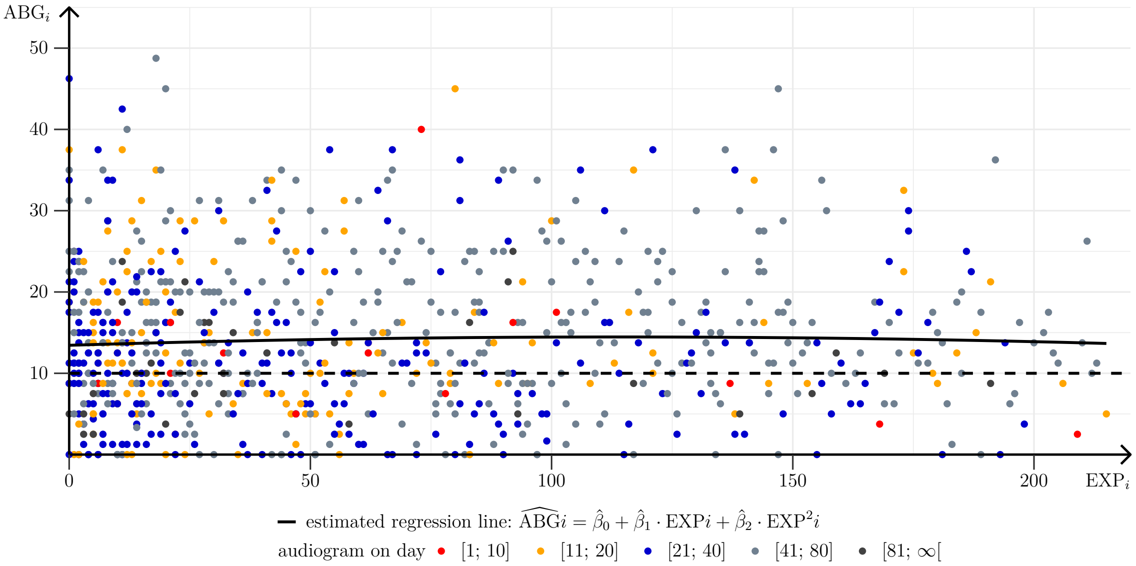

Figure 3.15: Postoperative ABG and Experience of Respective Surgeon – Primary Surgeries with Color-Coded Day of Audiogram

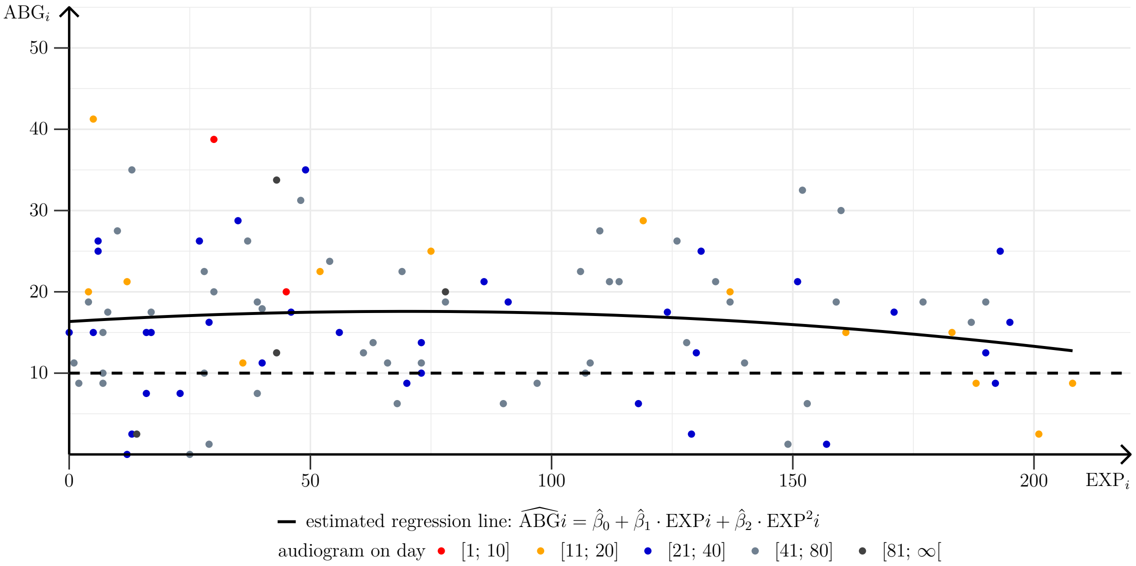

Figure 3.16: Postoperative ABG and Experience of Respective Surgeon – Revisions with Color-Coded Day of Audiogram

Appendix 3.B

Table 3.9: Estimation Results Model 3.1

Model 3.1 (ORT$_i$, OLS regression)

$\hat{\beta}_j$	$i = 1$		$i = 2$		$i = 3$	
(Intercept)	70.71704 ***	(10.09155)	71.29919 ***	(6.96510)	135.92829 ***	(14.06718)
EXP$_i$	−0.63108 **	(0.22411)	−0.64614 *	(0.25705)	−7.19968 **	(2.09254)
EXP$_i^2$	0.00476 *	(0.00196)	0.00294	(0.00260)	0.21350 *	(0.08358)
ECTOMY	−9.98176	(8.76809)	2.99415	(4.86540)	−12.90297	(9.84256)
REV	1.41155	(13.58501)	9.44610 *	(3.88685)	0.91093	(19.94320)
OOR	7.81460 *	(3.00496)	7.49565	(5.54316)	0.34658	(15.14548)

$\hat{\beta}_j$	$i = 4$		$i = 5$		$i = 6$	
(Intercept)	70.52394 ***	(8.04369)	117.40928 ***	(13.80187)	59.01622 ***	(3.79169)
EXP$_i$	−1.06044 *	(0.52182)	−3.72774 *	(1.54762)	−0.08151	(0.06954)
EXP$_i^2$	0.01292	(0.00882)	0.06120	(0.05166)	−0.00010	(0.00030)
ECTOMY	0.64792	(4.08482)	−9.06035	(12.17433)	0.13123	(2.06972)
REV	−7.46970 ***	(2.11594)	−16.98545	(19.03582)	0.40816	(3.15523)
OOR	3.85957 .	(1.93724)	−4.73170	(6.96545)	18.10493 ***	(4.61704)

$\hat{\beta}_j$	$i = 7$		$i = 8$		$i = 9$	
(Intercept)	43.33646 ***	(4.95150)	18.52922	(16.27660)	60.53858 ***	(4.73893)
EXP$_i$	−0.13092	(0.09181)	0.82206	(0.76637)	−0.71508 ***	(0.09881)
EXP$_i^2$	0.00050	(0.00044)	−0.01435	(0.01518)	0.00396 ***	(0.00063)
ECTOMY	−2.45466	(3.44503)	19.68836	(13.91558)	−2.17234	(3.69969)
REV	−6.48168	(4.08582)	−5.36344	(9.17211)	−8.08628 .	(4.58976)
OOR	1.88948	(2.19093)	3.85947	(9.57545)	14.29491 ***	(3.59327)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. For $i = \{1, 4, 6\}$, robust standard errors (White) due to heteroskedasticity.

n $\big(\text{model 3.1} \,|\, i = \{1, 2, ..., 9\}\big)$ = 100, 91, 25, 52, 30, 216, 196, 48, 153

R^2 $\big(\text{model 3.1} \,|\, i = \{1, 2, ..., 9\}\big)$ = 0.22247, 0.36032, 0.56957, 0.30548, 0.55136, 0.28123, 0.02696, 0.08211, 0.33898

Table 3.10: Estimation Results Model 3.2 and 3.3

$\hat{\beta}_j$	Model 3.2 (ORT$_b$, OLS regression)		Model 3.3 (ORT$_e$, OLS regression)	
(Intercept)	121.46087 ***	(8.98600)	54.45055 ***	(3.86443)
EXP$_i$	−1.42479 ***	(0.24318)	−0.18145 ***	(0.04701)
EXP$_i^2$	0.01378 ***	(0.00234)	0.00041 .	(0.00022)
ECTOMY	−1.95110	(3.41556)	−0.02963	(1.72108)
REV	−9.49692 *	(4.66879)	1.22336	(1.87328)
OOR	4.12213 *	(1.86387)	8.39912 ***	(1.36508)
SUR$_2$	—	—	6.49978	(4.20339)
SUR$_3$	—	—	—	—
SUR$_4$	−18.87505 ***	(4.53453)	—	—
SUR$_5$	−12.81890 *	(5.78411)	—	—
SUR$_6$	—	—	8.66319 *	(4.27706)
SUR$_7$	—	—	−10.71723 ***	(2.65784)
SUR$_8$	—	—	−5.47176	(4.22302)
SUR$_9$	−46.75053 ***	(7.04652)	−4.32654	(3.34233)
2007-10	−7.27992	(6.28949)	−1.07542	(2.88301)
2011-15	−25.66591 ***	(7.52149)	−2.60374	(4.19965)
2016-20	−19.00529 .	(9.84627)	0.53222	(5.20841)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. In model 3.2, robust standard errors (White) due to heteroskedasticity.

n (model 3.2) = 206, n (model 3.3) = 705

R^2 (model 3.2) = 0.65533, R^2 (model 3.3) = 0.26328

Surgeon reference category (model 3.2) = surg. 3, surgeon reference category (model 3.3) = surg. 1

Table 3.11: Estimation Results Model 3.4

$\hat{\beta}_j$	Model 3.4 (ORT, OLS regression)		
(Intercept)	77.35807	***	(4.28446)
EXP	−0.06745	***	(0.01504)
EXP^2	0.00003	*	(0.00001)
ECTOMY	−0.03682		(1.53103)
REV	0.37809		(1.72894)
OOR	8.60961	***	(1.42202)
SUR_2	−12.85576	***	(3.02989)
SUR_3	39.22938	***	(5.06598)
SUR_4	11.51020	***	(2.46285)
SUR_5	29.92084	***	(4.93679)
SUR_6	−8.55571	***	(2.18172)
SUR_7	−17.28111	***	(1.97327)
SUR_8	−24.65333	***	(3.92494)
SUR_9	−12.50851	***	(1.73082)
SUR_{10}	52.95169	***	(8.89797)
SUR_{11}	−29.12239	***	(3.74140)
SUR_{12}	27.53494	**	(8.38238)
SUR_{13}	40.65067	***	(2.35361)
SUR_{14}	−30.77390	***	(5.94541)
SUR_{15}	10.20117	**	(3.90063)
SUR_{16}	26.24453	***	(4.94854)
SUR_{17}	15.31839	.	(7.92467)
SUR_{18}	−13.13624		(11.54109)
SUR_{19}	9.57174	***	(2.79510)
SUR_{20}	−0.05661		(5.42079)
SUR_{21}	16.39351	***	(2.46937)
SUR_{22}	−13.18984	***	(3.01087)
SUR_{23}	20.16760	***	(3.38894)
SUR_{24}	2.00894		(2.29237)
2007-10	−0.10755		(2.45419)
2011-15	−0.42089		(3.73646)
2016-20	−0.83554		(5.26932)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. Robust standard errors (White) due to heteroskedasticity.

n = 1,042

$R^2 = 0.42640$

Table 3.12: Estimation Results Model 3.5

Model 3.5 (ABG$_i$, OLS regression)

$\hat{\beta}_j$	$i = 1$		$i = 2$		$i = 3$	
(Intercept)	-7.76335	(18.93502)	24.32025 **	(8.88253)	25.86835 .	(12.64705)
EXP$_i$	0.02345	(0.13111)	0.08930	(0.20691)	0.28662	(0.95055)
EXP$_i^2$	-0.00107	(0.00128)	-0.00104	(0.00218)	-0.00820	(0.03757)
L_ABG	0.55974	(1.93545)	1.95948	(2.88098)	7.85824 .	(4.14956)
ECTOMY	0.26263	(6.48481)	-0.43645	(3.94519)	6.25867	(5.31339)
L_DIA	-4.48505	(6.65686)	1.82680	(4.02498)	-11.40125	(9.16555)
REV	1.02807	(4.73380)	6.63185 *	(3.22586)	1.37280	(10.77465)
AGE	0.96296	(0.67352)	-0.44022	(0.34704)	-0.40255	(0.53331)
AGE^2	-0.00874	(0.00702)	0.00407	(0.00382)	0.00242	(0.00628)
FEMALE	0.69551	(2.03858)	-5.22958 .	(2.80007)	-2.75012	(3.86420)
BI	1.34258	(2.02699)	-1.54352	(3.12353)	-1.43034	(4.46561)

$\hat{\beta}_j$	$i = 4$		$i = 5$		$i = 6$	
(Intercept)	20.04993	(13.00229)	33.27415 *	(14.53006)	37.03904 ***	(6.00738)
EXP$_i$	-0.51075	(0.32954)	-1.41883 .	(0.67326)	0.05771	(0.04401)
EXP$_i^2$	0.00969	(0.00627)	0.04593 .	(0.02220)	-0.00028	(0.00018)
L_ABG	-0.14255	(2.93729)	-2.56142	(3.06526)	-1.35717	(1.24775)
ECTOMY	-1.12828	(3.17228)	3.11210	(4.20614)	0.06611	(1.33361)
L_DIA	-5.23686	(6.17015)	—	—	-6.43248	(4.10863)
REV	14.35895 .	(8.46416)	-9.83731	(6.85527)	1.32250	(1.74906)
AGE	0.18096	(0.47179)	-1.11899 *	(0.46555)	-0.68814 **	(0.22088)
AGE^2	-0.00174	(0.00465)	0.01436 *	(0.00559)	0.00773 **	(0.00235)
FEMALE	1.76663	(2.40349)	6.41635 .	(3.63246)	-3.26014 *	(1.32262)
BI	5.81276 .	(3.33541)	-7.81267 *	(3.62310)	-2.38058 .	(1.25838)

$\hat{\beta}_j$	$i = 7$		$i = 8$		$i = 9$	
(Intercept)	12.37317 .	(7.42529)	17.96184	(21.43783)	22.43992 **	(8.53476)
EXP$_i$	-0.01702	(0.04809)	-0.30529	(0.48923)	-0.09090	(0.08607)
EXP$_i^2$	0.00008	(0.00025)	0.00857	(0.00955)	0.00068	(0.00056)
L_ABG	2.55187 .	(1.34593)	0.62190	(3.62063)	1.27138	(1.86717)
ECTOMY	-0.91411	(1.59844)	-1.49656	(11.30212)	-4.46184	(3.40710)
L_DIA	3.18823	(2.37293)	5.19746	(12.13876)	0.48277	(3.05230)
REV	2.38764	(1.78609)	4.84553	(4.89014)	4.06262	(3.80851)
AGE	-0.17065	(0.28891)	-0.45398	(0.51202)	-0.08610	(0.30420)
AGE^2	0.00233	(0.00343)	0.00498	(0.00565)	0.00092	(0.00330)
FEMALE	1.65751	(1.26906)	-0.18027	(4.31427)	2.02902	(1.88610)
BI	-2.58381 .	(1.37469)	0.35775	(4.80617)	-1.67355	(2.08585)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. For $i = \{6, 7\}$, robust standard errors (White) due to heterosked.

n $\left(\text{model } 3.5 \,|\, i = \{1, 2, ..., 9\}\right)$ = 93, 75, 24, 48, 26, 193, 163, 44, 140

R^2 $\left(\text{model } 3.5 \,|\, i = \{1, 2, ..., 9\}\right)$ = 0.12575, 0.19507, 0.40756, 0.29769, 0.58416, 0.15379, 0.06872, 0.10681, 0.06983

Table 3.13: Estimation Results Model 3.6 and 3.7

$\hat{\beta}_j$	Model 3.6 (ABG_b, OLS regression)		Model 3.7 (ABG_e, OLS regression)	
(Intercept)	27.34925 **	(9.43896)	24.16441 ***	(4.24098)
EXP_i	−0.23017	(0.14905)	0.00673	(0.02911)
EXP_i^2	0.00318 .	(0.00163)	−0.00001	(0.00012)
L_ABG	0.46417	(1.40623)	0.67675	(0.76208)
ECTOMY	1.76013	(2.12896)	−0.94105	(0.94537)
L_DIA	−3.11631	(3.41357)	−0.39171	(1.67295)
REV	7.65071 *	(3.77018)	2.99525 **	(1.05583)
AGE	−0.21965	(0.24896)	−0.39381 **	(0.14188)
AGE^2	0.00213	(0.00254)	0.00441	(0.00161)
FEMALE	2.02615	(1.47626)	−0.76867	(0.77157)
BI	−0.45205	(1.57950)	−1.89390 *	(0.78025)
SUR_2	—	—	−0.46807	(2.61536)
SUR_3	—	—	—	—
SUR_4	5.17484 *	(2.51080)	—	—
SUR_5	−0.37926	(3.30664)	—	—
SUR_6	—	—	0.19088	(2.67188)
SUR_7	—	—	0.02996	(1.53571)
SUR_8	—	—	−2.38965	(2.52164)
SUR_9	−2.75021	(5.65602)	5.30063 *	(2.07097)
2007-10	−0.91599	(4.19149)	−0.99095	(1.65990)
2011-15	−8.11635	(6.12050)	−2.39887	(2.44025)
2016-20	−7.53405	(7.27171)	−4.32450	(2.87454)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. In model 3.7, robust standard errors (White) due to heteroskedasticity.

n (model 3.6) = 186, n (model 3.7) = 620

R^2 (model 3.6) = 0.10595, R^2 (model 3.7) = 0.08103

Surgeon reference category (model 3.6) = surg. 3, surgeon reference category (model 3.7) = surg. 1

Table 3.14: Estimation Results Model 3.8

Model 3.8 (SUCC_i, logit)

$\hat{\beta}_j$	$i = 1$			$i = 2$			$i = 3$		
(Intercept)	3.40631		(4.51553)	−1.80887		(2.32439)	−12.63454		(3,956.18263)
EXP_i	−0.01686		(0.03132)	−0.04808		(0.04158)	0.20688		(0.32499)
EXP_i^2	0.00039		(0.00031)	0.00043		(0.00047)	−0.00688		(0.01220)
L_ABG	−0.23698		(0.46059)	−0.42682		(0.63761)	−1.19895		(1.30514)
ECTOMY	−0.15089		(1.47526)	1.64113		(1.06293)	−3.50220		(2.23170)
L_DIA	0.37647		(1.68711)	−0.02002		(0.84645)	14.80494		(3,956.18082)
REV	0.39080		(1.07351)	−3.01385	**	(1.03391)	−20.58993		(3,956.18250)
AGE	−0.13607		(0.16035)	0.04301		(0.10410)	−0.10472		(0.21902)
AGE^2	0.00111		(0.00166)	−0.00020		(0.00112)	0.00223		(0.00312)
FEMALE	0.01837		(0.48346)	0.63885		(0.65640)	0.40439		(1.22849)
BI	−0.19342		(0.48012)	−0.22890		(0.66669)	−1.81066		(1.50569)
$\hat{\beta}_j$	$i = 4$			$i = 5$			$i = 6$		
(Intercept)	−16.48905		(2,695.54623)	−1,095.45394		(38,768,278.6)	−25.14093	***	(3.33418)
EXP_i	0.08846		(0.11959)	38.56051		(13,219.3057)	−0.02157	.	(0.01188)
EXP_i^2	−0.00163		(0.00237)	−1.06216		(368.64292)	0.00009		(0.00005)
L_ABG	−0.29277		(1.12038)	−11.06634		(14,190.0263)	−0.08462		(0.36785)
ECTOMY	0.08679		(0.97462)	−17.69214		(38,746,639.8)	0.34327		(0.40792)
L_DIA	16.58143		(2,695.54431)	—		—	15.35732	***	(0.54335)
REV	−16.10686		(3,956.18055)	432.04462		(381,081.446)	−0.66830		(0.49205)
AGE	−0.02784		(0.15048)	81.77053		(27,992.1313)	0.39744	**	(0.13340)
AGE^2	0.00008		(0.00151)	−1.19727		(404.03457)	−0.00430	**	(0.00141)
FEMALE	−0.99230		(0.75639)	−388.34051		(954,276.924)	1.21900	**	(0.43498)
BI	−1.65036		(1.42796)	1,077.64201		(1,009,000.32)	0.97881	*	(0.38258)
$\hat{\beta}_j$	$i = 7$			$i = 8$			$i = 9$		
(Intercept)	−1.43991		(1.95873)	−5.10192		(4.45266)	−2.15633		(2.10262)
EXP_i	0.01865		(0.01238)	0.06121		(0.09758)	0.02060		(0.01884)
EXP_i^2	−0.00008		(0.00006)	−0.00129		(0.00188)	−0.00015		(0.00012)
L_ABG	−0.42345		(0.33449)	0.21424		(0.71394)	−0.31573		(0.40136)
ECTOMY	0.60288		(0.50616)	16.82183	***	(1.26641)	−0.05269		(0.77606)
L_DIA	−0.08485		(0.68488)	−15.64864	***	(1.70383)	0.26446		(0.72851)
REV	−0.52651		(0.56582)	−1.20090		(1.10043)	−1.52212		(1.13363)
AGE	−0.00367		(0.07032)	0.13078		(0.13565)	0.05518		(0.07777)
AGE^2	0.00011		(0.00076)	−0.00132		(0.00166)	−0.00062		(0.00082)
FEMALE	−0.32280		(0.33577)	0.82157		(0.88625)	−0.47885		(0.40075)
BI	0.29666		(0.36415)	0.71402		(1.07004)	0.30185		(0.42758)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. For $i = \{2, 6, 7, 8\}$, robust standard errors (White) due to heteroskedasticity.

$\mathrm{n}\left(\text{model } 3.8 \,|\, i = \{1, 2, ..., 9\}\right) = 93,\ 75,\ 24,\ 48,\ 26,\ 193,\ 164,\ 44,\ 140$

$\mathrm{BIC}\left(\text{model } 3.8 \,|\, i = \{1, 2, ..., 9\}\right) = 167.71452,\ 124.68979,\ 57.90069,\ 87.38310,\ 32.58097,\ 251.40298,\ 267.35835,\ 94.55602,\ 217.33306$

McFadden's $\mathrm{R}^2\left(\text{model } 3.8 \,|\, i = \{1, 2, ..., 9\}\right) = 0.08578,\ 0.24577,\ 0.27753,\ 0.13302,\ 1.00000,\ 0.18565,\ 0.03704,\ 0.12704,\ 0.05633$

Table 3.15: Estimation Results Model 3.9 and 3.10

$\hat{\beta}_j$	Model 3.9 (SUCC_b, logit)		Model 3.10 (SUCC_e, logit)	
(Intercept)	−1.60563	(2.40727)	−3.64195 **	(1.20090)
EXP_i	0.02775	(0.03629)	−0.00324	(0.00680)
EXP_i^2	−0.00027	(0.00039)	−0.00002	(0.00003)
L_ABG	−0.23862	(0.34022)	−0.25348	(0.17729)
ECTOMY	−0.84951 .	(0.49751)	0.56574 *	(0.26576)
L_DIA	1.08555	(1.11713)	0.18108	(0.37522)
REV	−1.53558	(1.14709)	−0.96855 **	(0.29458)
AGE	0.03842	(0.06147)	0.07793 *	(0.03857)
AGE^2	−0.00038	(0.00063)	−0.00081 .	(0.00042)
FEMALE	−0.63047 .	(0.35639)	0.21980	(0.18415)
BI	0.27968	(0.37741)	0.50074 **	(0.18686)
SUR_2	—	—	1.32237 *	(0.62155)
SUR_3	—	—	—	—
SUR_4	−0.82981	(0.62285)	—	—
SUR_5	0.78489	(0.78250)	—	—
SUR_6	—	—	0.70581	(0.62524)
SUR_7	—	—	0.08404	(0.36164)
SUR_8	—	—	1.48158 *	(0.62078)
SUR_9	−0.20188	(1.35224)	−0.50239	(0.49471)
2007-10	−0.13059	(1.03875)	0.89829 *	(0.43164)
2011-15	0.02719	(1.47627)	1.23886 *	(0.61380)
2016-20	−0.19407	(1.75788)	2.16137 **	(0.76675)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. In model 3.10, robust standard errors (White) due to heteroskedasticity.

n (model 3.9) = 186, n (model 3.10) = 621

BIC (model 3.9) = 309.29119, BIC (model 3.10) = 888.74336

McFadden's R^2 (model 3.9) = 0.82380, McFadden's R^2 (model 3.10) = 0.38731

Surgeon reference category (model 3.9) = surg. 3, surgeon reference category (model 3.10) = surg. 1

Table 3.16: Estimation Results Model 3.11

$\hat{\beta}_j$	Model 3.11 (ABG, OLS regression)		
(Intercept)	21.96431	***	(3.72179)
EXP	−0.00270		(0.00965)
EXP^2	0.00000		(0.00001)
L_ABG	0.36382		(0.61514)
ECTOMY	−0.44025		(0.81972)
L_DIA	−0.59758		(1.34442)
REV	2.74747	**	(0.94902)
AGE	−0.30990	**	(0.10597)
AGE^2	0.00341	**	(0.00116)
FEMALE	−0.40941		(0.63247)
BI	−1.20004	.	(0.66144)
SUR_2	0.52326		(2.14290)
SUR_3	−0.52123		(1.81128)
SUR_4	4.66992	**	(1.56773)
SUR_5	1.05761		(1.88433)
SUR_6	1.55219		(1.41360)
SUR_7	0.53008		(1.17193)
SUR_8	−1.74338		(2.42979)
SUR_9	3.99703	**	(1.30651)
SUR_{10}	1.50211		(1.93167)
SUR_{11}	−4.47964	.	(2.64598)
SUR_{12}	1.33747		(4.67787)
SUR_{13}	7.58788	***	(1.51922)
SUR_{14}	−1.78821		(3.39090)
SUR_{15}	−0.90160		(2.28004)
SUR_{16}	7.85933	.	(4.40465)
SUR_{17}	7.29399	*	(3.25207)
SUR_{18}	2.06957		(4.15998)
SUR_{19}	1.67308		(2.04530)
SUR_{20}	−0.28491		(2.91138)
SUR_{21}	1.16471		(3.34618)
SUR_{22}	−9.00647	***	(2.10698)
SUR_{23}	4.80363	*	(2.39172)
SUR_{24}	−2.96067		(1.84711)
2007-10	−0.59448		(1.45170)
2011-15	−0.55660		(2.45566)
2016-20	−2.55729		(3.37307)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. Robust standard errors (White) due to heteroskedasticity.

n = 938

$R^2 = 0.07748$

Table 3.17: Estimation Results Model 3.12

$\hat{\beta}_j$	Model 3.12 (SUCC, logit)	
(Intercept)	−1.40841	(0.89808)
EXP	−0.00266	(0.00189)
EXP^2	0.00000	(0.00000)
L_ABG	−0.12914	(0.14232)
ECTOMY	0.21509	(0.21796)
L_DIA	0.14483	(0.30113)
REV	−0.80560 **	(0.25170)
AGE	0.05945 *	(0.02784)
AGE^2	−0.00065 *	(0.00030)
FEMALE	0.08289	(0.14832)
BI	0.33147 *	(0.15344)
SUR$_2$	0.10996	(0.43734)
SUR$_3$	−0.42969	(0.52863)
SUR$_4$	−1.34174 **	(0.43920)
SUR$_5$	−0.40464	(0.49570)
SUR$_6$	−0.51484	(0.32802)
SUR$_7$	−0.37931	(0.27660)
SUR$_8$	0.36310	(0.50082)
SUR$_9$	−0.65138 *	(0.28742)
SUR$_{10}$	−1.15216	(0.79171)
SUR$_{11}$	0.48974	(0.65842)
SUR$_{12}$	−0.11387	(0.82307)
SUR$_{13}$	−14.08002 ***	(1.06510)
SUR$_{14}$	−0.35417	(0.72120)
SUR$_{15}$	−0.04671	(0.56234)
SUR$_{16}$	−1.45152	(0.94236)
SUR$_{17}$	−1.87937 *	(0.75379)
SUR$_{18}$	−0.54580	(1.18947)
SUR$_{19}$	−14.09689 ***	(1.09231)
SUR$_{20}$	−0.21233	(1.12420)
SUR$_{21}$	0.47394	(0.70083)
SUR$_{22}$	14.61955 ***	(1.08435)
SUR$_{23}$	−14.65144 ***	(1.10965)
SUR$_{24}$	0.64199	(0.60419)
2007-10	0.54919	(0.34293)
2011-15	0.46748	(0.55168)
2016-20	0.93559	(0.77143)

Significance levels: *** 0.001 ** 0.01 * 0.05 . 0.1

Standard errors in parentheses. Robust standard errors (White) due to heteroskedasticity.

n = 939

McFadden's $R^2 = 0.06193$

4 Learning Effects in Ischemic Stroke Treatment

Carsten Bauer[171]

4.1 Introduction

Stroke is a major issue in health care. From 2018 to 2020, about 253,000 strokes were treated on average annually in German hospitals (Gesundheitsberichterstattung des Bundes, 2022). Moreover, it is forecasted that the number of strokes will rise by 37 % until 2030 and a staggering 62 % until 2050 in comparison to 2007 (Statista, 2010).[172] In 2019, stroke was the second most frequent cause of death worldwide (World Health Organization, 2020). In 2017, about 7 million disability-adjusted life years were lost due to stroke in the EU (Wafa et al., 2020).[173]

There are two major types of strokes, ischemic (80-85 %) and hemorrhagic ones (15-20 %). This review focuses on research with regard to ischemic strokes. Timely treatment of ischemic strokes is essential as it is associated with a high chance of good outcome.[174] This means there is an inherent need to reduce times from onset to treatment. This span of time can be divided into the time from onset to arrival at the hospital as well as from arrival to treatment. The latter one is referred to as the door-to-needle (DTN) time in case of a thrombolysis treatment.[175] There is much research concerned with the reduction of DTN times in ischemic stroke patients, especially in the recent past.[176]

---

[171] I thank PD Dr. Dr. Stefan Schenk for his valuable impulses, Mareike Seeger for providing medical advice and Hanna Jenzen for her support during the literature search.

[172] In 2007, about 209,000 strokes were treated in German hospitals (Gesundheitsberichterstattung des Bundes, 2022).

[173] Disability-adjusted life-years lost are caused by premature death and years of healthy life lost due to disability. Disability is considered by multiplying the years of healthy life lost with a disability weight (DW). In case of acute ischemic stroke, the DW ranges from 0.019 for mild long-term consequences to 0.588 for severe long-term consequences and cognition problems (Global Burden of Disease Collaborative Network, 2020), i.e. a stroke survivor living with mild long-term consequences loses 0.019 years of healthy life each year.

[174] See section 4.2 for the measurement of stroke outcome.

[175] “Needle” refers to the intravenous access necessary for a thrombolysis.

[176] The articles included in this review represent this, see figure 4.2.

In general, the evaluation of possible learning curve effects is meaningful from a medical as well as a managerial perspective. From the medical point of view, existing learning curves might primarily influence outcomes. It is conceivable that with rising experience in a specific type of treatment (surgical or conservative), outcomes improve along with staff's (surgeons' or physicians') skills. From the managerial perspective, possible learning curves can reduce the treatment time. Besides, high outcome quality is typically linked to lower overall costs of a case as complications are associated with a higher resource use, a prolonged length of stay and higher follow-up costs. It seems likely that with rising experience in a specific type of treatment, routine emerges which improves surgeons' or physicians' movements and consequently, treatment times get shorter.

In case of stroke treatment, learning curve effects might optimize in-hospital processes, accelerate the clarification of symptoms and eventually decrease DTN times as well as improve outcome quality.[177]

The paper at hand reviews research engaged in thrombolysis treatment of ischemic strokes. The focus of this review is on research methods and results with regard to staff's learning effects in stroke treatment which become manifest in a reduction of DTN times. Research on the pre-hospital phase is not subject of this review.[178] The aim is to synthesize relevant literature and concisely represent it. Furthermore, the review discusses the economic implications.

The remainder of this review is organized as follows: Section 4.2 contains background information about stroke, stroke treatment and stroke outcome. The search strategy and the inclusion criteria are described in section 4.3, followed by the results in section 4.4. A discussion of the results and their economic implications is presented in section 4.5. The review closes with a short summary and outlines the need for further research in section 4.6.

4.2 Background

In order to clarify assumed stroke symptoms, patients can be subjected to computed tomography or magnetic resonance tomography (also called magnetic resonance imag-

---

[177] Outcome in stroke cases is highly dependent on timely treatment, see section 4.2.

[178] Articles are for example engaged in the reduction of onset-to-treatment times through improved communication between the emergency medical service and the hospital or raising public awareness of stroke symptoms and their urgency.

ing). For these procedures, patients need to be admitted to hospital, since it is virtually impossible to clarify the symptoms elsewhere.[179]

Ischemic strokes account for 80-85 % and hemorrhagic ones for 15-20 %. Ischemic strokes are treated by a recanalization therapy which can be either a systemic thrombolysis (also called thrombolytic therapy) with a (recombinant) tissue plasminogen activator ((r)tPA) or a catheter-directed thrombolysis (pharmaceutical is applied directly to the affected vessel via a catheter) as well as a mechanical recanalization (thrombectomy) by means of recanalization devices. Neurological guidelines unrestrictedly recommend thrombolysis only up to 4.5 hours after symptom onset. Thrombectomy is recommended up to 6 hours after symptom onset or even beyond in particular cases (Ringleb et al., 2021).[180] Hemorrhagic strokes are attended to by surgery.

Time is the essential factor in treatment of ischemic strokes as it is generally accepted that the earlier the thrombolysis is applied the higher the chance of good outcome as well as full recovery. Saver (2006) finds that the average patient loses 1.9 million neurons and 13.8 billion synapses every minute in which a large vessel ischemic stroke is untreated. The brain loses as many neurons as it does in almost 3.6 years of normal aging in each hour which elapses without treatment. The investigation of Meretoja et al. (2014) reveals the long-term effects of a reduction of the per-minute brain damage quantified by Saver (2006): Reducing the onset-to-treatment time by a single minute grants on average 1.8 days of extra healthy life. Consequently, neurological guidelines recommend thrombolysis to be applied as soon as possible (Ringleb et al., 2021). Due to the 4.5-hour window in which thrombolysis is recommended, cutting delays within the hospital might lead to a higher proportion of patients receiving beneficial thrombolysis: Messé et al. (2016) compare cases thrombolysis was administered to with cases where it was not. They find that the median door-to-CT time[181] of 40 minutes for patients not receiving thrombolysis was twice as long as the median for patients receiving it.

Outcome in stroke cases is usually evaluated by the Modified Rankin Scale (mRS), a

---

[179] With the usual equipment of ambulances, diagnostics is not possible. In recent times, there are trials with mobile stroke units (MSUs, ambulances with diagnostics equipment) in order to be able to diagnose strokes and begin treatment on the road. In Germany, a first trial started in Berlin in 2011. However, MSUs are far from being in use comprehensively, neither in Germany nor worldwide. For more information on the trial in Berlin as well as further references on MSUs, see Alexandrov and Nilanont (2021).

[180] Though, a previous thrombolysis (which guidelines advise only up to 4.5 hours after symptom onset) is recommended.

[181] Time from arrival at the hospital to computed tomography (imaging for diagnostics).

measure of disability in everyday life.[182] Another measure of impairment is the National Institutes of Health Stroke Scale (NIHSS).[183] Besides specific scales, (in-hospital) mortality and occurrence of complications (especially bleedings, i.e. intercranial hemorrhage)[184] are frequently used to measure outcome quality. There are various studies of the relation between a reduction of the DTN time and the improvement of outcomes (e.g. Goyal et al., 2019; Yafasova et al., 2021; Xian et al., 2022).[185]

The everyday life focus of the mRS illustrates that outcome is directly linked to patients' future quality of life. This in turn is closely connected to costs incurred by the health care system. The lower the quality of life, the more treatment is necessary which increases costs. For society, further costs, for example care insurance costs, costs of home care by relatives or costs of productivity losses can be saved by improved outcome quality.

4.3 Methods

4.3.1 Search Strategy

As the evaluation of possible learning curve effects is meaningful from a medical as well as a managerial perspective, the search strategy comprised medical and managerial databases. To be exact, the databases were Business Source Premier (managerial), LIVIVO (managerial and medical), MEDLINE (medical), PubMed (medical), Science Direct (managerial and medical) and Web of Science (managerial and medical). The keywords “door-to-needle”, “door to needle” and “DTN” were applied in combination with “stroke”, “apoplexy” and “optimiz*”.[186] German equivalents of the English keywords were also used. Search results were restricted to articles as well as English and German language where filters allowed language selection. There was no restriction with regard to the year of publication. The search was conducted in late April and early May 2022. Bibliographies of eligible articles from the database search were scanned for further eligible articles.

---

[182] See table 4.1 for the description of the mRS values.

[183] See table 4.2 for the description of the NIHSS values. The NIHSS is also frequently used for pre-treatment stroke severity classification.

[184] Bleeding in the brain; a possible complication of thrombolysis.

[185] Goyal et al. (2019) as well as Xian et al. (2022) investigate the effect of a reduction on the short-term outcome, Yafasova et al. (2021) depict the effect on the long-term outcome. All three articles include a considerable number of patients (n = 601/6,252/185,501).

[186] In Science Direct, truncation (“optimiz*”) was not possible. “Optimization” and “optimizing” were used instead.

4.3.2 Inclusion Criteria

Publications in English and German were considered. The articles were supposed to be primary research, i.e. reviews and meta-analyses were excluded. Title and abstracts of remaining, potentially eligible records were scanned in order to filter inappropriate articles in terms of content. For inclusion, an article had to address learning effects; however, it did not have to address them explicitly. There are numerous articles comparing DTN times for example between two methods of diagnostics, between hub and spoke centers in a telemedicine network, presence of a pharmacist in the stroke team, etc. These articles were excluded from analysis, since they are concerned with the (reduction of) DTN times, but do not address any kind of learning.

In the context of this review, learning in a broader sense means boosting performance of the same persons or the same team by a quality improvement program, for example a revised treatment protocol or a reorganization of logistic processes within the hospital. This means there are no learning curve effects, but there is a change in the setting which triggers learning. Studying learning in a narrower sense means studying learning curve effects. The development of DTN times over time (at least by two points in time) is evaluated. Occasion might be given by a new treatment protocol having been introduced or beginners having entered the emergency department. In case of learning curves effects, learning happens by gaining experience in stroke treatment. The setting does not change. Some studies combine the investigation of learning in a broader sense and learning in a narrower sense. For example, they study the change in DTN times caused by an improvement initiative, this means they compare DTN times before and after the initiative, and additionally, they consider DTN times in the period following the initiative over several points in time.

Studies which are neither concerned with learning in a narrower sense nor in a broader sense, but indirectly engage in learning curve effects by comparing DTN times from low-volume hospitals[187] with high-volume hospitals were also included, as the differences in DTN times are probably the result of learning and economies of scale.

The articles have to deal with the in-hospital phase, learning in the pre-hospital phase was excluded. This means articles on learning for example by improved communication between emergency medical services and hospitals via (revised) notifications were not considered.

---

[187] Volume refers to the ischemic stroke volume.

4.4 Results

4.4.1 Literature Search

The literature search in databases detected 2,578 articles and resulted in 154 eligible ones. Scanning the bibliographies of these articles yielded 11 further, eligible articles, resulting in 165 articles being included in total. Figure 4.1 illustrates the selection process according to the inclusion criteria.

4.4.2 Included Studies

Stroke cases have been increasing and stroke treatment has been gaining more relevance. This is represented in research concerned with (the reduction of) DTN times. Figure 4.2 depicts the number of articles included in this review by year of publication. The articles are from the period 1998-2022. From around 2012 on, there was a strong increase in the number of correspondent articles per year.[188]

Subjects. The majority of articles addresses learning understood in a broader sense by reporting the result of a quality improvement program (n = 126); only a minority is concerned with learning in a narrower sense (n = 62). Of the latter articles, one investigates physicians' individual learning curve effects. A few studies are indirectly engaged in learning curve effects as they compare DTN times from low-volume hospitals with high-volume hospitals (n = 7).[189]

Hospitals. The study setting is often a university or tertiary care hospital (n = 71); only in a small number of articles, secondary care hospitals are considered (n = 4). Primary and comprehensive stroke centers are seldom explicitly mentioned (n = 3, n = 4).[190] The hospitals studied in the included articles are located in 38 countries.[191] Figure 4.3 highlights these countries in a world map.

---

[188] As the search was conducted in late April and early May 2022, only a fraction of 2022 (about one third) is covered. If the number of articles (10, simplified from four months) is extrapolated to the entire year, the number (30) marks the preliminary peak.

[189] Multiple keywords per article are possible. For a compact illustration of these and further numbers regarding subject keywords, see table 4.3.

[190] Numbers refer to articles, not to hospitals (some articles study more than one hospital). Multiple types of hospitals per article are possible. In some articles (especially in large multicenter studies or studies using large databases as for example the Get With The Guidelines-Stroke registry), there is no detailed information about the hospitals. For a compact illustration of these numbers regarding hospitals, see table 4.3.

[191] Some multinational studies do not give detailed information about the location of hospitals, so the number of countries may be higher.

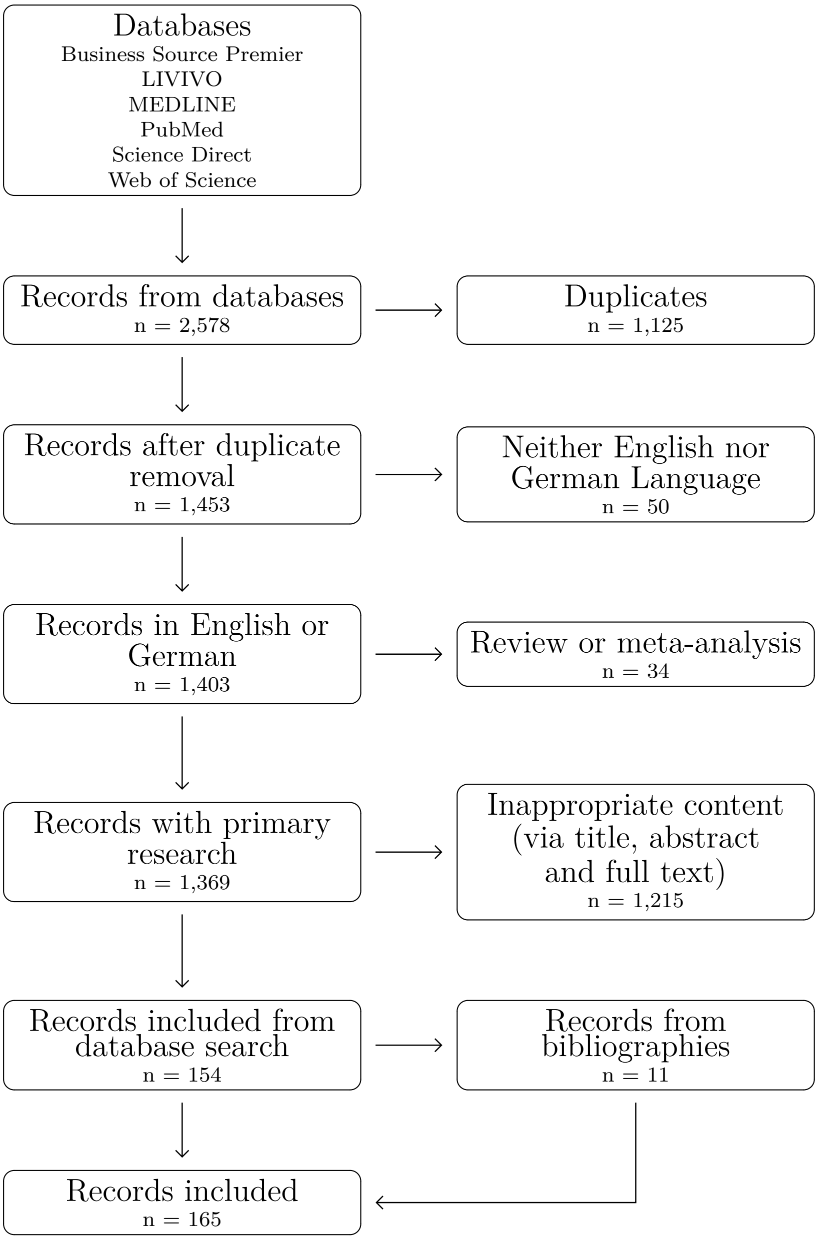

Figure 4.1: Literature Inclusion

Patients. One study has a focus on the elderly, the other studies do not have specific limitations with respect to patients' characteristics. The number of investigated cases strongly varies. There are single-sample studies as well as studies relying on large stroke databases with thousands of cases.

DTN Times. It is striking that virtually all articles report improvements in DTN times. Quality improvement programs change the percentage of patients receiving treat-

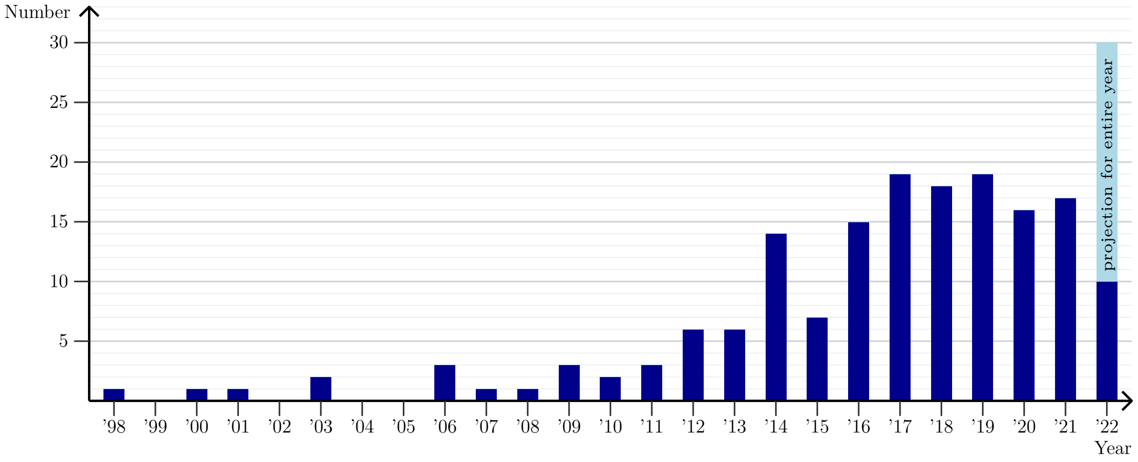

Figure 4.2: Year of Publication of Included Articles

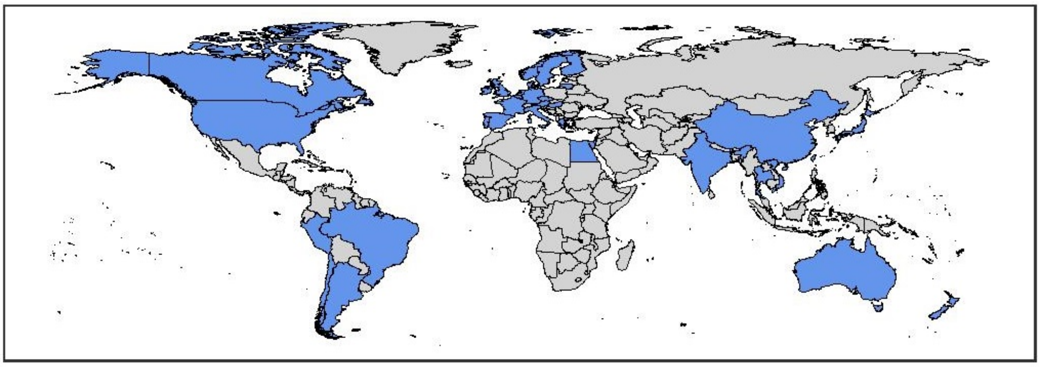

Figure 4.3: Countries of Hospitals from Included Articles

ment within one hour[192] from median 32 to 70 % (mean 38 to 67 %).[193] The median of reported median DTN times falls from 73 to 44 min. (mean of reported median DTN times 69 to 44 min.).[194] The median of reported mean DTN times falls from 76 to 48 min. (mean of reported mean DTN times 76 to 52 min.).[195]

Learning in a narrower sense increases the percentage of patients receiving treatment within one hour.[196] The median percentage reported in the beginning is 31 % (mean 26 %)

---

[192] Calculations use DTN $\leq$ 60 min. as well as $<$ 60 min. depending on the articles.

[193] These and the following numbers are regardless of the variations of time spans between studies and other differences such as the extent of the quality improvement measures.

[194] In two cases, median DTN times increase. However, the rise by 2 and 7 min. respectively is reported to be not significantly different from zero (in the following "significant" for short). In one case, there is no change in median DTN time.

[195] For a compact illustration of these and further DTN times, see table 4.4.

[196] Calculations use DTN $\leq$ 60 min. as well as $<$ 60 min. depending on the articles. There is one exception in which the percentage decreases over the study period. The study investigates the trend of residents'

and 48 % (mean 47 %) in the end. The median of reported median DTN times falls from 75 to 61 min. (mean of reported median DTN times 75 to 61 min.).[197] The median of reported mean DTN times falls from 102 to 71 min. (mean of reported mean DTN times 94 to 67 min.).[198]

Studies indirectly engaging in learning curve effects by comparing low-volume with high-volume hospitals report a negative correlation between volume and DTN times, i.e. the higher the hospital volume, the lower the DTN times. In low-volume hospitals, the median percentage of patients receiving treatment within one hour[199] is 30 % (mean 45 %), in high-volume hospitals, it is 63 % (mean 62 %). The median of reported median DTN times for low- and high-volume hospitals are 73 and 52 min. respectively (mean of reported median DTN times 62 and 51 min. respectively).

Clinical Outcome. As their focus is on DTN times, not all studies investigate clinical outcomes. Almost all articles doing so find improvements or no differences in clinical outcomes, at least no statistically significant ones. There are merely two studies reporting worse outcomes.[200] Improvements are for example quantified by a higher probability of favorable/good outcome in the mRS, lower post-treatment NIHSS, lower (in-hospital) mortality, fewer complications (especially intercranial hemorrhage), and higher rates of discharge to home.

Statistical Methods. Most studies do statistical testing, for example in order to identify significant changes in DTN times. A minority of articles also contains regression models (mostly linear or logistic regression models). Some articles are limited to descriptive statistics.

Journals. The included studies are from a wide range of journals (n = 77). Most articles are from the Journal of Stroke and Cerebrovascular Diseases (n = 19), Stroke (n =

DTN times in their first year starting each July, the so-called “July effect”. However, the change is only –0.10 percentage points. A statement with regard to significance is not possible with the given information.

[197] In two cases, median DTN times increase. In both cases, the “July effect” is studied. One article compares the first quarter of the year with the last one (change by 6 min.) and finds no significant difference. In one case, there is no change in median DTN time.

[198] In one case, mean DTN times increase by a mere 0.5 min., but this rise is not significant.

[199] Calculations use DTN ≤ 60 min. as well as < 60 min. depending on the articles.

[200] One study reports a higher rate of symptomatic intercranial hemorrhage, however simultaneously detects neurological improvement (measured by NIHSS). The second study investigates a possible “July effect”, but does not really find such one as DTN times remain almost unchanged throughout the year.

16) as well as Frontiers in Neurology (n = 8). The median Journal Impact Factor is 2.989 (mean 7.007).[201] Even though some journals also engage in health care management, all are medical journals in the first instance; none of the included articles is from a managerial journal.

Consideration of Costs. The great majority of the articles (from medical journals) does not pay attention to the economic implications of learning effects; only three studies consider costs at all: One article reports the cost savings due to less wasted tPA resulting from an optimized treatment process. Another study uses the estimated lifetime cost savings of beneficial rtPA administration to evaluate the cost effects of an increased share of patients treated with rtPA. The third article calculates the costs of reducing the DTN time by 1 min. per patient in a quality improvement project.

4.5 Discussion

4.5.1 Synthesis

It is important to notice that the reported DTN times must not be overrated; they are rather used to get a rough idea of the time spans and their changes. Due to a large variety in study settings as well as time periods, it is virtually impossible to directly compare the results of the studies. Because of the usually nonlinear learning curve (strongest effects of further “experience” in the beginning, effect of further “experience” decreases over time), it would not be possible to make the results comparable by adjusting to a single time period, even if the setting was comparable. For example, if the percentage of DTN times up to one hour increases by 10 percentage points within one year in study A and by 20 percentage points within two years in study B, it is not an appropriate approach to assume the annual change in the percentage in study B to be 10 percentage points. Most likely, the change is more than 10 percentage points in the first year and less in the second year. Besides, there is no uniform definition of “door” and even “needle” in DTN times (Kruyt et al., 2013). This means one and the same case can have diverging DTN times assigned to in different studies.

Nevertheless, the results show that, apart from a few exceptions, there is learning in stroke treatment, in a broader as well as in a narrower sense, which results in reduced DTN

---

[201] Based on 75 journals. For a compact illustration of these numbers regarding journals, see table 4.5.

times. Furthermore, the results of the articles which also investigate clinical outcomes substantiate the association of shorter DTN times with improved clinical outcomes.

The DTN time-reducing effect of hospital's volume may suggest centralized stroke treatment in order to have higher volumes at the remaining hospitals which treat strokes. However, there is a challenge: Centralized treatment means on average longer ways to hospital, which adversely affects the time from onset to arrival at the hospital and counteracts the reduction of DTN times within the hospital. Thus, a more centralized stroke treatment has to be carefully traded off. Otherwise, the attempt might effect the opposite of the intended purpose and increase onset-to-treatment times.

As already depicted, none of the included articles was detected in a managerial journal; all articles are contained in medical journals. Only three articles consider costs at all. However, these considerations seem to be isolated first approaches to taking costs into account. There are no studies which comprehensively analyze the impact of learning in stroke treatment on costs. For example, the study which calculates the costs of reducing the DTN time in a quality improvement project does not contain a calculation of cost savings associated with the DTN time reduction. However, this would be necessary in order to thoroughly evaluate the DTN time reduction in this setting. Managerial research is obviously not engaged in learning curve effects in ischemic stroke treatment, although stroke is a major issue in health care and an important cost factor with an upward trend.

4.5.2 Economic Implications

The economic relevance of stroke becomes apparent when having a look at stroke-related costs: Total costs roughly accounted for 45 billion € in the EU in 2015.[202] 20 billion € were incurred as direct health care costs, 16 billion € as costs of informal care of people with stroke and 9 billion € as costs due to productivity losses (Wilkins et al., 2017). Lifetime direct costs of ischemic stroke in Germany accounted for 43,129 € in 2004 (Kolominsky-Rabas et al., 2006).[203]

Efforts to cut costs might start with reducing DTN times via an improvement initiative, i.e. learning in a broader sense. However, such projects cause costs themselves.

---

[202] To make amounts more comparable, they are also given adjusted for inflation to 2022 and in Euro for amounts in other currencies. Using EU inflation rates from 2016 to 2021, costs would account for 49 billion € in 2022.

[203] Direct costs comprise inpatient care, outpatient care, rehabilitation, and nursing. Costs are discounted to the year 2004. Using German inflation rates from 2005 to 2021, costs would account for 55,360 € in 2022.

Ajmi et al. (2021) calculate the costs of reducing the DTN time by 1 min. per patient in a quality improvement project in a Norwegian university hospital setting. The costs amount to $ 29 when all costs are included and $ 13 when unpaid time is excluded.[204] These per-minute costs are accompanied by additional 1.8 days of extra healthy life (Meretoja et al., 2014). This in turn saves costs which means there has to be a comparison of costs in order to properly evaluate the initial economic effect of a DTN time reduction project. Though, it has to be considered that improvement initiatives usually have a limited time span and thus cause costs only in this period whereas reduced DTN times probably persist. Learning curve effects, i.e. learning in a narrower sense, can reduce DTN times without causing costs.

Fattore et al. (2012) investigate the costs of Italian stroke patients within the first year.[205] They find that the mRS is strongly associated with costs throughout this period, i.e. the better the mRS, the lower the costs. For example, if the mRS is 4 or 5, costs following discharge are almost seven times as high as if the mRS is 0 to 2 (13,382 vs. 1,965 €).[206] Barral et al. (2021) specifically evaluate the costs of informal care and productivity losses within the first year in France. They find costs of informal care to be more than five times higher in case of mRS 3 to 5 compared to mRS 0 to 2 (25,200 vs. 4,607 €) and the costs of productivity losses to be about 10 % higher (8,015 vs. 7,403 €). Both types of costs combined, costs are almost three times higher in case of mRS 3 to 5 compared to mRS 0 to 2 (33,215 vs. 12,010 €),[207] i.e. their findings are in line with the ones by Fattore et al. (2012).[208] Kim et al. (2020) engage in 5-year costs of stroke in South Korea and attest the positive correlation between the mRS and costs. If the mRS is 5, costs are almost five times as high as if the mRS is 0 (257,486 vs. $ 53,578).[209] Combining research results about the relation between a reduction of DTN times and the improvement of outcomes

---

[204] They use prices and an exchange rate NOK-USD of 2019. Unpaid time is considered to be opportunity costs. It is defined as personnel time spent for the DTN time reduction without causing any direct or extra personnel costs, either during working hours or outside of working hours. Using the mean exchange rate USD-EUR of 2019 and Norwegian inflation rates from 2020 and 2021, costs would account for 27 and 12 € respectively in 2022.

[205] They study ischemic as well as hemorrhagic stroke patients and consider health care costs, non-health care costs and productivity losses.

[206] See table 4.1 for the description of the mRS values. Using Italian inflation rates from 2008 to 2021 (data from 2005 to 2007), costs would account for 16,094 and 2,363 € respectively in 2022.

[207] Using French inflation rates from 2017 to 2021 (data from 2015 to 2016), costs would account for 35,660 and 12,894 € respectively in 2022.

[208] Although Barral et al. (2021) state considerably higher costs (it should be noted that the study scope is even smaller compared to Fattore et al. (2012)), both studies report a multiplication of costs with deteriorating mRS.

[209] Using the mean exchange rate USD-EUR of 2016 and South Korean inflation rates from 2017 to 2021 (data from 2011 to 2016), costs would account for 248,230 and 51,652 € respectively in 2022.

(e.g. Goyal et al., 2019; Yafasova et al., 2021; Xian et al., 2022) with the results that more favorable outcomes are accompanied by lower costs, a reduction of DTN times is supposed to reduce total costs.

The already mentioned centralization of stroke treatment would also have economic effects. Shorter DTN times mean less usage of in-hospital resources, which reduces costs. As shorter DTN times are associated with improved clinical outcomes, costs following discharge are cut. Besides shorter DTN times, economies of scale can reduce costs, for example through a higher utilization of equipment and fixed costs degression. However, onset-to-treatment times may be increased by a centralization strategy, which worsens clinical outcome.[210] As outcome and costs are closely related,[211] such strategy has to be carefully traded off not only from a medical perspective, but also from a managerial one, because a failed centralization strategy also has negative cost consequences resulting from deteriorating clinical outcomes due to longer onset-to-treatment times.

4.6 Conclusions

There are many articles concerned with learning in a broader as well as in a narrower sense regarding DTN times in ischemic stroke treatment. Virtually all studies report a reduction of DTN times, either through a quality improvement program or due to learning curve effects. Several studies depict the DTN time-reducing effect of hospital's volume. The results of this review have made apparent that the reduction of DTN times, i.e. speeding up processes, does not contradict the quality of treatment. In contrast, faster processes are the key to improved clinical outcomes in stroke treatment. Improvements in clinical outcomes in turn come along with cost reductions. Thus, DTN time reductions are beneficial from a medical as well as a managerial perspective.

The economic evaluation of DTN time reductions is challenging. There are studies engaging in the relation between the reduction of DTN times and the improvement of clinical outcomes as well as studies concerned with outcome measures and associated costs; however, there are no articles which comprehensively evaluate the economic implications and importance of learning in stroke treatment. Nevertheless, this review has indicated the enormous costs of stroke and by combining study results, it has been able to argue that DTN reductions are supposed to reduce total costs. The review has also made apparent

---

[210] This aspect is elaborated on in section 4.5.1.

[211] This relation is outlined in section 4.2.

that there are no studies which attend to learning in ischemic stroke treatment from a primarily managerial perspective. The great majority of the solely medical articles does not pay attention to the economic implications of learning effects. This means there is need for further managerial research.

Improvements in stroke treatment have the chance to counteract the expected rise in stroke burden as well as to make a contribution to the avoidance of an overstraining of health care systems, especially against the background of the demographic change. Therefore, there is need for further managerial research in order to meet the importance of an issue in health care that affects society as a whole.

Appendix 4

Table 4.1: Modified Rankin Scale

Value	Description
0*	No symptoms
1*	No significant disability. Able to carry out all usual activities, despite some symptoms
2*	Slight disability. Able to look after own affairs without assistance, but unable to carry out all previous activities
3	Moderate disability. Requires some help, but able to walk unassisted
4	Moderately severe disability. Unable to attend to own bodily needs without assistance, and unable to walk unassisted
5	Severe disability. Requires constant nursing care and attention, bedridden, incontinent
6	Dead

* Referred to as favorable, good or independent functional outcome.

Table 4.2: National Institutes of Health Stroke Scale

Value	Description
0	No stroke symptoms
1-4	Minor stroke
5-15	Moderate stroke
16-20	Moderate to severe stroke
21-42	Severe stroke

Scale consists of 13 items, answer options correspond to values 0-2, 0-3 or 0-4.

Table 4.3: Overview of Articles

Articles	n
	165
Subject keyword	**n[1]**
Learning broader sense	126
Learning narrower sense	62
Low vs. high volume	7
Training	13
Telemedicine	10
Cost consideration	3
Hospital	**n[2]**
University or tertiary care	71
Secondary care	4
Primary Stroke Center	3
Comprehensive Stroke Center	4

[1] Multiple keywords per article are possible.

[2] Numbers refer to articles, not to hospitals (some articles study more than one hospital). Multiple types of hospitals per article are possible.

Table 4.4: DTN Times

Quality Improvement Programs (Learning in a Broader Sense)		
DTN Time within one hour[1] (%)	Beginning	End
Median	32	70
Mean	38	67
Reported Median DTN Time (min.)	Beginning	End
Median	73	44
Mean	69	44
Reported Mean DTN Time (min.)	Beginning	End
Median	76	48
Mean	76	52
Learning Curve Effects (Learning in a Narrower Sense)		
DTN Time within one hour[1] (%)	Beginning	End
Median	31	48
Mean	26	47
Reported Median DTN Time (min.)	Beginning	End
Median	74	61
Mean	75	61
Reported Mean DTN Time (min.)	Beginning	End
Median	102	71
Mean	94	67
Volume Effects		
DTN Time within one hour[1] (%)	Low Volume	High Volume
Median	30	63
Mean	45	62
Reported Median DTN Time (min.)	Low Volume	High Volume
Median	73	52
Mean	66	51

[1] Calculations use DTN ≤ 60 min. as well as < 60 min. depending on the articles.

Table 4.5: Journal Metrics

Number of journals	77
Median Journal Impact Factor¹⁾	2.989
Mean Journal Impact Factor¹⁾	7.007
Weighted mean Journal Impact Factor²⁾	6.551

1) Based on 75 journals.

2) Based on 75 journals. Weighted by the number of articles.

Postface

The articles in this collection have made the first step toward utilization of learning curves as a viable option for cost reductions without deteriorating quality by measuring the learning effects. Such sophisticated cost containment strategies can be a blessing in times of steadily increasing health care costs and economic pressure as outlined in the preface. Nevertheless, the implementation up to the actual increase in efficiency requires further efforts in every case as the subject does not admit of easy, standardized solutions. Though, these efforts finally pay off: Such sophisticated strategies can reduce costs and keep quality stable or even improve it. This means, medical and cost containment objectives do not diverge, i.e. there is no ethical conflict.

The collection at hand has identified learning as a complex issue and evaluating learning as a complex set of problems. However, this complexity only rudimentarily lights up in the literature. There, a one-dimensional perspective predominates instead of an integral one. In case of learning in stroke treatment regarding a reduction of the time from the patient's arrival at the hospital until thrombolysis treatment, the review has vividly shown the focus to be solely a medical one.

Though, multiperspectivity is the precondition to cope with the complexity of the subject. Moreover, multiple perspectives can eventually benefit the patient's outcome. The managerial objectives do not need to be in opposition to the medical ones, but can be regarded as complementary. A classic managerial optimization does not serve its own purpose, i.e. optimizing for the sake of optimizing. In a medical context, it serves the medical progress and ultimately the patient, for example suffering from a stroke. Management makes its contribution and deploys synergy effects. The patient benefiting from multiperspectivity, health care systems and society also profit, as improved outcomes are associated with lower efforts and lower costs. This fact underlines the importance of the subject also in economic considerations.

Furthermore, in other, comparable treatments, for example in case of the other major infarction, myocardial infarction, or even in medical conditions which are not comparable, multiperspective approaches can be reasonable, too. And this does apparently not hold only for studies on learning; other research subjects can also profit by a multidisciplinary approach.

The collection has also hinted at how substantial an interdisciplinary cooperation in medical research is. To be on the cutting edge of medical research and to stay there respectively, it is indispensable to meet the demands for interdisciplinarity. It is worthwhile analyzing as well as reflecting with a medical as well as a managerial eye. One must not be underexposed or even blind on one eye—here preferably the managerial one—in order to fully comprehend medical issues in their complexity as for example learning in stroke treatment in case of the review.

The articles in this collection are not able to present final answers in their respective research topics; in every case, there is need for further, differentiated research. Admittedly, it would be somewhat presumptuous to declare the results of the articles to be ultimate. The German philosopher and sociologist Simmel (2011) appropriately describes the dilemma research has to face: “Somewhere knowledge may have an absolute basis, but we can never state irrevocably where this basis is; consequently, in order to avoid dogmatic thought, we have to treat each position at which we arrive as if it were the penultimate one” (p. 110). This statement must not be perceived as a gloomy view; rather, it “is a positive challenge, which the history of thought has illustrated many times” (p. 110).[212]

The remarks regarding the nature of science by the sociologist and economist Weber are consistent with the positive perception of science as an infinite process by his contemporary Simmel. Weber (2011) states in his innovative lecture “Science as a Vocation”[213] (p. 15) that everybody in science knows that his work will be outdated in 10, 20, 50 years. It is the fate of science, it is positively the meaning of scientific activity. Every scientific “fulfillment” means new “questions” and is to be outperformed and should become obsolete. Everybody who wants to serve science has to accept this. To be scientifically outperformed is not just scientists’ fate, it is their purpose. Scientists cannot work without hoping others to progress beyond.

Accepting Weber’s (2011) message, this collection has floated current insights in order to provide new impulses, promote scientific discourse and eventually become obsolete itself.

---

[212] This statement does not only apply to research, i.e. theory, it also does to the activities of hospitals, i.e. practice. Hospitals have to be regarded—like classical enterprises—as learning organizations. From a managerial, analytic perspective, the activities of hospitals can be classified into processes. For example, there is a treatment protocol for patients with assumed stroke symptoms which all respective patients go through. This protocol can be divided into several subprocesses itself, for example admission, diagnostics, etc. These processes in general are subject to suboptimality. To put it short: There is a constant need for process optimization.

[213] Original German publication “Wissenschaft als Beruf”.

Bibliography

Ajmi, S. C., Kurz, M. W., Ersdal, H., Lindner, T., Goyal, M., Issenberg, S. B. and Vossius, C. (2021). Cost-Effectiveness of a Quality Improvement Project, Including Simulation-Based Training, on Reducing Door-to-Needle Times in Stroke Thrombolysis. *BMJ Quality & Safety*.

Alexandrov, A. V. and Nilanont, Y. (2021). Improving Outcomes After Stroke: From Stroke Units to Mobile Stroke Units. *Stroke*, 52(9): 3072-3074.

American Society of Anesthesiologists (2019). ASA Physical Status Classification System. `https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system`.

Amin, M. R. and Friedmann, D. R. (2013). Simulation-Based Training in Advanced Airway Skills in an Otolaryngology Residency Program. *Laryngoscope*, 123(3): 629-634.

Arvidsson, S., Ouchterlony, J., Sjöoustedt, L. and Sväaurdsudd, K. (1996). Predicting Postoperative Adverse Events. Clinical Efficiency of Four General Classification Systems The Project Perioperative Risk. *Acta Anaesthesiologica Scandinavica*, 40(7): 783-791.

Baldwin, D. D., Dunbar, J. A., Parekh, D. J., Wells, N., Shuford, M. D., Cookson, M. S., Smith, J. A., Herrell, S. D., Chang, S. S. and Mcdougall, E. M. (2003). Single-Center Comparison of Purely Laparoscopic, Hand-Assisted Laparoscopic, and Open Radical Nephrectomy in Patients at High Anesthetic Risk. *Journal of Endourology*, 17(3): 161-167.

Barral, M., Rabier, H., Termoz, A., Serrier, H., Colin, C., Haesebaert, J., Derex, L., Nighoghossian, N., Schott, A.-M. and Viprey, M. (2021). Patients' Productivity Losses and Informal Care Costs Related to Ischemic Stroke: A French Population-Based Study. *European Journal of Neurology*, 28(2): 548-557.

Bauer, C., Möbs, N., Unger, O., Szczesny, A. and Ernst, C. (2020). Spillover Effects of Specialization Strategies in Hospitals: An Analysis of the Effects in the Short, Medium, and Long Term. Working paper. `https://www.econstor.eu/handle/10419/215783`.

Beldi, G., Bisch-Knaden, S., Banz, V., Mühlemann, K. and Candinas, D. (2009). Impact of Intraoperative Behavior on Surgical Site Infections. *The American Journal of Surgery*, 198(2): 157-162.

Bellini, V., Guzzon, M., Bigliardi, B., Mordonini, M., Filippelli, S. and Bignami, E. (2020). Artificial Intelligence: A New Tool in Operating Room Management. Role of Machine Learning Models in Operating Room Optimization. *Journal of Medical Systems*, 44: 20.

Berry, M., Berry-Stölzle and Schleppers, A. (2008). Operating Room Management and Operating Room Productivity: The Case of Germany. *Health Care Management Science*, 11(3): 228-239.

Blijleven, E. E., Wegner, I., Tange, R. A. and Thomeer, H. G. (2019). Revision Stapes Surgery in a Tertiary Referral Center: Surgical and Audiometric Outcomes. *Annals of Otology, Rhinology & Laryngology*, 128(11): 997-1005.

Blum, K., Heber, R., Löffert, S., Offermanns, M. and Steffen, P. (2020). Krankenhaus Barometer: Umfrage 2020. Deutsches Krankenhausinstitut.

Blum, K., Löffert, S., Offermanns, M. and Steffen, P. (2019). Krankenhaus Barometer: Umfrage 2019. Deutsches Krankenhausinstitut.

Childers, C. P. and Maggard-Gibbons, M. (2018). Understanding Costs of Care in the Operating Room. *JAMA Surgery*, 153(4): e176233.

Chung, W.-C., Lee, K.-T., Fan, P.-L., Tzeng, D.-S. and Chiu, H.-C. (2006). Operating Room Costs of Laparoscopic Cholecystectomy: Does Surgeon Volume Matter? *Kaohsiung Journal of Medical Sciences*, 22(3): 126-134.

Clark, J. R. and Huckman, R. S. (2012). Broadening Focus: Spillovers, Complementarities, and Specialization in the Hospital Industry. *Management Science*, 58(4): 708-722.

Collins, T. C., Daley, J., Henderson, W. H. and Khuri, S. F. (1999). Risk Factors for Prolonged Length of Stay After Major Elective Surgery. *Annals of Surgery*, 230(2): 251.

Com-Ruelle, L., Or, Z. and Renaud, T. (2008). The Volume-Outcome Relationship in Hospitals: Lessons from the Literature. *Issues in Health Economics*, no vol.(135): 1-4.

Contreras, J. M., Kim, B. and Tristao, I. M. (2011). Does Doctors' Experience Matter in LASIK Surgeries? *Health Economics*, 20(6): 699-722.

Deily, M. E. and McKay, N. L. (2006). Cost Inefficiency and Mortality Rates in Florida Hospitals. *Health Economics*, 15(4): 419-431.

Dimick, J. B., Chen, S. L., Taheri, P. A., Henderson, W. G., Khuri, S. F. and Campbell, D. A. (2004). Hospital Costs Associated with Surgical Complications: A Report from the Private-Sector National Surgical Quality Improvement Program. *Journal of the American College of Surgeons*, 199(4): 531-537.

Dimick, J. B., Weeks, W. B., Karia, R. J., Das, S. and Campbell, D. A. (2006). Who Pays for Poor Surgical Quality? Building a Business Case for Quality Improvement. *Journal of the American College of Surgeons*, 202(6): 933-937.

Dinh, J. V., Traylor, A. M., Kilcullen, M. P., Perez, J. A., Schweissing, E. J., Venkatesh, A. and Salas, E. (2006). Cross-Disciplinary Care: A Systematic Review on Teamwork Processes in Health Care. *Journal of the American College of Surgeons*, 202(6): 933-937.

Dudley, R. A., Johansen, K. L., Brand, R., Rennie, D. J. and Milstein, A. (2000). Selective Referral to High-Volume Hospitals: Estimating Potentially Avoidable Deaths. *Journal Of The American Medical Association*, 283(9): 1159-1166.

Durko, M., Kaczmarczyk, D. and Durko, T. (2007). Revision Stapes Surgery – Retrospective Analysis of Surgical Findings in a Series of 21 Otosclerosis Patients. In: Arnold, W. and Häusler, R. (Eds.): *Otosclerosis and Stapes Surgery*, 273-277. Karger, Basel. Advances in Oto-Rhino-Laryngology 65.

Ellis, R. P. (1998). Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins. *Journal of Health Economics*, 17(5): 537-555.

Ernst, C. and Szczesny, A. (2005). Cost Accounting Implications of Surgical Learning in the DRG Era – Data Evidence from a German Hospital. *Schmalenbach Business Review*, 57: 127-166.

Ernst, C. and Szczesny, A. (2006). Spezialisierungs- und Selektionsanreize fester Krankenhausbudgets 1993-2002: Eine empirische Analyse. *Betriebswirtschaftliche Forschung und Praxis*, 58: 566-584.

Ernst, C. and Szczesny, A. (2008). Capped Hospital Budgets, Risk-Influencing Activities and Financial Consequences. *Journal of Accounting and Public Policy*, 27(1): 38-61.

Fattore, G., Torbica, A., Susi, A., Giovanni, A., Benelli, G., Gozzo, M. and Toso, V. (2012). The Social and Economic Burden of Stroke Survivors in Italy: A Prospective, Incidence-Based, Multi-Centre Cost of Illness Study. *BMC Neurology*, 12(1).

Fisher, M. L. and Ittner, C. D. (1999). The Impact of Product Variety on Automobile Assembly Operations: Empirical Evidence and Simulation Analysis. *Management Science*, 45(6): 771-786.

Fleming, S. T. (1991). The Relationship Between Quality and Cost: Pure and Simple? *Inquiry*, 28(1): 29-38.

Flood, A. B., Shortell, S. M. and Scott, W. R. (1994). Organizational Performance: Managing for Efficiency and Effectiveness. In: Shortell, S. M. and Kalunzny, A. D. (Eds.): *Health Care Management: Organization Design and Behaviour*, 316-354. Delmar Publisher.

Gesundheitsberichterstattung des Bundes (2020a). Ausgaben, Kosten, Finanzierung: Betriebswirtschaftliche Eckdaten von Krankenhäusern. `http://www.gbe-bund.de/gbe10/trecherche.prc_them_rech?tk=19200&tk2=19500&p_uid=gast&p_aid=46905248&p_sprache=D&cnt_ut=8&ut=19585`.

Gesundheitsberichterstattung des Bundes (2020b). Ausgaben, Kosten, Finanzierung: Gesundheitsausgabenrechnung. `http://www.gbe-bund.de/gbe10/trecherche.prc_them_rech?tk=19200&tk2=19300&p_uid=gast&p_aid=46905248&p_sprache=D&cnt_ut=10&ut=19310`.

Gesundheitsberichterstattung des Bundes (2020c). Gesundheitsversorgung: Krankenhäuser. `http://www.gbe-bund.de/gbe10/trecherche.prc_them_rech?tk=14501&tk2=16301&p_uid=gast&p_aid=88582463&p_sprache=D&cnt_ut=1&ut=16301`.

Gesundheitsberichterstattung des Bundes (2020d). Gesundheitsversorgung: Personal im Gesundheitswesen, Gesundheitspersonalrechnung. `http://www.gbe-bund.de/gbe10/trecherche.prc_them_rech?tk=14501&tk2=15352&p_uid=gast&p_aid=8921668&p_sprache=D&cnt_ut=17&ut=16642`.

Gesundheitsberichterstattung des Bundes (2022). Krankheiten/ Gesundheitsprobleme: Krankenhäuser allgemein. `https://www.gbe-bund.de/gbe/trecherche.prc_them_rech?tk=8500&tk2=14300&p_uid=gast&p_aid=83515371&p_sprache=D&cnt_ut=1&ut=14300`.

GKV-Spitzenverband (2021). Bundesbasisfallwert. Mar. 31, 2021. `https://www.gkv-spitzenverband.de/krankenversicherung/krankenhaeuser/budgetverhandlungen/bundesbasisfallwert/bundesbasisfallwert.jsp`.

Glasscock, M. E., Storper, I. S., Hayes, D. S. and Bohrer, P. S. (1995). Twenty-Five Years of Experience with Stapedectomy. *Laryngoscope*, 105(9): 899-904.

Global Burden of Disease Collaborative Network (2020). *Global Burden of Disease Study 2019 (GBD 2019) Disability Weights.* Institute for Health Metrics and Evaluation, Seattle.

Goyal, M., Almekhlafi, M., Dippel, D. W., Campbell, B. C. V., Muir, K., Demchuk, A. M., Bracard, S., Davalos, A., Guillemin, F., Jovin, T. G., Menon, B. K., Mitchell, P. J., Brown, S., White, P., Majoie, C. B. L. M., Saver, J. L. and Hill, M. D. (2019). Rapid Alteplase Administration Improves Functional Outcomes in Patients With Stroke due to Large Vessel Occlusions: Meta-Analysis of the Noninterventional Arm From the HERMES Collaboration. *Stroke*, 50(3): 645-651.

Hall, J. C. and Hall, J. L. (1996). ASA Status and Age Predict Adverse Events after Abdominal Surgery. *Journal of Quality in Clinical Practice*, 16(2): 103-108.

Hautmann, R. E., De Petriconi, R. C. and Volkmer, B. G. (2010). Lessons Learned From 1,000 Neobladders: The 90-Day Complication Rate. *The Journal of Urology*, 184(3): 990-994.

Hoonhout, L. H. F., De Bruijne, M. C., Wagner, C., Zegers, H. W. M., Waaijman, R., Spreeuwenberg, P., Asscheman, H., Van Der Wal, G. and Van Tulder, M. W. (2009). Direct Medical Costs of Adverse Events in Dutch Hospitals. *BMC Health Services Research*, 9(1): 27.

Huckman, R. S. and Zinner, D. E. (2008). Does Focus Improve Operational Performance? Lessons from the Management of Clinical Trials. *Strategic Management Journal*, 29(2): 173-193.

Huesch, M. D. and Sakakibara, M. (2009). Forgetting the Learning Curve for a Moment: How Much Performance is Unrelated to Own Experience? *Health Economics*, 18(7): 855-862.

Hughes, G. B. (1991). The Learning Curve in Stapes Surgery. *Larnygoscope*, 101(12): 1280-1284.

HYER, N. L., WEMMERLÖV, U. and MORRIS, J. A. (2009). Performance Analysis of a Focused Hospital Unit: The Case of an Integrated Trauma Center. *Journal of Operations Management*, 27(3): 203-219.

IANNELLA, G. and MAGLIULO, G. (2016). Endoscopic Versus Microscopic Approach in Stapes Surgery: Are Operative Times and Learning Curve Important for Making the Choice? *Otology & Neurotology*, 37(9): 1350-1357.

INSTITUT FÜR DAS ENTGELTSYSTEM IM KRANKENHAUS (2014a). Fallpauschalen-Katalog 2014. Jan. 20, 2014. `https://www.g-drg.de/Archiv/DRG_Systemjahr_2014_Datenjahr_2012`.

INSTITUT FÜR DAS ENTGELTSYSTEM IM KRANKENHAUS (2014b). Fallpauschalen-Katalog 2015. Sep. 23, 2014. `https://www.g-drg.de/Archiv/DRG_Systemjahr_2015_Datenjahr_2013`.

INSTITUT FÜR DAS ENTGELTSYSTEM IM KRANKENHAUS (2015). Fallpauschalen-Katalog 2016. Oct. 29, 2015. `https://www.g-drg.de/Archiv/DRG_Systemjahr_2016_Datenjahr_2014`.

INSTITUT FÜR DAS ENTGELTSYSTEM IM KRANKENHAUS (2020). Fallpauschalen-Katalog 2021. Nov. 12, 2020. `https://www.g-drg.de/aG-DRG-System_2021/Fallpauschalen-Katalog/Fallpauschalen-Katalog_2021`.

INSTITUT FÜR DAS ENTGELTSYSTEM IM KRANKENHAUS (2021). aG-DRG-Report-Browser 2021. `https://www.g-drg.de/Datenbrowser_und_Begleitforschung/G-DRG-Report-Browser/aG-DRG-Report-Browser_2021`.

IURATO, S., BUX, G., MEVOLI, S. and ONORI, M. (2007). Stapes Surgery in the Elderly. In: ARNOLD, W. and HÄUSLER, R. (Eds.): *Otosclerosis and Stapes Surgery*, 231-236. Karger, Basel. Advances in Oto-Rhino-Laryngology 65.

JHA, A. K., ORAV, E. J., DOBSON, A., BOOK, R. A. and EPSTEIN, A. M. (2009). Measuring Efficiency: the Association of Hospital Costs and Quality of Care. *Health Affairs (Project Hope)*, 28(3): 897-906.

KALISH, R. L., DALEY, J., DUNCAN, C. C., DAVIS, R. B., COFFMAN, G. A. and IEZZONI, L. I. (1995). Costs of Potential Complications of Care for Major Surgery Patients. *American Journal of Medical Quality*, 10(1): 48-54.

KEIM, R., MLYNSKI, R., CEBULLA, M., RADELOFF, A. and HARNISCH, W. (2014). Integration einer neuen HNO-Datenbank am Universitätsklinikum Würzburg – ein Erfahrungsbericht. *Zeitschrift für Audiologie*, 53(3): 114-115.

KHAN, N. A., QUAN, H., BUGAR, J. M., LEMAIRE, J. B., BRANT, R. and GHALI, W. A. (2006). Association of Postoperative Complications with Hospital Costs and Length of Stay in a Tertiary Care Center. *Journal of General Internal Medicine*, 21(2): 177-180.

KHORSANDI A., M. T., JALALI, M. M. and SHOSHI D., V. (2018). Predictive Factors in 995 Stapes Surgeries for Primary Otosclerosis. *Laryngoscope*, 128(10): 2403-2407.

KIM, S.-E., LEE, H., KIM, J. Y., LEE, K.-J., KANG, J., KIM, B. J., HAN, M.-K., CHOI, K.-H., KIM, J.-T., SHIN, D.-I., YEO, M.-J., CHA, J.-K., KIM, D.-H., NAH, H.-W., KIM, D.-E., RYU, W.-S., PARK, J.-M., KANG, K., KIM, J. G., LEE, S. J., OH, M.-S., YU, K.-H., LEE, B.-C., PARK, H.-K., HONG, K.-S., CHO, Y.-J., CHOI, J. C., SOHN, S. I., HONG, J.-H., PARK, M.-S., PARK, T. H., PARK, S.-S., LEE, K. B., KWON, J.-H., KIM, W.-J., LEE, J., LEE, J. S., LEE, J., MERETOJA, A., GORELICK, P. B. and BAE, H.-J. (2020). Three-Month Modified Rankin Scale as a Determinant of 5-Year Cumulative Costs after Ischemic Stroke: An Analysis of 11,136 Patients in Korea. *Neurology*, 94(4): e978-e991.

KIM, Y.-J. L. and SOEKEN, K. L. (2005). A Meta-Analysis of the Effect of Hospital-Based Case Management on Hospital Length-of-Stay and Readmission. *Nursing Research*, 54(4): 255-264.

KLASK, J. and SCHMELZER, A. (2003). Postoperative Komplikationen der Stapesplastik: Eine Analyse unter medizinisch-ökonomischen Aspekten. *HNO*, 51(11): 893-897.

KOLOMINSKY-RABAS, P. L., HEUSCHMANN, P. U., MARSCHALL, D., EMMERT, M., BALTZER, N., NEUNDORFER, B., SCHOFFSKI, O. and KROBOT, K. (2006). Lifetime Cost of Ischemic Stroke in Germany: Results and National Projections from a Population-Based Stroke Registry: The Erlangen Stroke Project. *Stroke*, 37(5): 1179-1183.

KRUYT, N. D., NEDERKOORN, P. J., DENNIS, M., LEYS, D., RINGLEB, P. A., RUDD, A. G., VERMEULEN, M., STAM, J., HACKE, W. and ROOS, Y. B. (2013). Door-to-Needle Time and the Proportion of Patients Receiving Intravenous Thrombolysis in Acute Ischemic Stroke: Uniform Interpretation and Reporting. *Stroke*, 44(11): 3249-3253.

Lovato, A., Kraak, J., Hensen, E. F., Smit, C. F., Giacomelli, L., de Filippis, C. and Merkus, P. (2019). A Critical Look into Stapedotomy Learning Curve: Influence of Patient Characteristics and Different Criteria Defining Success. *Ear, Nose and Throat Journal*, 1-7. Sept. 26, 2019.

Lucidi, D., Molinari, G., Reale, M., Alicandri-Ciufelli, M. and Presutti, L. (2020). Functional Results and Learning Curve of Endoscopic Stapes Surgery: A 10-Year Experience. *Laryngoscope*, 1-7. Aug. 15, 2020.

Marchese, M. R., Conti, G., Cianfrone, F., Scorpecci, A., Fetoni, A. R. and Paludetti, G. (2009). Predictive Role of Audiological and Clinical Features for Functional Results after Stapedotomy. *Audiology and Neurotology*, 14(5): 279-285.

Meretoja, A., Keshtkaran, M., Saver, J. L., Tatlisumak, T., Parsons, M. W., Kaste, M., Davis, S. M., Donnan, G. A. and Churilov, L. (2014). Stroke Thrombolysis: Save a Minute, Save a Day. *Stroke*, 45(4): 1053-1058.

Messé, S. R., Khatri, P., Reeves, M. J., Smith, E. E., Saver, J. L., Bhatt, D. L., Grau-Sepulveda, M. V., Cox, M., Peterson, E. D., Gregg C. Fonarow, G. C. and Schwamm, L. H. (2016). Why are Acute Ischemic Stroke Patients Not Receiving IV tPA?: Results from a National Registry. *Neurology*, 87(15): 1565-1574.

Naderi, B., Roshanaei, V., Begen, M. A., Aleman, D. M. and Urbach, D. R. (2021). Increased Surgical Capacity without Additional Resources: Generalized Operating Room Planning and Scheduling. *Production and Operations Management*.

Nagayo, Y., Saito, T. and Oyama, H. (2021). A Novel Suture Training System for Open Surgery Replicating Procedures Performed by Experts Using Augmented Reality. *Journal of Medical Systems*, 45(60).

Nash, R., Patel, B. and Lavy, J. (2021). Changes to Hearing Levels Over the First Year After Stapes Surgery: An Analysis of 139 Patients. *Otology & Neurotology*, 39(7): 829-833.

OECD (2020). Health Expenditure and Financing: Health Expenditure Indicators. `https://www.oecd-ilibrary.org/content/data/data-00349-en`.

Papaspyrou, K., von Creytz, H., Kolonko, K., Mewes, T., Mann, W. and Scharrer, I. (2012). Nachblutungen und Gerinnungsstörungen bei Tonsillektomien. *HNO*, 60(6): 511-519.

Pernerstorfer, T. and Huemer, G. (2008). Effizientes OP-Management: Die richtigen Dinge tun und die Dinge richtig tun. *Pro Care*, 13: 8-14.

Poe, D. S. (2000). Laser-Assisted Endoscopic Stapedectomy. *Laryngoscope*, 110(5 Suppl. 95): 1-30.

Rahimi, I. and Gandomi, A. H. (2021). A Comprehensive Review and Analysis of Operating Room and Surgery Scheduling. *Archives of Computational Methods in Engineering*, 28(3): 1667-1688.

Ramsay, C. R., Grant, A. M., Wallace, S. A., Garthwaite, P. H., Monk, A. F. and Russell, I. T. (2000). Assessment of the Learning Curve in Health Technologies. A Systematic Review. *International Journal Of Technology Assessment In Health Care*, 16(4): 1095-1108.

Ringleb, P. A., Hametner, C., Köhrmann, M., Frank, B., Jansen, O., Berlis, A., Fischer, U., Laufs, U., Mordasini, P., Schellinger, P. D., Schubert, G. A., Sonnberger, M., Sykora, M., Urbach, H., Wachter, R., Wagner, M., Weiller, C., Harloff, A. and Langguth, P. (2021). Akuttherapie des ischämischen Schlaganfalls, S2e-Leitlinie, 2021. In: Deutsche Gesellschaft für Neurologie (Ed.): *Leitlinien für Diagnostik und Therapie in der Neurologie.*

Roberts, R. R., Frutos, P. W., Ciavarella, G. G., Gussow, L. M., Mensah, E. K., Kampe, L. M., Straus, H. E., Joseph, G. and Rydman, R. J. (1999). Distribution of Variable vs Fixed Costs of Hospital Care. *JAMA*, 281(7): 644-649.

Rosowski, J. J. and Merchant, S. N. (1995). Mechanical and Acoustic Analysis of Middle Ear Recontrsuction. *American Journal of Otology*, 16(4): 486-497.

Sargent, E. W. (2002). The Learning Curve Revisited: Stapedotomy. *Otolaryngology-Head and Neck Surgery*, 126(1): 20-25.

Saver, J. L. (2006). Time Is Brain—Quantified. *Stroke*, 37(1): 263-266.

Schön, F. and Müller, J. (2002). Der Würzburger Ohrbogen: Ein System zur Dokumentation von Operationen und Nachsorgeuntersuchungen in der HNO-Heilkunde. *Laryngorhinootologie*, 81(3): 171-177.

Sergi, B. and Paludetti, G. (2016). Can the Learning Curve in Stapes Surgery Predict Future Functional Outcome? *Acta Otorhinolaryngologica Italica*, 36(2): 135-138.

Sim, J. H., Chatzimichalis, M., Röösli, C., Laske, R. D. and Huber, A. M. (2012). Objective Assessment of Stapedotomy Surgery from Round Window Motion Measurement. *Ear & Hearing*, 33(5): e24-e31.

Simmel, G. (2011). *The Philosophy of Money* (T. Bottomore, D. Frisby and K. Mengelberg, Trans.). Routledge Classics. Abingdon, Oxon: Routledge. (Original work published 1900)

Skinner, W. (1974). The Focused Factory. *Harvard Business Review*, 52(3): 113-121.

Statista (2010). Prognose zum Anstieg von Zivilisationskrankheiten bis 2030 und 2050 gegenüber dem Jahr 2007. `https://de.statista.com/statistik/daten/studie/153966/umfrage/prognose-zum-anstieg-von-zivilisationskrankheiten-bis-2050/#professional`.

Statistisches Bundesamt (2020a). Erwerbstätigkeit: Eckzahlen zum Arbeitsmarkt, Deutschland. `https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbstaetigkeit/Tabellen/eckwerttabelle.html`.

Statistisches Bundesamt (2020b). Krankenhäuser: Einrichtungen, Betten und Patientenbewegung. `https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Krankenhaeuser/Tabellen/gd-krankenhaeuser-jahre.html;jsessionid=0352C77002628D12C0F20569F5588BC1.internet8741`.

Staudt, F. (2020). Ökonomie in der Medizin – ethische Aspekte. *Monatsschrift Kinderheilkunde*, 168(9): 828-833.

Tiwari, V. and Heese, H. (2009). Specialization and Competition in Healthcare Delivery Networks. *Health Care Management Science*, 12(3): 306-324.

Veit, C., Hertle, D., Bungard, S., Trümmer, A., Ganske, V. and Meyer-Hofmann, B. (2012). Pay-For-Performance im Gesundheitswesen: Sachstandsbericht zu Evidenz und Realisierung sowie Darlegung der Grundlagen für eine künftige Weiterentwicklung. Ein Gutachten im Auftrag des Bundesministeriums für Gesundheit. BQS-Institut.

Vincent, R., Sperling, N. M., Oates, J. and Jindal, M. (2006). Surgical Findings and Long-Term Hearing Results in 3,050 Stapedotomies for Primary Otosclerosis: A Prospective Study with the Otology-Neurotology Database. *Otology & Neurotology*, 27(8 Suppl. 2): S25-S47.

WAFA, H. A., WOLFE, C. D. A., EMMETT, E., ROTH, G. A., JOHNSON, C. O. and WANG, Y. (2020). Burden of Stroke in Europe: Thirty-Year Projections of Incidence, Prevalence, Deaths, and Disability-Adjusted Life Years. *Stroke*, 51(8): 2418-2427.

WATSON, G. J., BYTH, K. and DA CRUZ, M. (2015). Outcomes in Stapedotomy Surgery: The Learning Curve Redefined. *Otology & Neurotology*, 36(10): 1601-1603.

WEBER, M. (2011). *Wissenschaft als Beruf* (11th ed.). Berlin: Duncker und Humblot.

WEGNER, I. (2018). *Stapes Surgery for Otosclerosis: Towards Individualized Patient Care.* Utrecht University, Utrecht.

WEGNER, I., VERHAGEN, J. J., STEGEMAN, I., VINCENT, R. and GROLMAN, W. (2016). A Systematic Review of the Effect of Piston Diameter in Stapes Surgery for Otosclerosis on Hearing Results. *Laryngoscope*, 126(1): 182-190.

WEHKAMP, K.-H. (2002). DRG als medizinethische Herausforderung. *Deutsche Medizinische Wochenschrift*, 127(8): 395-398.

WILKINS, E., WILSON, L., WICKRAMASINGHE, K., BHATNAGAR, P., LEAL, J., LUENGO-FERNANDEZ, R., BURNS, R., RAYNER, M. and TOWNSEND, N. (2017). European Cardiovascular Disease Statistics 2017.

WOLTERS, U., WOLF, T., STÜTZER, H. and SCHRÖDER, T. (1996). ASA Classification and Perioperative Variables as Predictors of Postoperative Outcome. *British Journal Of Anaesthesia*, 77(2): 217-222.

WORLD HEALTH ORGANIZATION (2020). Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. `https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death`.

XIAN, Y., XU, H., SMITH, E. E., SAVER, J. L., REEVES, M. J., BHATT, D. L., HERNANDEZ, A. F., PETERSON, E. D., SCHWAMM, L. H. and FONAROW, G. C. (2022). Achieving More Rapid Door-to-Needle Times and Improved Outcomes in Acute Ischemic Stroke in a Nationwide Quality Improvement Intervention. *Stroke*, 53(4): 1328-1338.

YAFASOVA, A., FOSBØL, E. L., JOHNSEN, S. P., KRUUSE, C., PETERSEN, J. K., ALHAKAK, A., VINDING, N. E., TORP-PEDERSEN, C., GISLASON, G. H., KØBER, L. and BUTT, J. H. (2021). Time to Thrombolysis and Long-Term Outcomes in Patients With Acute Ischemic Stroke: A Nationwide Study. *Stroke*, 52(5): 1724-1732.

YUNG, M. W. and OATES, J. (2007). The Learning Curve in Stapes Surgery and Its Implication for Training. In: ARNOLD, W. and HÄUSLER, R. (Eds.): *Otosclerosis and Stapes Surgery*, 361-369. Karger, Basel. Advances in Oto-Rhino-Laryngology 65.

ZHAN, C. and MILLER, M. R. (2003). Excess Length of Stay, Charges, and Mortality Attributable to Medical Injuries During Hospitalization. *Journal Of The American Medical Association*, 290(14): 1868-1874.