# Qualitätstransparenz in Deutschland und deren Anwendung zur Steigerung der Behandlungsqualität im Krankenhaus

vorgelegt von
Christoph Pross, MSc
geboren in Salzwedel

vor der Fakultät VII – Wirtschaft und Management
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
– Dr. rer. oec. –

genehmigte Dissertation

**Promotionsausschuss:**

Vorsitzender: Prof. Dr. Frank Straube

Gutachter: Prof. Dr. Reinhard Busse

Gutachter: Prof. Dr. Jonas Schreyögg

Tag der wissenschaftlichen Aussprache: 17. Juli 2017

Berlin 2017

*„Es ist nicht genug zu wissen, man muss auch anwenden;*

*es ist nicht genug zu wollen, man muss auch tun“*

*“Knowing is not enough; we must apply.*

*Willing is not enough; we must do.”*

—Johann Wolfgang von Goethe, Wilhelm Meister’s Wanderjahre

# *Vorwort*

Die wissenschaftliche Arbeit für diese Dissertation wurde als Teil des Forschungsprojektes „Qualitätstransparenz in Deutschland und deren Anwendung zur Steigerung der Behandlungsqualität im Krankenhaus“ am Fachgebiet Management im Gesundheitswesen an der Technischen Universität Berlin ausgeführt. Die Arbeit und ihre Ergebnisse wurden gefördert durch Diskussionen am Fachgebiet Management im Gesundheitswesen und durch Aussprachen mit Forschern an der Georg-August-Universität Göttingen und der Universität zu Lübeck und Experten der Boston Consulting Group. Des Weiteren wurde die Arbeit fundiert durch Präsentationen und Diskussionen auf nationalen und internationalen Konferenzen und Seminaren. Finanziell unterstützt wurde die Arbeit durch ein Promotionsstipendium der Konrad-Adenauer-Stiftung, wofür ich sehr dankbar bin. Des Weiteren haben die Projektpartner – das Wissenschaftliche Institut der AOK (WIdO), die Deutsche Schlaganfallgesellschaft (DSG) und die Kooperation für Transparenz und Qualität im Gesundheitswesen (KTQ) – die Forschungsarbeit mit Daten und Know-how unterstützt. Den Projektpartnern danke ich für diese Hilfe.

Im Besonderen möchte ich meinem Betreuer Prof. Dr. med. Reinhard Busse für seine wertvolle Unterstützung und guten Ratschläge während der Forschungsarbeit und für seinen Beitrag beim Aufbau des Forschungsnetzwerkes danken. Ich möchte auch Dr. Alexander Geissler für hilfreiche Ideen, spannende Diskussionen und wichtige Kommentare danksagen. Des Weiteren danke ich den Kollegen am Fachgebiet Management im Gesundheitswesen für ihre Unterstützung und hilfreichen Impulse.

Die Arbeit möchte ich den Ärzten widmen, die sich täglich in einem System mit vielen Widerständen für die qualitativ bestmögliche Versorgung ihrer Patienten einsetzen.

# Inhaltsverzeichnis


| | | |
|---|---|---|
| Zusammenfassung | | v |
| Abstract | | vii |
| Einleitung | | 1 |
| Kapitel 1: | Qualität im Spannungsfeld zwischen Patienten und Leistungserbringern | 14 |
| Kapitel 2: | Qualitätstransparenz im internationalen Vergleich | 37 |
| Kapitel 3: | Qualitätsunterschiede zwischen deutschen Krankenhäusern | 87 |
| Kapitel 4: | Nutzerverhalten auf Public Reporting Portalen am Beispiel weisse-liste.de | 100 |
| Kapitel 5: | Spezialisierung und Zertifizierung im Krankenhaus | 117 |
| Kapitel 6: | Qualitätswettbewerb im deutschen Krankenhausmarkt | 147 |
| Kapitel 7: | Qualität und Ressourceneffizienz im Krankenhaus | 186 |
| Diskussion der Ergebnisse, Limitationen und Ausblick | | 231 |

# Zusammenfassung

Hintergrund und Ziele: Aufgrund hoher Informationsasymmetrien zwischen Patienten und behandelnden Ärzten haben Patienten nur geringe Möglichkeiten, die Behandlung im Krankenhaus und deren Erfolgsaussichten ex ante einzuschätzen. Qualitätstransparenz kann helfen, Informationsasymmetrien zu reduzieren und Patientenmündigkeit zu stärken. Jedoch ist das Wissen über Qualitätstransparenz und deren Wirkung fragmentiert, veraltet und im deutschen Kontext nur ansatzweise vorhanden. Die Qualität im Krankenhaus soll gesteigert werden, doch ist der Zusammenhang zwischen vielen Strukturfaktoren und Ergebnisqualität unzureichend erforscht. Diese Forschungsarbeit soll die wissenschaftliche Grundlage für den politischen Diskurs zu Transparenz und Verbesserung von Behandlungsqualität im Krankenhaus ausbauen und konkrete Empfehlungen für die Weiterentwicklung der Qualitätssicherung und den Ausbau qualitätsfördernder Strukturfaktoren in Deutschland geben.

Methodik: Als primäre empirische Grundlage fungiert eine umfassende Datenbank, welche die in Deutschland verfügbaren Struktur-, Prozess- und Ergebnisdaten auf Krankenhaus- und Fachabteilungsebene integriert. Zusätzlich werden Nutzerdaten des weisse-liste.de Public Reporting Portals verwendet. Methodisch interdisziplinär nutzt diese Arbeit verschiedene Ansätze und Techniken aus der Versorgungsforschung, der Gesundheitsökonomie und Big Data Marketing Forschung. Zum einen werden Qualitätstransparenzsysteme qualitativ untersucht. Dabei stehen insbesondere die Messung, Veröffentlichung und Incentivierung von Behandlungsqualität im Krankenhaus im Mittelpunkt eines internationalen Vergleiches. Zum anderen untersuchen Regressionsanalysen und eine stochastische Analyse der Effizienzgrenzen den Einfluss bestimmter Strukturfaktoren auf Qualität und Effizienz in der Krankenhausversorgung. Als Zielvariable wird primär Ergebnisqualität betrachtet.

Ergebnisse: Kapitel 1 untersucht die Einflussfaktoren auf Behandlungsqualität aus Patienten- und Leistungserbringersicht und erläutert die Konzepte Qualitätssicherung und Public Reporting. Als internationale Vergleichsarbeit identifiziert Kapitel 2 gesundheitspolitische Ansätze zur Steigerung von Qualitätstransparenz in fünf OECD Ländern. Unter anderem spielen Regierungen eine zentrale Rolle in der Etablierung von Standards und Anreizen für Qualitätstransparenz und integrierte Datensysteme in der Gesundheitsversorgung. Qualitätstransparenz benötigt einen stärkeren Fokus auf Ergebnisqualität und sollte sich auf Krankenhaus- und Leistungsbereichsebene konzentrieren, um Vergleichbarkeit und sinnvolle Wahlmöglichkeiten für Patienten zu garantieren. Kapitel 3 richtet den Fokus auf Deutschland und zeigt große Unterschiede in der risiko-adjustieren Qualität zwischen Krankenhäusern auf

nationaler und regionaler Ebene für ausgewählte Notfall- und elektive Leistungsbereiche. Je nach Leistungsbereich liegt das risiko-adjustierte Mortalitäts- bzw. Komplikationsrisiko bis zu 50-mal höher in Krankenhäusern mit schlechter Qualität im Vergleich zu den qualitativbesten Krankenhäusern. Kapitel 4 untersucht Nutzerverhalten in der Krankenhaussuche auf dem Public Reporting Portal „WeisseListe.de". Die Nutzung ist im Zeitraum von 2013 bis 2015 um 38% auf 2 750 Nutzer pro Tag angestiegen. Nutzer suchen insbesondere Qualitätsinformationen für onkologische und orthopädische Leistungsbereiche, doch sind für nur 5 der 20 meistgesuchten Leistungsbereiche Prozess- oder Ergebnisindikatoren verfügbar. Dicht besiedelte Gebiete in Westdeutschland mit höherer Krankenhausdichte weisen eine stärkere Portalnutzung auf als Ost- oder Süddeutschland. Unterschiedliche Nutzergruppen mit spezifischen Charakteristika konnten identifiziert werden und verdeutlichen die Notwendigkeit für nutzerorientiertes Public Reporting. Kapitel 5 untersucht den systematischen Einfluss von Spezialisierung der Krankenhausinfrastruktur und -prozesse und von Zertifizierung auf Qualität am Beispiel der Schlaganfallbehandlung. Krankenhäuser mit einer Stroke Unit haben eine 6% geringere Schlaganfallsterblichkeit. Ein zusätzlicher systematischer Einfluss für ein Stroke Unit Zertifikat kann nicht festgestellt werden. Kapitel 6 richtet den Fokus auf den Krankenhausmarkt als Ganzes und analysiert den Einfluss von Qualitätswettbewerb auf die Behandlungsqualität in den Leistungsbereichen Schlaganfall, Herzinfarkt und Hüft-Total-Endoprothetik. Ein signifikanter Einfluss von Krankenhauswettbewerb auf die Qualität der Versorgung kann nicht festgestellt werden, was möglicherweise auf das weiterhin unzureichende öffentliche Bewusstsein für Qualitätsunterschiede und gegenläufige Effekte durch Krankenhauszentralisierung und Spezialisierung zurückzuführen ist. Letztlich kombiniert Kapitel 7 Qualität und Ressourceneffizienz in einer neuen Methode zur Bestimmung qualitätsadjustierter technischer Krankenhauseffizienz. Dabei werden deutliche Effizienzpotentiale identifiziert, welche insbesondere durch Spezialisierung und Konzentration von Leistungen zu adressieren sind.

<u>Fazit</u>: Die Arbeit verdeutlicht Verbesserungspotentiale im Qualitätstransparenzsystem und in der Versorgungsqualität und Effizienz deutscher Krankenhäuser. Neben den aktuellen Bestrebungen im Rahmen des Krankenhausstrukturgesetzes kann ein internationaler Best Practice Austausch, Qualitätssicherung mit tatsächlichen Konsequenzen für schlechte Qualität, ein stärkerer Fokus auf Ergebnisqualität, Spezialisierung und Zentralisierung in der Versorgung, eine nutzenorientierte Verbesserung des Public Reportings und eine weitere Stärkung des Qualitätswettbewerbs die Versorgungsqualität in Deutschland nachhaltig steigern.

# Abstract

Background and objectives: Due to large information asymmetries between patients and physicians, patients often have limited opportunity to assess treatment options, process of care and their prospective treatment outcomes in advance. Quality of care transparency can help to reduce information asymmetries and increase patient autonomy. Yet, the scientific knowledge around quality of care transparency and its effects on outcomes is fragmented, outdated and, in the German context, rudimentary. In order to enhance quality transparency and to improve hospital outcome quality, the dynamics around quality of care transparency and the relationship between important structural factors and outcomes must be better understood. This research project aims to expand the scientific foundation for policymaking on quality transparency and quality improvement in hospital care. The goal is to provide concrete policy recommendations for improvement of quality transparency and the expansion of quality-enhancing structural factors in the German hospital sector and beyond.

Methods: The empirical foundation for this project is a comprehensive database, which integrates the available structural, procedural and outcome quality metrics at both a hospital and medical department level. In addition, website data for the weisse-liste.de public reporting portal is used. Based on an interdisciplinary health economics approach, this project employs different methodologies from health systems research, health economics and big data e-commerce marketing research. First, this project investigates quality of care transparency systems qualitatively by analyzing how they measure, report and incentivize better outcomes. Second, this project uses regression analysis and stochastic efficiency modeling to examine the influence of certain structural factors on quality and efficiency in hospital service delivery. Outcome quality, as opposed to process quality, is the main focus and dependent variable.

Results: Chapter 1 examines the factors impacting quality of care on the patient demand side and the care provider supply side, as well as reviews quality assurance and public reporting in the German context. As an international comparison, Chapter 2 examines health policy approaches to enhance quality of care in five mature OECD health systems. Among others, the results suggest governments play an essential role in establishing standards and incentives for quality of care transparency and integrated IT systems in heath service delivery. Furthermore, quality transparency needs a stronger emphasis on outcome quality and should focus at the medical condition and hospital level to ensure patients are empowered with accessible and meaningful information to guide healthcare decisions. Chapter 3 sets the focus on Germany and uncovers large differences in risk-adjusted outcomes between hospitals—both at a national and regional level—as well as for emergency and elective conditions. Depending on the treatment area, the risk-adjusted mortality or complication risk is up to 50 times higher in hospitals with poor outcome quality, relative to hospitals with the best outcome quality of care. Chapter 4 examines website user patterns for the hospital search function of the public reporting portal weisse-liste.de. Between 2013 and 2015, usage of the website increased by 38% annually, to 2,750 user sessions per day. In particular, users searched for quality of care information for oncological and orthopedic treatment areas. Process and outcome quality metrics are only available for 5 of 20 most searched-for treatment areas. In Western Germany, users from densely populated areas with a high number of hospitals conducted relatively more searches than users in Southern and Eastern Germany. From the website data, user groups with distinct characteristics could be identified and confirm the need for customized public reporting options. Chapter 5 evaluates the systematic influence of hospital certification, hospital infrastructure, and process specialization on stroke care outcome quality. Hospitals with a dedicated stroke unit have, on average, a 6% lower risk-adjusted 30-day stroke mortality, ceterius paribus. However, a stroke unit certification showed no systematic influence on 30-day stroke mortality. Chapter 6 focuses on the provider market at large—examining the influence of provider competition on outcomes for hip replacement surgery, and stroke and heart attack emergency care. The influence of provider competition on outcomes remains unclear and insignificant, which can be partially attributed to currently insufficient awareness about outcome differences and where to access the right information. It also hints at the potentially countervailing outcomes influences of hospital centralization and specialization. Chapter 7 combines the concepts of quality and efficiency of care in hospital service provision in a new stochastic frontier analysis model, estimating quality-adjusted technical hospital efficiency. Large inefficiencies are identified, which can be addressed in part through specialization and concentration in care delivery.

Conclusions: This research project illustrates substantial potential for improvement in the German quality transparency system, as well as opportunities to enhance the quality of care and efficiency in German hospital service delivery. Possible solutions to increase quality of care in Germany may include, among others: an international best practice exchange, stronger focus on outcome quality, actual regulatory consequences for hospitals with insufficient quality of care, specialization and centralization in hospital care, a user oriented public reporting system, and a substantial strengthening of hospital quality of care competition.

# Einleitung

Die Qualität in der Leistungserbringung ist in den letzten Jahren national und international zu einem der wichtigsten Themen der Versorgungsforschung und der Gesundheitspolitik geworden. Ein wichtiges Anzeichen für die unzureichende Qualität in den Gesundheitssystemen ist die hohe Anzahl vermeidbarer Sterbefälle aufgrund von medizinischen Behandlungsfehlern. Der Report *`To Err is Human´* des *Institute of Medicine* (IOM) aus dem Jahr 2000 hat erstmals das Ausmaß der vermeidbaren, iatrogenischen (d.h. durch ärztliche Behandlung verursachten) Sterbefälle in den USA aufgezeigt. Er schätzt diese auf 44 000 bis 98 000 Fälle [1]. Eine weitere, aktuellere Studie aus den USA erhöht die Zahl auf mindestens 210 000 – 400 000 vermeidbare Sterbefälle [2], womit iatrogenische Sterbefälle die dritthäufigste Todesursache in den USA sind.

Auch Deutschland hat trotz des hohen Ressourceneinsatzes im Gesundheitssystem Nachholbedarf bezüglich der Behandlungsqualität [3, 4]. Nach Schätzungen des Aktionsbündnisses Patientensicherheit treten in Deutschland in 5-10% aller Krankenhausfälle vermeidbare medizinische Fehler auf, welche in 0,1% aller Behandlungsfälle – bei 19 Millionen stationären Patienten[1] im Jahr 2015 19 000 Fälle – zu vermeidbaren Sterbefällen führen [5–7]. Mehr als 70% der vermeidbaren medizinischen Fehler verteilen sich dabei auf chirurgische und internistische Fächer [8]. Medizinische Behandlungsfehler sind eine häufige Todesursache und erfordern einen stärkeren Fokus auf die unzureichende Qualität und eine Fehlervermeidung in der Leistungserbringung.

Als Reaktion auf die unzureichende Qualität in der Leistungserbringung hat die deutsche Bundesregierung im Zuge ihrer Qualitätsoffensive Ende 2015 das Krankenhausstrukturgesetz (KHSG) zur Stärkung der Qualität im Krankenhaus verabschiedet [9]. Deutschland folgt damit

[1] Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten für beide Geschlechter.

einem Trend, der auch in anderen Ländern zu beobachten ist. In den USA treibt das *Center for Medicare and Medicaid Services* (CMS) im Zuge der Implementierung des *Affordable Care Acts* (ACA) die qualitätsorientierte Vergütung voran [10]. Die US-amerikanischen privaten Krankenkassen und Versorger entwickeln immer umfassendere Erstattungsansätze, in denen der komplette Behandlungsverlauf inklusive der Nachversorgung in einem Erstattungsbündel enthalten ist [11, 12]. Ebenso hat die niederländische Regierung das Jahr 2015 zum Jahr der Qualitätstransparenz erklärt und weitreichende Reformen zur verbesserten Messung und Veröffentlichung von Krankenhaus-Qualitätsindikatoren angestoßen [13].

Neben der hohen Anzahl an vermeidbaren Sterbefällen reagieren Deutschland und andere Länder damit auch auf andere Erkenntnisse. Erstens wachsen die Ausgaben für Gesundheit weiterhin stark [14]. Einmal mehr setzt sich die Erkenntnis durch, dass die umfangreichen Mittel im Gesundheitssystem effizienter eingesetzt werden müssen, um den langfristig steigenden Gesundheitsausgaben zu entgegnen [15, 16]. Zweitens zeigen sich immer deutlicher Qualitätsunterschiede zwischen und innerhalb einzelner Gesundheitssysteme [17]. Unterschiede zwischen Leistungserbringern innerhalb eines Landes beunruhigen Patienten, Krankenkassen und Versorgungsforscher [18–21].

Als eine der wichtigsten und erfolgversprechendsten Reformen im Gesundheitswesen wird dabei die Qualitätstransparenz – die Messung und Veröffentlichung von Qualitätsindikatoren wie Mortalitäts- und Komplikationsraten – angesehen [15, 21–24]. Transparenz soll den Qualitätswettbewerb fördern und Leistungsanbieter so zu mehr Investitionen in Qualitätsverbesserung bewegen [25]. Dies soll die Behandlungsqualität erhöhen [26]. Qualitätstransparenz kann helfen, Strukturfaktoren mit besonderer Bedeutung für eine hohe Ergebnisqualität (d.h. Mortalitäts- oder Komplikationsraten) zu identifizieren. Durch induzierte Qualitätssteigerung kann Transparenz darüber hinaus die Folgekosten schlechter medizinischer Behandlung reduzieren [15, 24] Zusätzlich kann Qualitätstransparenz den ineffizienten Einsatz von Ressourcen aufzeigen [15]. Dennoch sind Qualitätstransparenz und Qualitätswettbewerb zwischen Leistungserbringern in Deutschland als auch international umstritten und deren wissenschaftliche Fundierung wie im Folgenden erläutert unzureichend [27, 28]. Die stärkere Integration von Qualität in den Krankenhauswettbewerb führt zu Unsicherheiten über Potentiale und Konsequenzen bei den Leistungserbringern [29, 30].

Obwohl alle Gesundheitssysteme mit Qualitätsunterschieden und dem ineffizienten Einsatz von Ressourcen zu kämpfen haben, hat sich die Forschung zum Thema Qualitätsmessung und -transparenz lange hauptsächlich auf das US-amerikanische Gesundheitssystem und den britischen *National Health Service* (NHS) konzentriert. Die vorliegende Arbeit schließt an diese Studien an und ergänzt die internationale Forschung zum Thema Qualitätstransparenz um einen umfassenden Beitrag zum deutschen Gesundheitswesen.

In Deutschland gibt es zwar einzelne Arbeiten zum Thema, so zum Beispiel die Krankenhausreports 2005 und 2014 [6, 23] oder zur Entwicklung der Qualität im Zuge der Qualitätssicherung [31, 32], aber Fundierung der wissenschaftlichen Grundlagen fehlen in wichtigen Bereichen. Dies ist u.a. darin begründet, dass die standardisierte Qualitätsdatenmessung erst 2006 gestartet wurde und Zeitreihenanalysen über einen längeren Zeitraum daher erst seit Kurzem möglich sind. Daher soll nun den aktuellen gesundheitspolitischen Bemühungen zur Steigerung der Qualitätstransparenz in Deutschland eine umfangreichere wissenschaftliche Grundlage geboten werden. Der Wirkungsmechanismus von Qualitätstransparenz und Qualitätsverbesserung im Krankenhaus soll dabei möglichst ganzheitlich analysiert werden. Wichtige Strukturfaktoren für eine positive Qualitätsentwicklung werden identifiziert. Darauf basierend werden fundierte Empfehlungen für den Ausbau von Qualitätstransparenz und qualitätssteigernden Strukturfaktoren in Deutschland gegeben.

*Relevante Forschungslücken zu Qualitätstransparenz und Qualitätssteigerung*

Gemessen an dem hohen Grad der Abhängigkeit zwischen Messung, Veröffentlichung und Incentivierungsmöglichkeiten von Behandlungsqualität gibt es auch international bisher wenig Überblicksarbeiten, die mehrere der Kernelemente von Qualitätstransparenz gleichzeitig betrachten [33–36]. Zum Beispiel sind sowohl Veröffentlichungen als auch Systeme zur qualitätsorientierten Vergütung abhängig davon, welche Qualitätsindikatoren gemessen und durch Risikoadjustierung vergleichbar gemacht werden [37, 38].

In Deutschland gibt es trotz der Verfügbarkeit von Daten über einen Zeitraum von 10 Jahren im Rahmen der externen Qualitätssicherung und anderer Qualitätstranparenzinitiativen bis jetzt nur wenige Studien, welche die Qualitätsentwicklung, insbesondere für die Ergebnisqualität, über die Zeit hinweg betrachten [31, 32]. Die AQUA und IQTIG Qualitätsberichte vergleichen die Qualität ausschließlich zum vorherigen Jahr und referenzieren nur vereinzelt längere Zeiträume [39–41], obwohl eine Vergleichbarkeit der Indikatoren über die Jahre gerade auch ein Anliegen der externen Qualitätssicherung ist. Eine Analyse der Indikatoren der externen Qualitätssicherung für die Jahre 2006 bis 2008 fand eine Verbesserung für die Prozessindikatoren, aber eine Verschlechterung für nicht risikoadjustierte Ergebnisindikatoren [42]. Eine für 2006-2013 durchgeführte Trendanalyse der Indikatoren für die gesetzliche Qualitätssicherung in Bayern hat für eine Mehrheit der 123 untersuchten Indikatoren einen signifikanten positiven Trend festgestellt [32]. Jedoch gab es auch einige Indikatoren, insbesondere Ergebnisindikatoren, welche sich negativ entwickelt haben. Eine Analyse von sechs Indikatoren – 2 Diagnose-, 3 Prozess- und 1 Ergebnisindikator – der externen Qualitätssicherung hat eine positive Veränderung ergeben und diese auch in den Zusammenhang mit der Veröffentlichung der Daten gebracht [31]. Untersuchungen auf Basis von DRG Routinedaten und nationale Registerdaten haben für verschiedene Leistungsbereiche positive als auch negative Ergebnisqualitätstrends identifiziert [43–46]. Auch im internationalen Kontext liefern Studien widersprüchliche Ergebnisse in Bezug auf die Verbesserung von Ergebnisqualität [19, 47–54] und den Einfluss von Public Reporting Systemen [37, 55, 56].

Die Bundesregierung und die Organe der Selbstverwaltung gehen davon aus, dass die Informationen für Patienten schnell und einfach zugänglich sind und diese die Informationen in ihre Krankenhauswahlentscheidung als wichtige Kriterien integrieren [57]. Doch gibt es sowohl in Deutschland als auch international keine Studien, die untersuchen, wie Public Reporting Portale tatsächlich genutzt werden und wie sich typische Nutzergruppen auf Onlineportalen verhalten, welche Informationen sie suchen und wann und warum sie ihre Informationssuche abbrechen [37, 58, 59]. Als eine Ausnahme gilt eine aktuelle Studie aus den USA, welche aggregierte Website Analytics Daten von USA-basierten Public Reporting Portalen und Umfrageergebnisse von Nutzern dieser Portale untersucht hat [60]. Diese Studie identifizierte eine hohe Bereitschaft zur Krankenhauswahl auf Basis von Public Reporting Informationen und eine tendenziell bessere Erfahrung für Nutzer mit medizinischer Fachexpertise. Insgesamt gibt es großen Nachholbedarf bzgl. des Nutzungsverhaltens von Patienten und Ärzten auf Reporting Portalen und des damit verbundenen Informationsgewinns. Patienten haben ein starkes Eigeninteresse an der Auswahl der besten Leistungsanbieter. Und gerade weil Preissignale im deutschen als auch anderen europäischen Gesundheitssystemen fehlen, würden sich Patienten bei entsprechender Informationsverfügbarkeit und -verständlichkeit zusätzlich auf das Auswahlkriterium Qualität fokussieren[61].

Ein funktionierendes Public Reporting gilt als Schlüssel für den Qualitätswettbewerb zwischen Krankenhäusern. Die Auswirkungen von Krankenhauswettbewerb zwischen Kliniken wurden bisher primär im US-amerikanischen Gesundheitssystem und britischen NHS untersucht [62–66]. Für Einheitspreissysteme, wie dem amerikanischen Medicare oder dem NHS System, wurde oft ein positiver Effekt des Wettbewerbs auf die Qualität identifiziert [64], doch gibt es auch widersprüchliche Ergebnisse [67]. Neben den unklaren Resultaten auf internationaler Ebene gibt es in Deutschland nur wenige Studien, die sich mit der Krankenhausmarktstruktur beschäftigen [68–70] und keine Studien, welche die Auswirkungen von Krankenhauswettbewerb auf Behandlungsqualität empirisch im deutschen Krankenhausmarkt untersuchen. Diese Forschungslücke wird auch durch eine großangelegt Studie des Bundeskartellamtes zu den Auswirkungen von Krankenhauswettbewerb auf die Qualität der Behandlung und Wahlmöglichkeit von Patienten evident [71].

Neben dem Public Reporting von Qualitätsindikatoren und dem darauf basierenden Qualitätswettbewerb zwischen Leistungserbringern wird Behandlungsqualität auch durch Spezialisierung auf bestimmte Leistungsbereiche gesteigert [72]. Diese Spezialisierung wird oft durch eine Zertifizierung auf Leistungsbereichsebene, wie z.B. die Stroke Unit Zertifikate der Deutschen Schlaganfallgesellschaft (DSG) und die Zentren Zertifikate der Deutschen Krebsgesellschaft, von außen verifiziert. Zertifikate bestimmen Infrastruktur-, Personal-, und Prozessanforderungen, welche zu besserer Behandlungsqualität führen sollen. Sie haben Signalwirkung und sollen damit, wie auch Public Reporting, Patienten besondere Leistungszentren mit hoher Erfahrung und Behandlungsqualität verdeutlichen. Oft werden die hohen Kosten für Zertifikate auch mit dadurch verbesserter Behandlungsqualität begründet. Doch fragen sich viele Krankenhausmanager und Gesundheitspolitiker, wie stark Zertifizierung zur Verbesserung der Qualität beitragen kann, zusätzlich zur Spezialisierung, die meist vorher schon stattgefunden hat. Aber gerade in der fehlenden Unterscheidung zwischen Spezialisierung und Zertifizierung und deren jeweilige Auswirkung auf die Ergebnisqualität besteht eine wichtige Forschungslücke. Zusätzlich wird in der Literatur Spezialisierung oft gleichgesetzt mit der Fallzahl, doch ist es auch wichtig zu unterscheiden zwischen Infrastruktur- und Prozessspezialisierung und reiner Volumensteigerung [73]. Letztlich sind Spezialisierung und Zertifizierung mit hohen Investitionsmitteln und operativen Kosten verbunden und rufen im Zusammenhang mit dem steigenden Kostendruck im Gesundheitswesen die Frage nach Effizienz in der Krankenhausbehandlung auf.

Die Effizienzliteratur im Bereich Krankenhaus auf Basis von Dateneinhüllanalyse (DEA) und stochastischer Grenzwertanalyse (Stochastic-Frontier Analysis, SFA) Modellen ist umfangreich, doch findet Behandlungsqualität hier bisher unzureichende Betrachtung. [74, 75]. Dies ist einer der Hauptgründe, warum bisher keine Übersetzung der Erkenntnisse der Effizienzliteratur in gesundheitspolitische Maßnahmen stattgefunden hat [75]. Zusätzlich basieren existierende Studien meist auf limitierten Krankenhausstichproben und Datenzeiträumen, untersuchen Effizienz auf Ebene des Gesamtkrankenhauses und nicht auf Ebene eines Leistungsbereiches und haben meist räumliche Zusammenhänge außen vorgelassen [74, 76–79]. Eine Adressierung dieser Defizite in der Krankenhauseffizienzschätzung ist wichtig, um mögliche Fehlerquellen auszuschließen und die Relevanz für Anwender in Gesundheitspolitik und Unternehmen zu erhöhen.

Zum Zeitpunkt der Einführung der externen Qualitätssicherung hat der Krankenhausreport 2004 schon ähnliche grundsätzliche Fragen formuliert [23]:

1. Welchen Wert haben Qualitätsindikatoren und deren Darstellungsform für Patienten?
2. Welche Auswirkung haben die interne Verwendung und externe Veröffentlichung von Qualitätsindikatoren auf die Versorgung?
3. Funktioniert der durch Qualitätstransparenz angestoßene Qualitätswettbewerb?

Obwohl diese Fragen eine hohe und weiter steigende Bedeutung für das deutsche Gesundheitssystem haben, sind sie bisher unzureichend beantwortet wurden.

Um die Arbeit des IQTIGs zu unterstützen und die Möglichkeiten und Gefahren von Qualitätstransparenz in Deutschland umfänglich zu beleuchten, hat eine Adressierung der identifizierten Forschungslücken eine hohe Priorität für die deutsche Versorgungsforschung und Gesundheitsökonomie. Gleichzeitig gilt es die bisherigen Maßnahmen im Bereich der Qualitätssicherung zu evaluieren und darauf basierend Methoden zu entwickeln, um relevante Struktur- und Prozessfaktoren mit klarer positiver Wirkung auf die Behandlungsqualität zu identifizieren und damit auch einen wichtigen internationalen Forschungsbeitrag zu leisten.

## *Ziele und Methoden*

Die Arbeit adressiert Forschungslücken entlang dreier Dimensionen von Qualität im Krankenhaus (vgl. Abbildung 1). Die erste Dimension – die Messung von Qualität – bereitet die Grundlage von Qualitätstransparenz. Sie entscheidet über Umfang (Anzahl der Krankenhäuser) und Detaillevel (Leistungsbereiche und Indikatoren von Qualitätstransparenz). Dimension zwei – die Veröffentlichung (Public Reporting) – ermöglicht eine qualitätsorientierte Krankenhaussuche und entsprechende Krankenhausvergleiche. Die Verbesserung der Qualität als dritte Dimension und ergebnisorientierte Anwendung der Dimensionen eins und zwei beinhaltet Qualitätssteigerung aufgrund eines von Transparenz beförderten Qualitätswettbewerbs und die durch Transparenz ermöglichte Identifikation von wichtigen Strukturfaktoren für die Steigerung der Ergebnisqualität.

**Abbildung 1**: Übersicht der Arbeit und der einzelnen Beiträge


**Legende**: Die Werkzeuge fungieren als Grundlage für die Dissertation. Die Forschungsergebnisse unterteilen sich in sechs Beiträge in internationalen Fachzeitschriften. Quelle: Eigene Darstellung

Als Hintergrund für die Evaluation und Anwendung des deutschen Qualitätstransparenzsystems wird in Kapitel 1 die theoretische Grundlage für die gesundheitsökonomische Betrachtung der Qualität im Krankenhaus gelegt. Daran anschließend wird in Kapitel 2 die Messung, Veröffentlichung und Incentivierung von Qualität in Deutschland und vier weiteren Ländern betrachtet und wichtige Maßnahmen zur Steigerung der Qualitätstransparenz abgeleitet. Im nächsten Schritt werden in Kapitel 3 und 4 wichtige Aspekte der aktuellen Entwicklung und Performanz des Qualitätstransparenzsystems untersucht und Vorschläge zur Verbesserung des Systems aufgezeigt.

Qualitätstransparenz ermöglicht die Evaluation von Strukturfaktoren und deren Einfluss auf die Ergebnisqualität im Krankenhaus frei nach der Managementmaxime ‚Was man nicht messen kann, kann man nicht lenken'. Daher soll die bisher erreichte Transparenz über Struktur- und Ergebnisqualität dafür genutzt werden, Strukturfaktoren zu identifizieren, die einen systematischen und positiven Einfluss auf die Ergebnisqualität haben und damit zu einer

Verbesserung (vgl. Abbildung 1) der Qualität und Ressourceneffizienz im Krankenhaus beitragen können.

Auf der Ebene des Krankenhauses wird in Kapitel 5 exemplarisch am Beispiel Schlaganfall der Effekt von Stroke Unit und Zertifizierung auf die Schlaganfallmortalität untersucht. In Kapitel 6 wird untersucht ob es in Deutschland bereits einen Qualitätswettbewerb zwischen Krankenhäusern gibt. Letztlich wird in Kapitel 7 eine Methode zur Schätzung von qualitätsadjustierter, technischer Krankenhauseffizienz entwickelt und am Beispiel Schlaganfall angewandt, um Strukturfaktoren zur Steigerung von Qualität und Ressourceneffizienz zu evaluieren.

Die Umsetzung der Forschungsziele basiert auf der Unterstützung durch ein Forschungsnetzwerk, einer umfangreichen Krankenhausdatenbank und der Anwendung von einem interdisziplinären Spektrum an Methoden aus der Gesundheitsökonomie und Versorgungsforschung. Im ersten Schritt wurde ein Forschungsnetzwerk (vgl. Abbildung 1, Werkzeug a) mit verschiedenen Partnern aufgebaut, die als Wissensträger und Datenlieferanten das Forschungsprojekt unterstützten. Dazu zählen das Wissenschaftliche Institut der Ortskrankenkassen (WIdO) mit der Initiative Qualitätssicherung mit Routinedaten (QSR), die Weisse Liste.de gGmbH (WL.de), die Deutsche Schlaganfallgesellschaft (DSG) und die Kooperation für Transparenz und Qualität im Gesundheitswesen (KTQ). Mit Hilfe dieser Partner, den öffentlich-verfügbaren G-BA Qualitätsberichten der deutschen Krankenhäuser und der Unterstützung von Programmierern und Datenbankexperten der Boston Consulting Group (BCG) wurde eine umfangreiche MySQL Krankenhaus-Datenbank (vgl. Abbildung 1, Werkzeug b) aufgebaut. Sie integriert einen Großteil der in Deutschland für Krankenhäuser verfügbaren Struktur-, Prozess- und Ergebnisdaten und macht sie über die Jahre 2006 - 2014 vergleichbar. Gleichzeitig wurde eine weitere MySQL Datenbank (vgl. Abbildung 1, Werkzeug c) erstellt, die Nutzerdaten der Public Reporting Webseite WL.de analysierbar macht.

Hauptbestandteil der Krankenhausdatenbank sind die im XML-Format freizugänglichen G-BA Qualitätsberichte. Die Datengrundlage von sechs Jahresintervallen (2006, 2008, 2010, 2012, 2013, 2014) umfasst Strukturdaten wie Art des Trägers, Adressdaten, Anzahl der Ärzte und Pflegekräfte, medizinische Ausstattung und Fallzahlen. Zusätzlich umfassen die Qualitätsberichte Prozessindikatoren wie Antibiotikaprophylaxe bei Hysterektomie und Bestimmung der postoperativen Beweglichkeit nach Neutral-Null-Methode und Ergebnisindikatoren wie Mortalität und risikoadjustierte Mortalität, Komplikationsrate, Re-operationsrate für 30 Leistungsbereiche wie Cholezystektomie, Hüft- und Knie-Endoprothesen-Implantation und Herzschrittmacher-Implantation [80, 81]. Die Qualitätsberichte wurden auf der Basis von Institutskennzeichen und anderen Krankenhausinformationen über die Jahre hinweg integriert. Sie enthalten auch Informationen auf Fachabteilungsebene wie Anzahl ICD Diagnosen, OPS Prozeduren und Arzt- und Pflegepersonal.

Auf Krankenhausebene wurde die Datenbank um weitere relevante Struktur- und Qualitätsparameter erweitert. QSR Ergebnisindikatoren für die Leistungsbereiche Schlaganfall und Herzinfarkt wurden integriert. Die Informationen der QSR Initiative enthalten für beide Leistungsbereiche die Anzahl behandelter Patienten, die 30-Tage Mortalität, den Erwartungswert für die 30-Tage Mortalität und die Anzahl an Patienten mit Wiedereinlieferung nach 30, 90 und 365 Tagen. Der Erwartungswert wurde auf zwei Arten berechnet. Auf der einen Seite wurde eine jährliche Risikoadjustierung auf Basis eines jahresspezifischen Modells durchgeführt. Auf der anderen Seite wurden die Krankenhaus- und jahresspezifischen Erwartungswerte auf Basis des 2014er Modells zur Risikoadjustierung für alle anderen Jahre rückwirkend berechnet, um eine Vergleichbarkeit sicherzustellen. Aus dem Verhältnis der realisierten mit der erwarteten 30-Tages-Mortalität errechnet sich die standardisierte Mortalitätsrate (SMR) [82, 83].

Neben den QSR Indikatoren wurden Informationen zu DSG und KTQ Zertifikaten für den Zeitraum 2006 – 2014 eingespielt. Diese Informationen geben wieder, welche Krankenhäuser in welchem Jahr die Zertifikate erhalten haben und erlauben die Bildung von Variablen zu Zertifizierungsstatus und Alter des Zertifikats. Zusätzlich wurden Daten aus den AOK Krankenhaus Directories und als letzter Bestandteil Indikatoren zur räumlichen Entwicklung aus der INKAR Datenbank des Bundesamtes für Raumwesen und Raumordnung integriert.

Die primäre Analyseebene dieser Arbeit sind das Krankenhaus und die entsprechenden Leistungsbereiche. Regionale Versorgungsqualität oder Qualität auf der Ebene eines einzelnen Arztes wird nicht betrachtet. Qualitätsdaten sind nur auf Krankenhausebene vorhanden, da die Daten schon vor Erhalt von der Patientenebene auf die Krankenhausebene aggregiert wurden. Da in den Qualitätsberichten der deutschen Krankenhäuser wichtige Strukturfaktoren, wie Anzahl von Ärzten und Pflegekräften, auch auf Fachabteilungsebene zur Verfügung stehen, können auch Mehrebenenmodelle geschätzt werden. Die Betrachtung der relevanten Fachabteilungen ist zum Beispiel für eine Effizienzanalyse der Schlaganfallversorgung wichtig.

Das Thema Qualitätstransparenz zeichnet sich durch eine hohe Interdisziplinarität aus. Dementsprechend sind auch die verwendeten Methoden interdisziplinär und umfassen einfache und komplexere quantitative Methoden sowie qualitative Ansätze. Auf der einen Seite helfen Experteninterviews, Recherchen wissenschaftlicher und grauer Literatur und Analysen von Public Reporting Portalen für den Ländervergleich zu Qualitätstransparenz in Kapitel 2. Auf der anderen Seite werden einfache deskriptive und komplexere quantitative Methoden angewandt. Kapitel 3 nutzt Box Plots und einfache lineare Regressionen zur Qualitätstrendanalyse. Kapitel 4 nutzt die hierarchische Clustering Ward-Minimum-Varianz-Methode auf Basis von Nutzungsvariablen zur Erstellung von Nutzerprofilen. Mit Hilfe der Markov-Ketten-Modellierung zeigen wir die häufigsten Navigationspfade und Ausstiegspunkte auf den Public Reporting Portal WeisseListe.de. Kapitel 5 und 6 nutzen Fixed-Effets-

Regressionsmodelle, um die Wirkung von Krankenhauspezialisierung, Zertifizierung und Krankenhauswettbewerb auf Ergebnisqualität zu untersuchen.

Neben bestehenden Methoden werden auch methodische Grundlagen entwickelt, mit denen fortlaufend die Qualität und die wichtigsten Einflussfaktoren auf Qualitätsverbesserung in verschiedenen Leistungsbereichen evaluiert werden können. So wird zum Beispiel in Kapitel 7 ein geoadditives Stochastic Frontier Analyse (SFA) Modell zur Schätzung qualitätsadjustierter Effizienz im Krankenhaus entwickelt. Diese Methoden können in anderen Ländern mit ähnlicher Datengrundlage angewandt werden.

Der weitere Aufbau der Arbeit ist wie folgt: Kapitel 1 präsentiert die Determinanten von Qualitätsentscheidungen für Patienten (d.h. Nachfrage) und Leistungserbringer (d.h. Anbieter). Kapitel 2-7 beinhalten die zur Veröffentlichung eingereichten Artikel in englischer Sprache. Jedem Artikel vorangestellt ist eine deutsche Zusammenfassung. Abschließend diskutiert Kapitel 8 Ergebnisse und Empfehlungen, beschreibt die Limitationen und gibt einen Ausblick zur Forschungsagenda.

# <u>Kapitel 1</u>: Qualität im Spannungsfeld zwischen Patienten und Leistungserbringern

Ein zentrales Anliegen in Bezug auf die Funktion des Gesundheitsmarktes ist, ob der Wettbewerb zwischen Leistungserbringern und dessen Regulierung das gesellschaftlich optimale Maß an Versorgungsqualität liefert. Dazu gilt es die verwendete Definition von Qualität zu klären, die Determinanten von Qualität auf Nachfrage- und Angebotsseite zu beleuchten, die Grundfunktion des Public Reporting und den aktuellen Stand der Qualitätssicherung in Deutschland zu betrachten. Mit diesem Hintergrund können dann die identifizierten Forschungslücken adressiert werden.

## *Was ist Qualität im Gesundheitswesen?*

Für die meisten Menschen ist Gesundheit das höchste Gut[2] und dementsprechend ist der Anspruch hoch, Gesundheit bei Krankheit wiederherzustellen. Dieser Anspruch begründet auch das in Deutschland solidarisch organisierte Gesundheitssystem, in dem jedem Menschen ein hohes Maß an Quantität und Qualität der Gesundheitsleistungen zugesichert wird, egal ob er dafür zahlen kann oder nicht. Doch ist Qualität im Gesundheitswesen keineswegs eindeutig definiert. Definitionen sind nicht statisch, sondern von Interessen, Wertvorstellungen und Zielen der definierenden Akteure abhängig [84, 85]. Auch wird Qualität in den verschiedenen medizinischen Anwendungsbereichen präventive, Akut- oder palliative Versorgung von akuten oder chronischen Krankheiten und der anschließenden Rehabilitation unterschiedlich definiert.

In Deutschland wird oft auf Definitionen der *International Organization of Standardization* (ISO) und des US-amerikanischen IOMs verwiesen [86]. Die ISO definiert Qualität als „die Gesamtheit von Eigenschaften und Merkmalen eines Produkts oder einer Dienstleistung, die

[2] „Gesundheit ist zwar nicht alles, aber ohne Gesundheit ist alles nichts.“ Arthur Schopenhauer

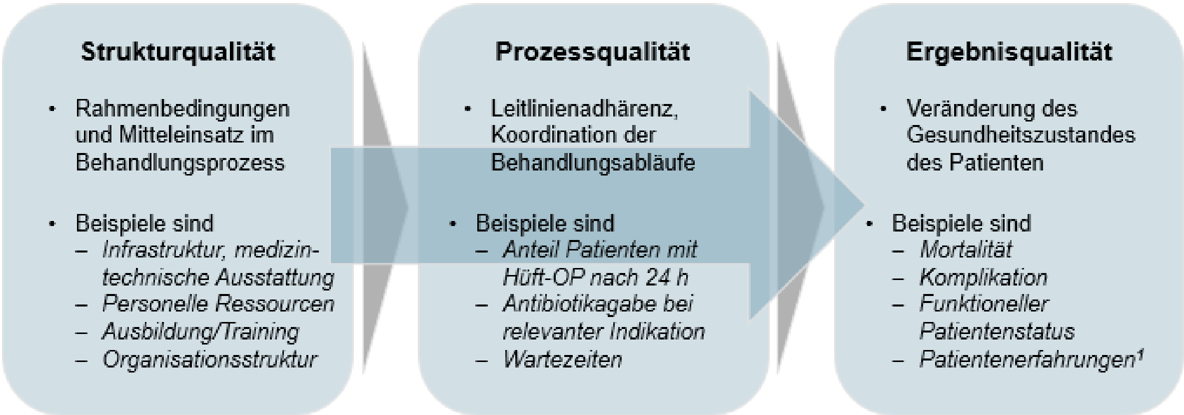
sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen". Das IOM beschreibt Qualität von Gesundheitsleistungen als das „Maß in dem Gesundheitsleistungen für einzelne Personen und ganze Bevölkerungen die Wahrscheinlichkeit von erwünschten Ergebnissen erhöhen und dem aktuellen professionellen Wissensstand entsprechen" [87]. Dazu definiert das IOM sechs Dimensionen, die Qualität auszeichnen: 1. Patientensicherheit, 2. Behandlungseffektivität, 3. Patienten-Zentriertheit, 4. Versorgung entsprechend dem aktuellen Wissensstand, 5. Effizienz und 6. gerechter Zugang zu Gesundheitsleistungen [88].

Die im Sozialgesetzbuch (SGB) V festgehaltenen Qualitätsansprüche berücksichtigend, definiert Ruprecht (1993) Qualität im Gesundheitswesen als „eine ausreichende und zweckmäßige, d.h. patienten- und bedarfsgerechte, an der Lebensqualität orientierte, fachlich qualifizierte, aber auch wirtschaftliche medizinische Versorgung mit dem Ziel, die Wahrscheinlichkeit erwünschter Behandlungsergebnisse bei Individuen und in der Gesamtbevölkerung zu erhöhen" [89]. In der Lesehilfe für die Qualitätsberichte der Krankenhäuser schreibt der Gemeinsame Bundesausschuss (G-BA) deutlich offener „Unter Qualität kann man die Eignung einer Sache zu einem vorgegeben Zweck verstehen" [90]. Sowohl die Eigenschaft der Sache als auch der Zweck ist anwendungsspezifisch zu definieren und zu evaluieren. Weiterhin verwendet der G-BA zur Qualitätssicherung in Deutschland das Qualitätsmodell von Alexis Donabedian [90], welches Qualität in der Medizin grundsätzlich in die drei Elemente Struktur-, Prozess- und Ergebnisqualität unterteilt [91].

Die vorliegende Arbeit beschäftigt sich primär mit der Qualität der Leistungserbringung im stationären Sektor. Sie schließt dabei an die Praxis des G-BA an und bedient sich der Donabedianischen Unterscheidung zwischen Struktur-, Prozess- und Ergebnisqualität und dem Zusammenhang dieser drei Elemente [91, 92] (vgl. Abbildung 2). Strukturqualität bezeichnet das Vorhandensein von relevanten Versorgungsstrukturen wie die Anzahl der Ärzte und

Krankenschwestern, die Verfügbarkeit von medizintechnischen Instrumenten und die Organisationsstruktur. Prozessqualität beschreibt die Leitlinienadhärenz der Versorgungsabläufe und beinhaltet z.B. den Anteil der Patienten, die entsprechend der Indikation und Leitlinie mit Antibiotika behandelt werden oder den Anteil der Patienten, bei denen bei entsprechender Indikation innerhalb von 48 Stunden eine Hüftoperation initiiert wurde. Zusätzlich umfasst die Prozessqualität auch die Koordination der medizinischen Versorgung innerhalb und über Abteilungen hinweg und den korrekten Umgang mit medizintechnischen Geräten. Ergebnisqualität beschreibt die Auswirkungen bzw. den Erfolg der medizinischen Versorgung. Als Ergebnisindikatoren gelten Sterblichkeits-, Komplikations- und Infektionsraten sowie Patienteneinschätzungen (engl. patient-reported outcome measures, PROMs) wie Schmerzempfinden oder funktionale Verbesserung nach Knieoperationen (z.B. Oxford-Knee-Scores oder Euro-Quol 5D).

**Abbildung 2**: Struktur-, Prozess- und Ergebnisqualität im Gesundheitswesen




**Legende**: Struktur- und Prozessqualität beeinflussen die letztendlich im Behandlungsprozess erreichte Ergebnisqualität., eigene Darstellung nach Donabedian (1966 und 1992)

Donabedian legt seinem Modell eine gegenseitige Beeinflussung der drei Elemente zugrunde Doch sind letztlich gute Struktur- und Prozessqualität notwendige, aber keine hinreichenden Bedingungen für gute Ergebnisqualität. Neben Patienten-rapportierten Ergebnisindikatoren werden oft auch Patientenerfahrungen als Qualitätsindikatoren herangezogen, doch ist die medizinische Aussagekraft und Relevanz dieser umstritten, da Patientenerfahrungen subjektiv und durch andere Faktoren als lediglich die Behandlung beeinflussbar sind [93]. Als Achillesferse der Ergebnisindikatoren gilt die Risikoadjustierung. Durch diese wird die Ergebnisqualität verschiedener Krankenhäuser mit Patientenpopulationen mit unterschiedlicher Altersstruktur, Begleiterkrankungen und anderen Risikofaktoren vergleichbar.

Die vorliegende Arbeit legt den Fokus auf diese Ergebnisindikatoren als Zielvariablen. Die wissenschaftliche Evidenz für den kausalen Zusammenhang zwischen Prozess – und Ergebnisindikatoren ist limitiert [15, 94, 95]. Die medizinische Versorgung kann in fast allen Prozessvariablen optimal sein, doch kann ein einzelner Fehler zu desaströsen Behandlungsergebnissen für Patienten und zusätzlichen unnötigen Kosten für das Gesundheitssystem führen. Da Behandlungsergebnisse einen konkreten Endpunkt darstellen, haben sie auch den höchsten Nutzen für Patienten. Letztlich sind Ergebnisse die ultimative Bestätigung von Wirkungsgrad und Qualität einer Behandlung [96].

### *Einflussfaktoren auf die Ergebnisqualität in der Gesundheitsversorgung*

Der Gesundheitsmarkt ist im Vergleich zu Märkten für andere Produkte und Dienstleistungen ist sowohl auf der Nachfrageseite als auch auf der Angebotsseite mit zahlreichen Besonderheiten versehenen. Diese verstärken Informationsasymmetrien im Prinzipal-Agenten-Modell zwischen Patienten und Leitungserbringern und verhindern ein Marktgleichgewicht in Bezug auf das wichtige Produktkriterium Ergebnisqualität. Staatliche Maßnahmen zur Qualitätssicherung adressieren dieses Ungleichgewichte im Prinzipal-Agenten-Modell und daraus resultierenden Marktversagen. Dabei soll Qualitätstransparenz die im verborgenen liegende Qualität der Leistungserbringung für alle Marktteilnehmer sichtbar machen.

#### Qualitätsregulativ auf Nachfrageseite limitiert

Entsprechend der gewöhnlichen volkswirtschaftlichen Rolle des Konsumenten sollten Patienten im Gesundheitssystem in der Lage sein, durch ihre Nachfragepräferenzen das angebotene Maß an Qualität mitzubestimmen. In der Gesundheitsversorgung und insbesondere im Akutbereich ist diese Qualitätsregulierung durch die Nachfrageseite jedoch stark limitiert. Ein Hauptproblem liegt darin, dass die Beurteilung medizinischer Qualität ein hohes Maß an medizinischem Fachwissen erfordert [97]. Dieses Fachwissen ist bei Patienten in der Regel nicht vorhanden, was zu starken Informationsasymmetrien zwischen Patient (Prinzipal) und Leistungserbringer (Arzt) führt [98]. Diese Informationsasymmetrien wiederrum führen zu den in der Gesundheitsversorgung besonders relevanten und in der Prinzipal-Agenten-Theorie als ‚Verborgene Eigenschaften', ‚Verborgenes Handel' und ‚verborgende Information' bekannten Problemen [97, 100]. ‚Verborgene Eigenschaften' beschreiben das Problem, dass zu Beginn der Behandlung der Leistungserbringer und dessen Eigenschaften dem Patienten in der Regel relativ unbekannt sind. ‚Verborgendes Handeln' und ‚verborgende Informationen' führen dazu, dass Patienten während und nach der Behandlung die Behandlung selbst als auch das Ergebnis nur bedingt einschätzen können. Sie können nicht unterscheiden, ob die Behandlung selbst oder äußere Umstände zu dem Ergebnis geführt haben und ob das Ergebnis unter den gegebenen Umständen das maximal erreichbare positive Ergebnis ist.

Die besondere Relevanz dieser Thematik in der Gesundheitsversorgung kann an einem beispielhaften Vergleich mit anderen Gütern – einem T-Shirt und einem Auto – verdeutlicht werden [99]. Konsumenten können mit großer Sicherheit die Qualität eines T-Shirts einschätzen. Falls die Qualität im Markt nicht ausreicht, werden neue Produzenten in den Markt eintreten und T-Shirts mit höherer Qualität anbieten. Es findet eine Auffächerung des Angebots mit entsprechend höherem Preis für bessere Qualität statt. T-Shirts gelten als Suchgüter (engl. search goods).

Beim Autokauf können die meisten Konsumenten relativ klar sagen, was für ein Auto sie suchen. Dennoch können sie sich erst durch Testfahrten und die häufige Nutzung des Autos von der Qualität des Produktes überzeugen. Autos gelten daher als Erfahrungsgüter (engl. experience goods). Die Verfügbarkeit von privaten und staatlichen Verbrauchs-, Sicherheitsund Fahrtests und Zertifizierungen wandelt aber auch Autos zunehmend zu Suchgütern[3] und fördert so ein Marktgleichgewicht zwischen nachgefragter und angebotener Qualität.

In Bezug auf Gesundheitsdienstleistungen können Konsumenten aufgrund des fehlenden medizinischen Fachwissens oft weder vor noch nach der Behandlung deren Qualität richtig einschätzen. Die Qualitätskontrollfunktion der Nachfrage wird bei Gesundleistungen weiter erschwert, da diese Nachfrage auf Ebene des individuellen Konsumenten irregulär, meist unvorhersehbar und mit einem hohen Maß an Unsicherheit verbunden ist [97] – ex ante Information des Konsumenten findet kaum statt. Medizinische Behandlungen gelten daher als Vertrauensgüter (engl. credence goods). Bei Vertrauensgütern wiegt das Problem der Informationsasymmetrien zwischen den Patienten und dem ärztlichen Fachpersonal besonders schwer, welches zu volkswirtschaftlichem Marktversagen führen kann [97, 101].

Im deutschen Gesundheitssystem als auch in anderen europäischen Ländern ergeben sich weitere, spezifische Probleme für die qualitätssichernde Wirkung der Nachfrage: Patienten sind durch das DRG Fallpauschalensystem, welches keine Korrelation zwischen Preis und Qualität zulässt, und dem GKV System, als Mittler zwischen konsumierenden Patienten und produzierenden Leistungserbringern, vom Preissignal und der damit einhergehenden automatischen Produktevaluation abgeschnitten. Patienten können den Preis nicht als Zeichen für Qualität verwenden und ihre unterschiedliche Zahlungsbereitschaft für das Gut Qualität nicht zum Ausdruck bringen.

Obwohl sich GKV Patienten rechtlich deutschlandweit behandeln lassen können, ist die Gesundheitsversorgung aufgrund der mit Distanz steigenden Transportkosten lokal begrenzt [68, 102, 103]. Auch wenn die Reisebereitschaft je nach Patient und Erkrankung variiert, führen Transportkosten und das Bedürfnis in Erreichbarkeit der Familie behandelt zu werden zu lokal

[3] Bestimmte Autosegmente wie Sportwagen sind aber weiterhin klare Erfahrungsgüter.

begrenzten Gesundheitsmärkten und damit auch zu einer Begrenzung eines möglichen Qualitätswettbewerbs zwischen Leistungserbringern [104–106].

Ökonomische Qualitätsanreize auf der Angebotsseite limitiert

Wie auf der Nachfrageseite prägen auch auf der Angebotsseite mehrere entscheidende Aspekte die Ausprägung von Qualität im deutschen Gesundheitswesen. Die freie Arzt- und Krankenhauswahl ermöglicht im Grundsatz allen Krankenhäusern, durch ein mehr an Qualität zusätzliche Patienten anzulocken und das finanzielle Ergebnis des Krankenhauses, so lange der marginale Ertrag pro zusätzlichem Patient positiv ist, als wirtschaftende Einheit zu steigern. Doch sorgen Kollektivverträge gleichzeitig dafür, dass die Krankenkassen jede im Rahmen des Leistungskataloges erbrachte Leistung vergüten müssen, unabhängig von deren Qualität [107].

Informationsasymmetrien und der Einfluss von statistischem Zufall führen zur Differenzierung zwischen Aspekten, die in Verträgen festgehalten (engl. contractible) und die nicht festgehalten werden (engl. non-contractible) können [108]. Bei einem Krankenhausaufenthalt gehören zum Beispiel die Verpflegung, die Größe des Bettes, die Präsenz und der Ausbildungsgrad von medizinischem Fachpersonal zu erster Kategorie. Fleiß und Sorgfalt, Fähigkeit des Arztes und die Häufigkeit des Händewaschens gehören dagegen zu zweiter Kategorie. Da sowohl vertraglich festlegbare als auch nicht-vertragliche greifbare Qualität mit Ressourceneinsatz verbunden sind und Patienten als Konsumenten und Krankenversicherer als Kostenerstatter nicht kontrahierbare Leistungen nur schwer bis gar nicht einschätzen können, kann es bei den wichtigen nicht-vertragstauglichen Aspekten von Qualität zu einer Qualitätsunterversorgung kommen [64, 99, 108].

Dagegen spricht die intrinsische Motivation von Ärzten und medizinischem Personal, welche in der ökonomischen Literatur als Altruismus zusammengefasst wird. Wenn Leistungserbringer altruistisch agieren, haben sie eine intrinsische Motivation die Qualität möglichst hoch zu halten [109]. Auf Ebene einzelner Ärzte oder Krankenschwestern mag diese intrinsische Motivation zwar noch ausgeprägt sein, aber auf der aggregierten Krankenhausebene, wo Sachzwänge wie wirtschaftliche Ziele und Personalnot vorherrschen, ist die Steigerung von Qualität aufgrund von Altruismus fraglich [110–112].

Auch langfristig wirkt sich fehlende Transparenz auf der Angebotsseite negativ auf die Qualität aus: Wenn gute Behandlungsqualität nicht erkennbar bzw. kontrahierbar ist und Qualitätsunterschiede im Verborgenen bleiben, kann es zu einem Investitionsmangel kommen, da privatwirtschaftlichen Akteuren ohne zusätzliche Incentivierung der Anreiz fehlt, in qualitätsfördernde, kostenintensive Infrastruktur, medizinisch-technische Ausstattung oder Personalstrukturen zu investieren. So leiden zum Beispiel viele Krankenhäuser an einem Investitionsmangel im Bereich der elektronischen Patientenakte, obwohl diese Technologie Qualität durch standardisierte Speicherung und Weitergabe von Patienteninformationen und den Einsatz als Arzt-Entscheidungshilfe-System fördert [113, 114]. Sobald Qualität aber für Patienten und Krankenkassen, welche zunehmend die Patientensteuerung übernehmen, sichtbar wird, haben Leistungserbringer einen Anreiz in Qualität zu investieren, um im Qualitätswettbewerb erfolgreich zu sein.

Der entstehende Qualitätswettbewerb in Deutschland wird aktuell primär durch private Träger wie die HELIOS Kliniken vorangetrieben [115]. Zu diesem Zweck hat Helios die Initiative Qualitätsmedizin federführend gegründet.[4] Da viele Wettbewerbsparameter wie Standort, Preis, Personal und Ausstattung im deutschen Krankenhausmarkt stark reguliert sind, wird Qualität von einigen Krankenhäusern zusehends als Wettbewerbsvariable anerkannt. Der mögliche Qualitätswettbewerb ist natürlich abhängig von der Anzahl der Krankenhäuser und deren Leistungsbereiche in einem lokalen Markt [62, 102]. Damit Qualitätswettbewerb

[4] Gleichzeitig haben sich die Helios Kliniken im Rahmen der Dachorganisation Stiftung Initiative Qualitätskliniken klar gegen Referenzportale positioniert.

funktioniert, bedarf es einer umfänglichen Transparenz über die Qualität der Leistungserbringung für alle Marktteilnehmer.

Gute Leistungserbringer mit hoher Behandlungsqualität haben einen inhärenten Anreiz, ihre Qualität sichtbar und vergleichbar zu machen, um Reputation und zusätzliche Patienten durch ihre bessere Qualität zu gewinnen [57]. Dementsprechend können auch auf der Angebotsseite Drittanbieter helfen, medizinische Behandlungsqualität sichtbar zu machen. Dies geschieht primär über Zertifikate, die an Krankenhäuser vergeben werden, wenn diese bestimmte Struktur- und Prozessvoraussetzungen erfüllen. In Deutschland sind es primär die Qualitätsmanagement Zertifikate der Kooperation für Qualität und Transparenz im Gesundheitsweisen (KTQ) und die ISO 9001 Zertifikate [116]. Gleichzeitig gibt es auch auf Fachabteilungsebene weitere Zertifikate, die eine Spezialisierung und damit höhere Qualität aufzeigen sollen. Dazu gehören das Stroke Unit Zertifikat der Deutschen Schlaganfallhilfe [117] und die Zentren-Zertifikate der Deutschen Krebsgesellschaft [118]. Allen Zertifikaten ist gemein, dass in Peer Reviews die Behandlungsvoraussetzungen und das Qualitätsmanagement eines Krankenhauses bzw. einer Fachabteilung durch `Peers´ (d.h. Fachärzte aus dem gleichen Leistungsbereich aus anderen Krankenhäusern) evaluiert und bestätigt werden [119].

Aufgrund der hohen Bedeutung von Informationen zur Qualität von Gesundheitsdienstleistungen wird Qualitätstransparenz zu einem öffentlichen Gut [99], das entweder von Leistungserbringern selbst, Dritten privatwirtschaftlich oder vom Staat durch Regulierung oder eigenes Angebot bereitgestellt wird. Zwar gibt es z.B. in den USA privatwirtschaftliche Lösungen wie healthgrades.com und consumerreports.org; doch stellen die Komplexität der Daten [99], häufig fehlende strukturierte Informationsstandards für den Informationsaustausch [15] und die Interessenkonflikte der verschiedenen Stakeholder Hürden für eine privatwirtschaftliche Lösung dar. In Deutschland wirken zusätzlich der hohe Regulierungsgrad, die langsamen Prozesse in der Selbstverwaltung, die hohen Ansprüche an

Datenschutz und das stärkere Vertrauen in und die Erwartungen an den Staat gegen stärkere privatwirtschaftliche Lösungen zur Qualitätstransparenz und eine Reduktion der Informationsasymmetrien auf der Nachfrageseite [120, 121]

## Qualitätssicherung in Deutschland

In Deutschland sind die Begriffe der Behandlungsqualität und Qualitätstransparenz eng mit dem Konzept der Qualitätssicherung verknüpft. Erste Qualitätssicherungsansätze in Deutschland waren die Münchener Perinatalstudie (1975-1977) und einzelne Register auf Länderebene Ende der 1970er Jahre [4]. Seit Anfang der 1990er Jahre sind Krankenhäuser gesetzlich verpflichtet, in einer minimalen und nicht-standardisierten Form Qualitätsmanagement durchzuführen. Mitte der 1990er Jahre haben sie erste Qualitätsindikatoren zur Qualitätsmessung eingeführt [4]. Im Zuge der Einführung des DRG Systems wurde die Qualitätssicherung durch die GKV-Gesundheitsreform im Jahr 2000 institutionalisiert und ausgebaut, um der Gefahr einer Verschlechterung der Patientenversorgung durch Kostendruck und Volumenanreize im DRG Fallpauschalensystem entgegenzuwirken [86, 122] und Mindestanforderungen an die Qualität sicherzustellen.

Im Zuge des in 2016 verabschiedeten KHSG wurden die Anforderungen der im Aufgabenbereich des G-BA liegenden externen Qualitätssicherung deutlich erweitert. Das fünfte Sozialgesetzbuch (SGB V, Paragraphen 135 – 139d) reguliert verpflichtende Anforderungen im Rahmen der externen Qualitätssicherung an Krankenhäuser, die nach §108 SGB V für die Behandlung von GKV Patienten zugelassenen sind. Dazu gehören insbesondere folgende in §136b zusammengefasste Anforderungen:

1. Erfüllungsnachweise zu den Fortbildungspflichten des medizinischen Fachpersonals;
2. Mindestmengen für Leistungsbereiche, in denen Behandlungsergebnisse abhängig von der Menge der erbrachten Leistungen sind;

3. jährliche, strukturierte und vergleichbare Qualitätsberichte;

4. vier auszuwählende Leistungsbereiche, in denen qualitätsbasierte Selektivverträge erprobt werden und

5. eine qualitätsorientierte Vergütung für geeignete Leistungsbereiche.

Für die Umsetzung und Weiterentwicklung der externen Qualitätssicherung hat der G-BA 2015 das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) gegründet. Das IQTIG soll die staatlichen Maßnahmen zur Qualitätssicherung konsolidieren und weiterentwickeln. Die Qualitätssicherung soll zukünftig sektorenübergreifend stattfinden, Patientenbefragungen einbeziehen und möglichst sparsam bezüglich des zusätzlichen Dokumentationsaufwandes erfolgen. Die zu verwendenden Qualitätsindikatoren sollen justiziable sein, um darauf aufbauend qualitätsorientierte Vergütungssysteme zu entwickeln und Qualität noch stärker in die auf Länderebene stattfindende Krankenhausplanung zu integrieren. Das IQTIG soll außerdem Qualitätsergebnisse allgemeinverständlich und zwischen Krankenhäusern vergleichend veröffentlichen. Letztlich soll das IQTIG Kriterien für die Bewertung von Zertifikaten und Qualitätssiegeln entwickeln und auf Basis dieser Kriterien über die Aussagekraft der Zertifikate urteilen (SGB V § 137a). Der Gesetzgeber und der G-BA haben damit dem IQTIG einen anspruchsvollen Aufgabenkatalog definiert. Dieser kann vom IQTIG nur über mehrere Jahre und unter Einbezug des nationalen und internationalen Forschungsstandes umgesetzt werden. Der erste Bericht zu planungsrelevanten Qualitätsindikatoren und die starke Kritik daran verdeutlichen die Komplexität des Unterfangens [123, 124].

Zusätzlich zur gesetzlichen Qualitätssicherung gibt es noch weitere privatwirtschaftliche Ansätze. So errechnet die AOK QSR Ergebnisindikatoren auf Basis von Patientendaten für 16 Leistungsbereiche [125, 126]. Des Weiteren messen die von privaten Leistungserbringern gegründeten Organisationen Initiative für Qualitätsmedizin (IQM) und Qualitätskliniken.de

Prozess- und Ergebnisindikatoren, die auf Basis von Routinedaten berechnet werden [127, 128]. Neben der Sicherstellung von minimalen Qualitätsstandards verfolgen alle Initiativen zusätzlich auch einen Public Reporting Ansatz.

## Public Reporting zur Steigerung der Transparenz von Ergebnisqualität

Public Reporting beschreibt die Veröffentlichung von Informationen zur Behandlungsqualität – wie Indikatoren, Berichte und Bewertungen – einzelner Leistungserbringer. Diese Daten sollen Patienten und Einweiser ebenso informieren wie Leistungserbringer. Letztere sollen in die Lage versetzt und incentiviert werden, ihre eigene Behandlungsqualität zu verbessern für ein bestehendes Patientenaufkommen und um zusätzliche Patienten zu gewinnen [37, 58].

Der Wirkungsmechanismus zwischen Ergebnistransparenz und Qualitätsverbesserung wird in der Literatur durch drei Wirkungspfade beschrieben:

Die *ex-ante Verbesserung* tritt bereits durch die Erwartung der Veröffentlichung ein. Diese Erwartung bringt Leistungserbringer dazu, die Behandlungsqualität durch Kontrolle und stärkere Befolgung der etablierten Behandlungsprotokolle, zusätzliche Investitionen in Prozesse und medizinische Ausstattung sowie durch qualitätsfokussierte Anreizstrukturen zu steigern [129, 130].

In der *ex-post Systemverbesserung* evaluieren die Mediziner und das Krankenhausmanagement die Leistung der eigenen Fachabteilungen anhand veröffentlichter Ergebnisse und auf Basis der Vergleichsmöglichkeiten. Sie versuchen die Behandlungsergebnisse durch Adressierung der identifizierten Schwachstellen zu verbessern. Zusätzlich gibt es die Möglichkeit, die Best Practices der Qualitätsführer in die eigenen Behandlungsabläufe zu integrieren. Oftmals enthalten die Transparenzinitiativen auch Mechanismen, um nach einer Identifizierung der Schwachstellen die unzureichende Qualität durch klinikübergreifende Evaluations- und Verbesserungsinitiativen zu steigern (z.B. Strukturierte Dialoge im Rahmen der externen stationären Qualitätssicherung des G-BA und IQM „Peer Reviews“) [131–133].

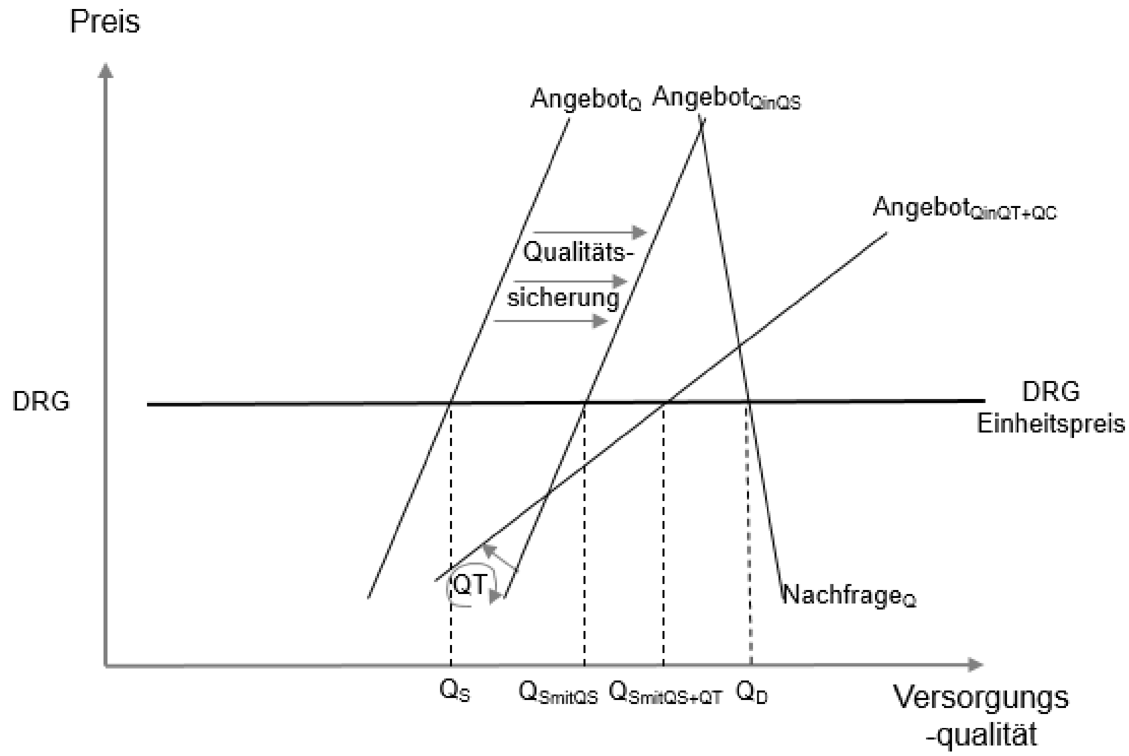
In der *Angebotsselektion* wählen Patienten und Einweiser nach Analyse und Abwägung öffentlich verfügbarer Qualitätsinformationen von Krankenhäusern, Leistungsbereichen und anderen Entscheidungskriterien (z. B. Distanz) das Krankenhaus mit den vergleichsweise besseren Behandlungsergebnissen (sog. „voting with feet"). Über die Zeit wird Qualitätsführerschaft durch einen Zuwachs an Patienten und einen höheren Marktanteil belohnt. Qualitätsrückstände werden durch entsprechende Rückgänge bestraft. Dies setzt allerdings voraus, dass Patienten direkten und einfachen Zugang zu veröffentlichten Qualitätsinformationen haben, diese verstehen und einordnen können sowie der neue Informationsgehalt von den bereits bestehenden Auffassungen zur Qualität der Krankenhäuser abweicht und dadurch zu einer Verhaltensänderung führen kann [28, 36, 37, 56, 134, 135].

Neben dem positiven Potential zur Steigerung des Qualitätswettbewerbs zwischen Leistungserbringern gibt es auch mögliche negative Nebeneffekte von Qualitätstransparenz. Als wichtigste negative Folge ist die Adverse Selektion durch Ärzte und Krankenhäuser anzuführen: Um möglichst gute Qualitätsergebnisse zu präsentieren, können Leistungserbringer Hochrisikopatienten abweisen [28]. Dieses kann zu negativen Wohlfahrtseffekten und insgesamt schlechterer Behandlungsqualität führen. Ferner können Leistungserbringer zusätzliche Begleiterkrankungen kodieren, welche ihre Behandlung relativ zu anderen Leistungserbringern besser erscheinen lässt (sog. Upcoding) [136]. Ferner sind die hohen Kosten zu nennen, die mit der Messung von klinischen Daten im Zuge der externen Qualitätssicherung für behandelnde Ärzte und das Krankenhaus entstehen. Qualitätssicherung auf Basis von Routinedaten kann diese Kosten reduzieren [29, 137]. Des Weiteren ist es aufgrund statistischer Zufälle und der für valide Aussagen notwendigen höheren Fallzahlen oft schwer, juristisch rechtssichere Indikatoren zu entwickeln ohne dabei zusätzlich auch auf Prozess- oder Strukturfaktoren zurückzugreifen [123].

Zusammenfassend verdeutlicht Abbildung 3 schematisch das Marktversagen in der Bereitstellung von Qualität im Gesundheitswesen und das Wirken von Qualitätssicherung und Qualitätstransparenz. Im DRG Einheitspreisumfeld und ohne Qualitätssicherung und Qualitätstransparenz liegt das von den Leistungserbringern angebotene Qualitätsniveau bei $Q_S$ und damit weit unter dem von Patienten nachgefragten Level $Q_D$. Die Qualitätsnachfrage ist relativ unelastisch, da Konsumenten Qualität nur sehr limitiert einschätzen können und in Bezug auf Gesundheit grundsätzlich ein hohes Maß an Qualität fordern. Gleichzeitig ist die Nachfrage nicht gänzlich unelastisch, da Konsumenten immer auch eine Abwägung treffen zwischen ihrer Nachfrage nach Gesundheit und den damit verbundenen Opportunitätskosten Transport, Suchkosten (inklusive Einschätzung des Qualitätsniveaus) und dem finanziellen Aufwand, auch wenn Patienten die direkten Kosten nicht selbst tragen. Zum Beispiel variiert zwischen Patienten die Bereitschaft weite Fahrtwege für ein bestimmtes Level an Behandlungsqualität auf sich zunehmen. Patienten investieren auch unterschiedlich viel Zeit in die Suche nach Qualitätsinformationen über Leistungserbringer [106, 122, 138].

Neben der Nachfragekurve ist auch die Qualitätsangebotskurve relativ unelastisch, da Anreize für zusätzliche Qualitätsanstrengungen auf Seiten der Leistungserbringer nur sehr limitiert vorhanden sind. Qualitätssicherung kann die Qualitätsanforderungen an Leistungserbringer erhöhen und die Qualitätsangebotskurve nach außen verschieben, so dass bei jedem Preislevel, inklusive des DRG Einheitspreises, ein höheres Level an Qualität ($Q_{SinQT}$) angeboten wird. Zusätzlich zu Minimalanforderungen durch eine Qualitätssicherung kann eine privat oder staatlich organisierte Qualitätstransparenz die Qualitätsangebotskurve elastischer machen und zu einem Qualitätsniveau über den Minimalanforderungen führen ($Q_{SinQC+QT}$). Für Leistungserbringer lohnt es sich bessere Qualität bereitzustellen, weil die Qualität im Zuge der Qualitätstransparenz für Patienten und Krankenkassen erkennbar wird und stärker in Wahl bzw. Kaufentscheidung einfließen kann. Abbildung 3 lässt auch eine kurze Veranschaulichung der Qualitätsorientierten Vergütung (d.h. Pay-for-Performance) und des Qualitäts-Kosten-Trade-offs zu. Bei ersterem wird ein höherer Preis für bessere Qualität gezahlt, was den DRG Einheitspreis um einen entsprechenden Qualitätszuschlag nach oben schieben würde. Bei letzterem könnte bessere Qualität durch eine kostensenkende Wirkung auch einen negativen Einfluss auf den DRG Einheitspreis (d.h. Kostendurchschnitt einer Gruppe von mehr als 250 DRG Referenzkrankenhäusern) bei geringer und mittlerer Qualität haben. Bei sehr hoher Qualität lässt sich wiederum ein positiver Einfluss auf den DRG Einheitspreis durch höhere Kosten bei der Bereitstellung von Spitzenqualität vermuten. Dies würde zu einer U-Kurve für den DRG Einheitspreis führen.

**Abbildung 3**: Schematische Darstellung Qualitätsnachfrage und -angebot ohne und mit Qualitätssicherung und Qualitätstransparenz




*Legende: Darstellung des Marktungleichgewichtes in Nachfrage nach und Angebot von Qualität in der Krankenhausbehandlung. Preis beinhaltet nicht nur direkte Kosten, sondern auch indirekte Kosten (z.B. Suchkosten), eigene Darstellung*

## References


1. Kohn L, Corrigan J, Donaldson M. To Err is Human: Building a Safer Health System. Washington D.C.: National Academic Press; 2000.
2. James JT. A new, evidence-based estimate of patient harms associated with hospital care. J Patient Saf. 2013;9:122–8. doi:10.1097/PTS.0b013e3182948a69.
3. OECD. Health Care Quality Indicators - Acute Care. 2015. http://www.oecd.org/els/health-systems/hcqi-acute-care.htm. Accessed 31 Oct 2015.
4. Busse R, Blümel M, Ognyanova D. Das deutsche Gesundheitssystem: Akteure, Daten, Analysen. Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft; 2013.
5. Schrappe M, Lessing C, Jonitz G, Conen D, Gerlach F, Hart D, et al. Agenda Patientensicherheit 2007. 2007. http://www.aps-ev.de/fileadmin/fuerRedakteur/PDFs/Agenda_Patientensicherheit/Agenda_2007_mit_Titelblatt.pdf. Accessed 12 Sep 2015.
6. Klauber J, Geraedts M, Friedrich J, Wasem J. Krankenhaus-Report 2014: Schwerpunkt Patientensicherheit. 1st ed. Stuttgart: Schattauer GmbH; 2014.
7. Destatis. Krankenhausstatistik. 2016. http://www.gbe-bund.de/oowa921-install/servlet/oowa/aw92/WS0100/_XWD_FORMPROC?TARGET=&PAGE=_XWD_2&OPINDEX=2&HANDLER=XS_ROTATE_ADVANCED&DATACUBE=_XWD_30&D.000=DOWN&D.001=PAGE&D.922=PAGE&D.100=ACROSS. Accessed 26 Dec 2016.
8. Vries EN de, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care. 2008;17:216–23. doi:10.1136/qshc.2007.023622.
9. Deutscher Bundestag. Gesetz zur Reform der Strukturen der Krankenhausversorgung: Krankenhausstrukturgesetz -KHSG; 2015.
10. Committee on Accounting for Socioeconomic Status in Medicare Payment Programs, Board on Population Health and Public Health Practice, Board on Health Care Services, Institute of Medicine, National Academies of Sciences, Engineering, and Medicine. Accounting for Social Risk Factors in Medicare Payment: Identifying Social Risk Factors. Washington D.C.: National Academic Press; 2016.
11. Elbuluk A, Bosco JA. Private payer bundled payment arrangements. Seminars in Arthroplasty. 2016;27:201–6. doi:10.1053/j.sart.2016.10.011.
12. Dummit LA, Kahvecioglu D, Marrufo G, Rajkumar R, Marshall J, Tan E, et al. Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes. JAMA. 2016;316:1267–78. doi:10.1001/jama.2016.12717.
13. Oldenburg C, Van den Berg J, Leistikow I. All aboard, getting nationwide indicators on the rails: Collaborative governance as a strategy for developing effective national quality indicators for hospital care. British Medical Journal Outcomes. 2015.
14. Morgan D. Focus on Health Spending: OECD Health Statistics 2015; 2015.
15. Porter ME, Teisberg EO. Redefining health care: Creating value-based competition on results. Boston, Mass.: Harvard Business School Press; 2006.
16. Pauly MV. Analysis & commentary: The trade-off among quality, quantity, and cost: how to make it--if we must. Health Aff (Millwood). 2011;30:574–80. doi:10.1377/hlthaff.2011.0081.
17. OECD. Health at a Glance 2015: OECD Indicators: OECD Publishing; 2015.

18. Chung S-C, Sundström J, Gale CP, James S, Deanfield J, Wallentin L, et al. Comparison of hospital variation in acute myocardial infarction care and outcome between Sweden and United Kingdom: population based cohort study using nationwide clinical registries. BMJ. 2015;351:h3913. doi:10.1136/bmj.h3913.
19. Fonarow GC, Smith EE, Reeves MJ, Pan W, Olson D, Hernandez AF, et al. Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. Stroke. 2011;42:159–66. doi:10.1161/STROKEAHA.110.601831.
20. Wouters MW, Siesling S, Jansen-Landheer ML, Elferink, M A G, Belderbos J, Coebergh JW, Schramel, F M N H. Variation in treatment and outcome in patients with non-small cell lung cancer by region, hospital type and volume in the Netherlands. European Journal of Surgical Oncology. 2010;36 Suppl 1:S83-92. doi:10.1016/j.ejso.2010.06.020.
21. GKV Spitzenverband. Qualitätsorientierte Versorgungssteuerung und Vergütung. 2014. https://www.gkv-spitzenverband.de/krankenversicherung/qualitaetssicherung_2/qualitaetssicherung_2.jsp . Accessed 23 Dec 2016.
22. Porter ME. What Is Value in Health Care? The New England Journal of Medicine. 2010;26:2477–81.
23. Klauber J, Bevan G, editors. Schwerpunkt: Qualitätstransparenz - Instrumente und Konsequenzen. 1st ed. Stuttgart [u.a.]: Schattauer; 2005.
24. Yong PL, Saunders RS, Olsen L, editors. The healthcare imperative: Lowering costs and improving outcomes : workshop series summary. Washington, D.C.: National Academies Press; 2010.
25. Fung CH, Lim Y-W, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med. 2008;148:111–23.
26. Totten AM, Wagner J, Tiwari A, O'Haire C, Griffin J, Walker M. Closing the quality gap: revisiting the state of the science (vol. 5: public reporting as a quality improvement strategy). Evid Rep Technol Assess (Full Rep). 2012:1–645.
27. Dranove D, Kessler D, McClellan M, Satterthwaite M. Is More Information Better?: The Effects of “Report Cards” on Health Care Providers. Journal of Political Economy. 2003;111:555–88. doi:10.1086/374180.
28. Dafny L, Dranove D. Do report cards tell consumers anything they don’t already know? The case of Medicare HMOs. The RAND Journal of Economics. 2008;39:790–821.
29. DKG. Position der Deutschen Krankenhausgesellschaft zur Weiterentwicklung der Qualitätssicherung und der Patientensicherheit. Das Krankenhaus. 2014;5.
30. Costa S-D. Qualitätsmanagement im Krankenhaus - Nicht zum Nutzen der Patienten: Qualitätsmanagement ist für die Medizin ähnlich nützlich wie Ornithologie für die Vögel. Deutsches Ärzteblatt. 2014;111:1344–5.
31. Kraska RA, Krummenauer F, Geraedts M. Impact of public reporting on the quality of hospital care in Germany: A controlled before-after analysis based on secondary data. Health Policy. 2016;120:770–9. doi:10.1016/j.healthpol.2016.04.020.
32. Held J, Lack N, Hermanek P, Eßer M. Trendanalyse der Qualitätsindikatoren der Krankenhäuser einer Region von 2006/07 bis 2013. Gesundheitswesen 2015. doi:10.1055/s-0041-110674.
33. Cacace M, Ettelt S, Brereton L, Pedersen J, Nolte E. How health systems make available information on service providers: Experience in seven countries. Cambridge: Rand Corporation; 2011.

34. Cordasev H, Björnberg A, Hjertqvist O. Cross border care EU: How to choose the best hospital?: - A study of hospital information portals in five EU countries. Health Consumer Powerhouse; 2010.
35. Milstein R, Schreyoegg J. Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. Health Policy. 2016;120:1125–40. doi:10.1016/j.healthpol.2016.08.009.
36. Boyce T, Dixon A, Fasolo B, Reutskaja E. CHOOSING A HIGH-QUALITY HOSPITAL. 2010.
37. Kumpunen S, Trigg L, Rodrigues R. Public reporting in health and long-term care to facilitate provider choice; 2014.
38. Pross C, Geissler A, Busse R. Measuring, Reporting, and Rewarding Quality of Care in 5 Nations: 5 Policy Levers to Enhance Hospital Quality Accountability. Milbank Q. 2017;95:136–83. doi:10.1111/1468-0009.12248.
39. AQUA. Qualitätsreport 2013. Göttingen: AQUA; 2014.
40. AQUA, editor. Qualitätsreport 2012. Göttingen: AQUA; 2013.
41. IQTIG. Qualitätsreport 2015. 2016. https://www.iqtig.org/ergebnisse/qualitaetsreport/. Accessed 17 Oct 2016.
42. Stausberg J, Berghof K. Qualität der stationären Versorgung in Deutschland. Eine Analyse der Entwicklung zwischen 2004 und 2008 aus Daten der externen vergleichenden Qualitätssicherung. Dtsch Med Wochenschr. 2014;139:181–6. doi:10.1055/s-0033-1359990.
43. Freisinger E, Fuerstenberg T, Malyar NM, Wellmann J, Keil U, Breithardt G, Reinecke H. German nationwide data on current trends and management of acute myocardial infarction: discrepancies between trials and real-life. Eur Heart J. 2014;35:979–88. doi:10.1093/eurheartj/ehu043.
44. Krogias C, Bartig D, Kitzrow M, Weber R, Eyding J. Trends of hospitalized acute stroke care in Germany from clinical trials to bedside. Comparison of nation-wide administrative data 2008-2012. J Neurol Sci. 2014;345:202–8. doi:10.1016/j.jns.2014.07.048.
45. Gondos A, Holleczek B, Arndt V, Stegmaier C, Ziegler H, Brenner H. Trends in population-based cancer survival in Germany: to what extent does progress reach older patients? Ann Oncol. 2007;18:1253–9. doi:10.1093/annonc/mdm126.
46. Kallmayer MA, Tsantilas P, Knappich C, Haller B, Storck M, Stadlbauer T, et al. Patient characteristics and outcomes of carotid endarterectomy and carotid artery stenting: analysis of the German mandatory national quality assurance registry - 2003 to 2014. J Cardiovasc Surg (Torino). 2015;56:827–36.
47. Kroch EA, Duan M, Silow-Carroll S, Meyer JA. HOSPITAL PERFORMANCE IMPROVEMENT: TRENDS IN QUALITY AND EFFICIENCY. 2007.
48. Shih T, Nicholas LH, Thumma JR, Birkmeyer JD, Dimick JB. Does pay-for-performance improve surgical outcomes? An evaluation of phase 2 of the Premier Hospital Quality Incentive Demonstration. Ann Surg. 2014;259:677–81. doi:10.1097/SLA.0000000000000425.
49. Cohen ME, Liu Y, Ko CY, Hall BL. Improved Surgical Outcomes for ACS NSQIP Hospitals Over Time: Evaluation of Hospital Cohorts With up to 8 Years of Participation. Ann Surg. 2015;263:267–73. doi:10.1097/SLA.0000000000001192.
50. Sugiyama T, Hasegawa K, Kobayashi Y, Takahashi O, Fukui T, Tsugawa Y. Differential time trends of outcomes and costs of care for acute myocardial infarction

hospitalizations by ST elevation and type of intervention in the United States, 2001-2011. J Am Heart Assoc. 2015;4:e001445. doi:10.1161/JAHA.114.001445.
51. Jacques RM, Fotheringham J, Campbell MJ, Nicholl J. Did hospital mortality in England change from 2005 to 2010? A retrospective cohort analysis. BMC Health Serv Res. 2013;13:216. doi:10.1186/1472-6963-13-216.
52. Neuburger J, Currie C, Wakeman R, Tsang C, Plant F, Stavola B de, et al. The impact of a national clinician-led audit initiative on care and mortality after hip fracture in England: an external evaluation using time trends in non-audit data. Med Care. 2015;53:686–91. doi:10.1097/MLR.0000000000000383.
53. van den Berg, M, de Boer D, Gijsen R, Heijink R, Limburg L, Zwakhals S. Dutch Health Care Performance Report 2014. Bilthoven; 2015.
54. Falcoz P-E, Puyraveau M, Rivera C, Bernard A, Massard G, Mauny F, et al. The impact of hospital and surgeon volume on the 30-day mortality of lung cancer surgery: A nation-based reappraisal. J Thorac Cardiovasc Surg. 2014;148:841-8; discussion 848. doi:10.1016/j.jtcvs.2014.01.030.
55. Vos M de, Graafmans W, Kooistra M, Meijboom B, van der Voort P, Westert G. Using quality indicators to improve hospital care: a review of the literature. Int J Qual Health Care. 2009;21:119–29. doi:10.1093/intqhc/mzn059.
56. Ketelaar N, Faber MJ, Flottorp S, Rygh LH, Deane KHO, Eccles MP. Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. Cochrane Database Syst Rev. 2011:CD004538. doi:10.1002/14651858.CD004538.pub2.
57. Dranove D, Jin GZ. Quality Disclosure and Certification: Theory and Practice. Journal of Economic Literature. 2010;48:935–63. doi:10.1257/jel.48.4.935.
58. Schwenk U, Schmidt-Kaehler S. Public Reporting. Spotlight Gesundheit. 2016;1.
59. Werner RM, Norton EC, Konetzka RT, Polsky D. Do consumers respond to publicly reported quality information? Evidence from nursing homes. J Health Econ. 2012;31:50–61. doi:10.1016/j.jhealeco.2012.01.001.
60. Bardach NS, Hibbard JH, Greaves F, Dudley RA. Sources of traffic and visitors' preferences regarding online public reports of quality: web analytics and online survey results. J Med Internet Res. 2015;17:e102. doi:10.2196/jmir.3637.
61. Gaynor M, Moreno-Serra R, Propper C. Can competition improve outcomes in UK health care? Lessons from the past two decades. J Health Serv Res Policy. 2012;17 Suppl 1:49–54. doi:10.1258/jhsrp.2011.011019.
62. Bloom N, Propper C, Seiler S, van Reenen J. The Impact of Competition on Management Quality: Evidence from Public Hospitals. The Review of Economic Studies. 2015;82:457–89. doi:10.1093/restud/rdu045.
63. Propper C, Burgess S, Gossage D. Competition and Quality: Evidence from the NHS Internal Market 1991-9*. The Economic Journal. 2008;118:138–70. doi:10.1111/j.1468-0297.2007.02107.x.
64. Pauly MV, McGuire TG, Barros PP, editors. Handbook of Health Economics. Amsterdam, London: North Holland; 2012.
65. Kessler DP, Geppert JJ. The Effects of Competition on Variation in the Quality and Cost of Medical Care. J Economics Management Strategy. 2005;14:575–89. doi:10.1111/j.1530-9134.2005.00074.x.
66. Chou S-Y, Deily ME, Li S, Lu Y. Competition and the impact of online hospital report cards. J Health Econ. 2014;34:42–58. doi:10.1016/j.jhealeco.2013.12.004.

67. Gowrisankaran G, Town RJ. Competition, Payers, and Hospital Quality1. Health Serv Res. 2003;38:1403–22. doi:10.1111/j.1475-6773.2003.00185.x.
68. Hentschker C, Mennicken R, Schmid A. Defining hospital markets - an application to the German hospital sector. Health Econ Rev. 2014;4:28. doi:10.1186/s13561-014-0028-0.
69. Schmid A, Ulrich V. Consolidation and concentration in the German hospital market: the two sides of the coin. Health Policy. 2013;109:301–10. doi:10.1016/j.healthpol.2012.08.012.
70. Schmid A, Varkevisser M. Hospital merger control in Germany, the Netherlands and England: Experiences and challenges. Health Policy. 2016;120:16–25. doi:10.1016/j.healthpol.2015.11.002.
71. Bundeskartellamt. Sector inquiry into hospitals. 2016. https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2016/31_05_2016_Sektoruntersuchung_Krankenhaeuser.html;jsessionid=4F31E76D16CDB29A0B546BFDEEB4944C.1_cid387?nn=3591568. Accessed 16 Nov 2016.
72. Hillner BE, Smith TJ, Desch CE. Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care. J Clin Oncol. 2000;18:2327–40. doi:10.1200/jco.2000.18.11.2327.
73. Lee KCL, Sethuraman K, Yong J. On the Hospital Volume and Outcome Relationship: Does Specialization Matter More Than Volume? Health Serv Res. 2015;50:2019–36. doi:10.1111/1475-6773.12302.
74. Hollingsworth B. Revolution, evolution, or status quo?: Guidelines for efficiency measurement in health care. J Prod Anal. 2012;37:1–5. doi:10.1007/s11123-011-0221-7.
75. Hollingsworth B, Street A. The market for efficiency analysis of health care organisations. Health Econ. 2006;15:1055–9. doi:10.1002/hec.1169.
76. McKay NL, Deily ME. Cost inefficiency and hospital health outcomes. Health Econ. 2008;17:833–48. doi:10.1002/hec.1299.
77. Nayar P, Ozcan YA, Yu F, Nguyen AT. Benchmarking urban acute care hospitals: efficiency and quality perspectives. Health care management review. 2013;38:137–45. doi:10.1097/HMR.0b013e3182527a4c.
78. Felder S, Tauchmann H. Federal state differentials in the efficiency of health production in Germany: an artifact of spatial dependence. European Journal of Health Economics. 2013;14:21–39.
79. Herwartz H, Strumann C. On the effect of prospective payment on local hospital competition in Germany. Health Care Managment Science. 2012;15:48–62.
80. G-BA. Regelungen zum Qualitätsbericht der Krankenhäuser, 2013. 2014. https://www.g-ba.de/downloads/62-492-906/Qb-R_2014-06-19.pdf. Accessed 9 Sep 2015.
81. G-BA. Regelungen zum Qualitätsbericht 2012; 2013.
82. QSR. Statistische Verfahren. 2014. http://www.qualitaetssicherung-mit-routinedaten.de/methoden/stat_verfahren/index.html. Accessed 29 Jul 2015.
83. QSR Verfahren. QSR-Indikatorenhandbuch 2016 für Leistungsbereiche mit Berichterstattung im AOK-Krankenhausnavigator. 2016.
84. Sutherland K, Dawson S. Power and quality improvement in the new NHS: the roles of doctors and managers. Qual Health Care. 1998;7 Suppl:S16-23.

85. Currie V, Harvey G, West E, McKenna H, Keeney S. Relationship between quality of care, staffing levels, skill mix and nurse autonomy: literature review. J Adv Nurs. 2005;51:73–82. doi:10.1111/j.1365-2648.2005.03462.x.
86. Prütz F. Was ist Qualität im Gesundheitswesen? Ethik Med. 2012;24:105–15. doi:10.1007/s00481-012-0189-5.
87. Lohr KN. Medicare: A strategy for quality assurance. Washington, D.C.: National Academy Press; 1990.
88. National Institute for Medicine. Crossing the quality chasm: A new health system for the 21st century. Washington, D.C.: National Academy Press; 2001.
89. Ruprecht T. Von der Qualitätssicherung zum Qualitätsmanagement. Entwicklung in der vertragsärztlichen Versorgung. Zeitschrift für Allgemeinmedizin. 1993:963–7.
90. G-BA. Die gesetzlichen Qualitätsberichte 2012 der Krankenhäuser lesen und verstehen. 2014. https://www.g-ba.de/downloads/17-98-3049/2014-03-21_Lesehilfe-Qb.pdf=nDElXdo-sR27bLVhuoQa2g&cad=rja. Accessed 14 Sep 2015.
91. Donabedian A. Evaluating the Quality of Medical Care. Milbank Q. 1966;44.
92. Donabedian A. The role of outcomes in quality assessment and assurance. QRB Qual Rev Bull. 1992;18:356–60.
93. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. The New England Journal of Medicine. 2013;368:201–3. doi:10.1056/NEJMp1211775.
94. Copnell B, Hagger V, Wilson SG, Evans SM, Sprivulis PC, Cameron PA. Measuring the quality of hospital care: an inventory of indicators. Intern Med J. 2009;39:352–60. doi:10.1111/j.1445-5994.2009.01961.x.
95. Gandjour A, Kleinschmit F, Littmann V, Lauterbach KW. An evidence-based evaluation of quality and efficiency indicators. Qual Manag Health Care. 2002;10:41–52.
96. Donabedian A. Evaluating the quality of medical care. Milbank Q. 2005;83:691–729. doi:10.1111/j.1468-0009.2005.00397.x.
97. Arrow K. Uncertainty and the Welfare Economics of Medical Care. American Economic Review. 1963;53:941–73.
98. Buchanan A. Principal/agent theory and decision making in health care. Bioethics. 1988;2:317–33.
99. Sloan FA, Hsieh C-R. Health economics. Cambridge, Mass.: MIT Press; 2012.
100. Pratt JW, Zeckhauser RJ. Principals and agents: The structure of business. Boston, Mass.: Harvard Business School Press; 1985.
101. Haas-Wilson D. Arrow and the information market failure in health care: the changing content and sources of health care information. J Health Polit Policy Law. 2001;26:1031–44.
102. Gaynor M. What Do We Know About Competition and Quality in Health Care Markets? NBER Working Paper 2006. doi:10.3386/w12301.
103. Kessler DP, McClellan MB. Is Hospital Competition Socially Wasteful? The Quarterly Journal of Economics. 2000;115:577–615. doi:10.1162/003355300554863.
104. Morrisey MA, Sloan FA, Valvona J. Defining Geographic Markets for Hospital Care. Law and Contemporary Problems. 1988;51:165. doi:10.2307/1191730.
105. Dranove D, Shanley M. A note on the relational aspects of hospital market definitions. J Health Econ. 1990;8:473–8. doi:10.1016/0167-6296(90)90028-2.

106. Chandra A, Finkelstein A, Sacarny A, Syverson C. Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector. Am Econ Rev. 2016;106:2110–44. doi:10.1257/aer.20151080.
107. GKV Spitzenverband. 14 Positionen für 2014. 2013. https://www.gkv-spitzenverband.de/presse/themen/krankenhausreform/krankenhausreform.jsp.
108. McGuire T. Physician agency. In: Culyer AJ, Newhouse JP, editors. Handbook of health economics. Amsterdam: Elsevier; 2000. p. 461–536.
109. Ellis RP, McGuire TG. Provider behavior under prospective reimbursement. Cost sharing and supply. J Health Econ. 1986;5:129–51.
110. Jones R. Declining altruism in medicine. BMJ. 2002;324:624–5.
111. Workman S. Declining altruism in medicine. Altruism is not equal to self sacrifice. BMJ. 2002;324:1398.
112. Chang T, Jacobson M. What is the Mission of a Not-For-Profit Hospital? NBER Working Paper. 2010.
113. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy. 2011;4:47–55. doi:10.2147/RMHP.S12985.
114. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. Health Aff (Millwood). 2015;34:2174–80. doi:10.1377/hlthaff.2015.0992.
115. Richter-Kuhlmann E, Flintrop J. Das Gespräch mit Prof. Dr. med. Ralf Kuhlen, medizinischer Konzerngeschäftsführer der Helios Kliniken GmbH, Berlin: Der Skepsis Qualität entgegensetzen. Deutsches Ärzteblatt. 2014;111:35–6.
116. Lindlbauer I, Schreyögg J, Winter V. Changes in technical efficiency after quality management certification: A DEA approach using difference-in-difference estimation with genetic matching in the hospital industry. European Journal of Operational Research. 2016;250:1026–36. doi:10.1016/j.ejor.2015.10.029.
117. Wiedmann S, Heuschmann PU, Hillmann S, Busse O, Wietholter H, Walter GM, et al. The quality of acute stroke care- an analysis of evidence-based indicators in 260 000 patients. Dtsch Arztebl Int. 2014;111:759–65. doi:10.3238/arztebl.2014.0759.
118. Wesselmann S, Beckmann MW, Winter A. The concept of the certification system of the German Cancer Society and its impact on gynecological cancer care. Arch Gynecol Obstet. 2014;289:7–12. doi:10.1007/s00404-013-3084-5.
119. Chop I. Leitfaden Ärztliches Peer Review. 2014. http://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/Leitfaden_Aerztliches-Peer-Review_2014.pdf. Accessed 14 Dec 2016.
120. Gottweis H. Verwaltete Körper: Strategien der Gesundheitspolitik im internationalen Vergleich. Wien: Böhlau; 2004.
121. Spahn J, Müschenich M, Debatin JF. App vom Arzt: Bessere Gesundheit durch digitale Medizin. Freiburg: Verlag Herder; 2016.
122. Breyer F, Zweifel P, Kifmann M. Gesundheitsökonomik. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013.
123. DNVF. Stellungnahme des Deutschen Netzwerk Versorgungsforschung (DNVF) e.V. zum Vorbericht Planungsrelevante Qualitätsindikatoren des Institutes für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) in der Fassung vom 18.7.2016. 2016. http://www.netzwerk-versorgungsforschung.de/uploads/Stellungnahmen/DNVF-Stellungnahme_IQTIG_PlanQI.pdf. Accessed 19 Dec 2016.

124. KMA. Bundesärztekammer kritisiert IQTIG-Qualitätsindikatoren. KMA. 19.08.2016.
125. WIdO. QSR-Indikatorenhandbuch 2016 für Leistungsbereiche mit Berichterstattung im AOK-Krankenhausnavigator. 2016. http://www.qualitaetssicherung-mit-routinedaten.de/methoden/bereiche/index.html. Accessed 15 Oct 2016.
126. WIdO. QSR-Indikatorenhandbuch 2016 für Leistungsbereiche ohne Berichterstattung im AOK-Krankenhausnavigator. 2016. http://www.qualitaetssicherung-mit-routinedaten.de/methoden/bereiche/index.html. Accessed 15 Oct 2016.
127. IQM. Sachbericht. 2014. http://www.initiative-qualitaetsmedizin.de/mediapool/1056/media_file/. Accessed 22 Feb 2015.
128. 4QD. Qualitätskliniken.de Klinikvergleich. 2015. http://qualitätskliniken.de/klinikdetails.php?clinic=1586&distance=11. Accessed 31 Aug 2015.
129. Berwick DM, James B, Coye MJ. Connections between Quality Measurement and Improvement. Med Care. 2003;41.
130. Debatin J. Krankenhäuser: Mehr Qualität und Effizienz durch Wettbewerb. In: Schumpelick V, Vogel B, editors. Medizin zwischen Humanität und Wettbewerb: Probleme, Trends und Perspektiven : Beiträge des Symposiums vom 27. bis 30. September 2007 in Cadenabbia. Freiburg: Herder; 2008.
131. Busse R, Nimptsch U, Mansky T. Measuring, monitoring, and managing quality in Germany's hospitals. Health Aff (Millwood). 2009;28:w294-304. doi:10.1377/hlthaff.28.2.w294.
132. Nimptsch U, Mansky T. Quality measurement combined with peer review improved German in-hospital mortality rates for four diseases. Health Aff (Millwood). 2013;32:1616–23. doi:10.1377/hlthaff.2012.0925.
133. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. Qual Saf Health Care. 2006;15:433–6. doi:10.1136/qshc.2006.018549.
134. Dranove D, Sfekas A. Start spreading the news: a structural estimate of the effects of New York hospital report cards. J Health Econ. 2008;27:1201–7. doi:10.1016/j.jhealeco.2008.03.001.
135. Pope DG. Reacting to rankings: evidence from "America's Best Hospitals". J Health Econ. 2009;28:1154–65. doi:10.1016/j.jhealeco.2009.08.006.
136. Bruch HP, Bürk C, Hänisch E, Weigel T. Qualitätssicherung mit Routinedaten – P4P durch die Hintertür? 2015. http://www.bdc.de/qualitaetssicherung-mit-routinedaten-p4p-durch-die-hintertuer-3/. Accessed 15 Dec 2016.
137. WIdO. Qualitätssicherung mit Routinedaten (QSR). 2007. http://qs-mit-routinedaten.de/entwicklung/. Accessed 12 Sep 2014.
138. Romley JA, Goldman DP. How costly is hospital quality? A revealed-preference approach. J Ind Econ. 2011;59:578–608.

# Kapitel 2: Qualitätstransparenz im internationalen Vergleich

Christoph Pross, Alexander Geissler, Reinhard Busse

Postprint, erschienen im März 2017 als: Pross C, Geissler A, Busse R. Measuring, Reporting, and Rewarding Quality of Care in 5 Nations: 5 Policy Levers to Enhance Hospital Quality Accountability. *Milbank Quarterly*. 2017;95:136–83. http://dx.doi.org/10.1111/1468-0009.12248

**Hintergrund:** In den letzten Jahren haben zahlreiche Studien in verschiedenen Ländern gravierende Unterschiede in der Ergebnisqualität zwischen Leistungserbringern festgestellt. Als Antwort darauf haben Gesundheitssysteme die Qualitätstransparenz durch systematische Messung und Veröffentlichung von Struktur-, Prozess- und Ergebnisindikatoren erhöht. Patienten und einweisende Ärzte sollen befähigt werden, bessere Krankenhäuser auszuwählen. Der Handlungsdruck auf Krankenhäuser mit unterdurchschnittlicher Qualität soll erhöht werden. Neben der Veröffentlichung von Indikatoren werden zunehmend qualitätsorientierte Vergütungsmodelle entwickelt und implementiert. Obwohl es zahlreiche Initiativen in verschiedenen Ländern, erste Erfahrungen mit diesen Initiativen und ein hohes Potential für die Identifizierung und Adaption länderübergreifender Best Practices gibt, sind aktuelle internationale Studien, welche Qualitätstransparenz holistisch entlang der Ebenen Messung, Veröffentlichung und Incentivierung betrachten, Mangelware.

**Ziele**: Für die fünf in der Qualitätstransparenz führenden Länder Deutschland, England, Niederlande, Schweden und die USA werden die wichtigsten Initiativen und deren Umfang entlang der Dimensionen Messung, Veröffentlichung und Incentivierung von Qualität identifiziert. Die Ergebnisse werden grafisch anschaulich und komprimiert dargestellt, um die Zugänglichkeit für Gesundheitspolitiker und Experten zu erhöhen. Auf Basis der

Länderanalysen werden fünf wichtige gesundheitspolitische Ansätze identifiziert, mit denen die Qualitätstransparenz entscheidend verbessert werden kann.

**Methoden**: Auf Basis von Experteninterviews und einer umfangreichen Literaturrecherche wurde ein Analysenrahmen entlang der Dimensionen Messung, Veröffentlichung und Incentivierung von Qualität entwickelt. Für jede der drei Dimensionen wurden zwei wichtige Subdimensionen identifiziert. Für Qualitätsmessung sind das Art der Indikatoren und Datenquellen. Für Qualitätsveröffentlichung sind das Grad der Zentralisierung und Art des Datenzugangs. Für Qualitätsanreizstrukturen sind das Grad der Anwendung und Art der qualitätsorientierten Vergütung. Die Ergebnisse der Länderanalysen werden in drei verschiedenen Detaillierungsgraden (grafische Übersicht, tabellarische Übersicht und detaillierter Bericht) dargestellt. Die Politikempfehlungen werden auf dieser Grundlage hergeleitet und mit Positiv- und Negativbeispielen aus den Länderanalysen illustriert.

**Ergebnisse**: Die Messung von Qualität ist über die Länder hinweg vergleichbarer als die Veröffentlichung und die Incentivierung. Viele Länder fokussieren ihre Veröffentlichung und Incentivierung stark auf Prozessindikatoren. Deutschland hat bisher nur kleine Pilotprojekte zur qualitätsorientierten Vergütung gestartet; in Schweden, den Niederlanden und England sind diese Pilotprojekte schon deutlich umfangreicher und finden teilweise auf nationaler Ebene statt. In den USA ist qualitätsorientierte Vergütung bereits ein wichtiger und umfangreicher Teil des Krankenhausvergütungssystems. Um Qualitätstransparenz zu fördern, sind fünf Ansätze wichtig: 1) Staatliche Führung ist notwendig um Standards und Anreize festzulegen. 2) Die richtige Balance zwischen Zentralisierung und Dezentralisierung ist entscheidend für nationale Vergleichbarkeit und gleichzeitig für Innovation und lokale Teilhabe. 3) Gesundheitssysteme sollten sich stärker auf Ergebnisindikatoren und weniger auf Prozessindikatoren konzentrieren. 4) Leistungserbringer spielen eine zentrale Rolle in der Qualitätstransparenz und müssen als Unterstützer gewonnen werden. 5) Qualitätstransparenz sollte auf der Ebene eines Krankenhauses und von medizinischen Leistungsbereichen ansetzen.

THE
MILBANK QUARTERLY
A MULTIDISCIPLINARY JOURNAL OF POPULATION HEALTH AND HEALTH POLICY

*Original Investigation*

# Measuring, Reporting, and Rewarding Quality of Care in 5 Nations: 5 Policy Levers to Enhance Hospital Quality Accountability

CHRISTOPH PROSS,* ALEXANDER GEISSLER,* and REINHARD BUSSE*,†

**Berlin University of Technology; †European Observatory on Health Systems and Policies*

**Policy Points:**

- Similarities and disparities between countries and initiatives are identified. Measuring, reporting, and rewarding quality is heavily focused on process measures. Hospital-level benchmarking is not always available publicly. Quality-related payment schemes vary widely, with several countries only piloting small-scale initiatives.
- To increase quality accountability, the government has to set standards and incentives. The right balance between system centralization and decentralization has to be struck. Accountability needs to be based on outcomes, not process measures, and focus should be on hospital and medical condition levels. Providers have a central role as quality accountability advocates.

**Context:** Studies have documented wide quality variation among hospitals within and across countries. Increasing quality-of-care accountability for hospitals, especially for patients and the general public, is an important policy objective, but no study has yet systematically and comprehensively compared leading countries' initiatives in this regard.

**Methods:** Based on expert interviews and an extensive literature review, we investigate hospital quality accountability in England, Germany, the Netherlands, Sweden, and the United States. The underlying framework includes 3 elements: measuring quality, reporting quality, and rewarding quality. Each element is subdivided into 2 dimensions, with measuring composed of indicator type and data source, reporting composed of degree of reporting centralization

The Milbank Quarterly, Vol. 95, No. 1, 2017 (pp. 136-183)

and data accessibility, and rewarding composed of extent of application and type of quality-related payments.

**Findings:** The results show a wide spectrum of approaches and progress levels. Measuring strategies are more similar across countries, while quality reporting and financial rewards are more dissimilar. Reporting of process indicators is more prevalent than reporting of outcomes. Most countries have introduced some quality-related payment schemes, with the United States having the most comprehensive approach. Based on the cross-country assessment, 5 policy levers to enhance quality transparency are identified and illustrated through country-specific examples: (1) the government should take a central role in establishing standards and incentives for quality transparency and health IT system integration; (2) system centralization and decentralization need to be balanced to ensure both national comparability and local innovation; (3) health systems need to focus more on outcome transparency and less on process measures; (4) health systems need to engage providers as proponents of quality transparency; and (5) reporting should focus on hospital and condition levels to ensure comparability and enable meaningful patient choice.

**Conclusions:** The findings facilitate cross-country learning and best-practice adoption by assessing hospital quality accountability strategies in 5 countries in a structured and comparative manner. The identified policy levers are relevant for enhancing breadth, depth, and value of quality accountability.

**Keywords:** health care quality assessment, quality accountability, pay-for-performance, hospital quality of care, public reporting, comparative health policy.

Recent studies indicate that quality of care varies considerably across providers and medical conditions.[1-3] Advanced health systems are thus encouraging quality accountability both to stimulate provider competition around quality and to support patient choice through public reporting of quality variation among providers. Many providers, however, resist the increased pressure and resource impact associated with collection of clinical quality data and quality-related payments.[4,5] Likewise, patients often lack awareness of existing accountability initiatives or encounter confusing results for the same provider across different initiatives.[6,7]

Since the early 2000s, England, Germany, Sweden, the United States, and, more recently, the Netherlands have particularly championed quality accountability. There are similarities in how these countries organize measurement, reporting, and incentivization of quality; there are also important differences, including underlying data and indicators, degrees of mandatory versus voluntary reporting, and approaches to quality-related payments (QRP).[8] These are partially determined by disparities in health system features such as insurance and payment schemes, size of inpatient versus outpatient sector, and private sector service provision.[9,10]

Health systems still lack standardization and integration across countries regarding quality accountability, with limited learning and adaptation from international comparison.[11,12] Thorough cross-country analysis of quality accountability approaches is rare and, where existent, outdated.[8] Furthermore, the dimensions of quality accountability have primarily been examined in isolation,[13] and a broad framework to describe and analyze the policy status of quality accountability in a comparative manner across countries is lacking. Likewise, an up-to-date analysis of key policy levers to improve quality accountability is needed, as health systems have gathered first experiences with quality accountability[14,15] and key questions on the mechanisms of quality accountability remain unclear.[6,7,16,17]

Policymakers and other stakeholders in the 5 countries are examining their hospital quality accountability approaches and preparing to initiate the next phase, in which quality will become more transparent for all stakeholders and relevant for provider payment. Countries can benefit substantially from an international comparative analysis of policy approaches and experiences; best-practice identification and adoption; and harmonization of indicators measured, reported, and included in QRP schemes to facilitate comparison.

This study addresses this research gap by introducing a more comprehensive, policy-oriented hospital quality accountability framework, which is then used to provide a cross-country perspective on the state of the art of measuring, reporting, and rewarding hospital quality, informed by in-depth country research. We also identify 5 policy levers important for the advancement of quality accountability and illustrate these with country examples.

## Methods

The key elements of quality accountability are (1) measuring provider quality; (2) reporting provider quality; and (3), based on the obtained data, incentivizing providers through QRPs. Each of these elements comprises a range of components and approaches, heavily influenced by individual health system parameters. To comprehensively assess the quality accountability systems and the 3 key elements, we conducted country reviews based on expert interviews, a review of academic and gray literature, and an examination of public reporting online portals.

In each of the countries (other than Germany), we conducted 4–6 interviews with academic, regulatory, and industry experts (displayed in Appendix Table A1). We engaged experts both at the start of our country analysis to understand key elements and institutions of quality accountability, as well as at later points to collect detailed information and to review and comment on finalized sections. We reviewed the relevant academic literature based on a PubMed search strategy that included terms such as "quality of care," "outcome measurement," "quality reporting," and "value-based payments" (including different varieties thereof) linked with the terms "hospital" and "health care" and the respective country names. In addition, we investigated gray literature (eg, reports, documentations, and press releases), which we found through online desk research and on the websites of the relevant portals and agencies. We mostly included literature from 2005 onward, with a focus on more recent material to account for the constantly changing policy environment and evolving research and academic positions. We also received literature from our interviewed country experts. Lastly, we reviewed and tested the public reporting and benchmarking portals (eg, by conducting test hospital benchmarking) in the different countries.

We synthesized the assessment along an a priori outlined and continuously refined framework that captures the aforementioned elements and their most important dimensions:

1. **Measuring Quality**: Composed of *indicator type* (structural, process, outcome, or risk-adjusted outcome) and *data source* (clinical, administrative, or patient-reported).
2. **Reporting Quality**: Composed of *reporting centralization* (individual by medical condition, some grouping of conditions, or centralized for all conditions) and *data accessibility* (internal reporting, public reporting, or public benchmarking).
3. **Rewarding Quality**: Composed of *extent of QRP application* (pilots only, regional application or for several conditions,

large-scale or multicondition) and *type of quality-related payment* (bonus-malus, bundled payments, or capitation payments).

For all 3 quality accountability elements, we separately assessed the status for each country to paint a comprehensive picture for measurement, reporting, and rewarding of quality of care. Thus, we aimed to inform the use and further development of each. The 3 elements and their respective subdimensions across the 5 countries are also illustrated in 2-dimensional matrices, which convey the scope of the approaches to measuring, reporting, and rewarding quality, and indicate overlaps among country strategies.

This analysis focused on inpatient hospital care, which is most comparable across countries and it is where quality accountability is most advanced. Additionally, the analysis concentrated on hospital-level quality measuring and reporting, often captured in provider report cards, and less on national, regional, or physician-level reporting.

The framework was initially drafted after a preliminary literature review and after first discussions with experts in each country. After testing and with additional research insights, we refined the framework and used it to structure, record, and report both the academic and gray literature review as well as the expert interviews. As an organizing structure, we targeted a simple and easily accessible framework to maximize accessibility and comprehensibility for policymakers.

## Results

### *Measuring Quality*

Quality measurement is the foundation of quality accountability. The national approach to evaluate hospital quality determines which indicators can be utilized for reporting as well as rewarding providers. Countries make choices regarding voluntary versus mandatory, national versus regional, and public versus private measurement initiatives. Moreover, countries choose between using administrative, clinical, and patient-reported data as well as patient reporting and a combination of structural, process, outcome, and risk-adjusted outcome indicators. These decisions influence the width (ie, number of hospitals covered) and depth (ie, number of indicators measured) of quality measurement. They also determine the resource impact on providers necessary to collect the data. Based on different choices along these parameters, each country has a distinct quality measurement ecosystem.

*England.* The National Health Service (NHS) standard contract requires all NHS trusts and contractors to participate in all clinical audits, clinical outcome review programs, and registries on the NHS England Quality Accounts List. The audits and registries included in the Healthcare Quality Improvement Partnership (HQIP) Quality Accounts oblige providers to measure results across a variety of treatment areas and indicators that are based on administrative, clinical, and patient-reported data. For 2015-2016, the Quality Accounts List included 41 hospital-level programs such as the National Joint Registry (NJR) and the National Adult Cardiac Surgery Audit (see Appendix Table A2 for a list of abbreviations by country). In addition, the Commissioning for Quality and Innovation (CQUIN), set up by the Department of Health in 2009, measures 52 indicators, consisting mostly of structural and process metrics and a few selected outcomes. The CQUIN indicators, however, are sometimes seen as complex and not comparable across trusts. Furthermore, Public Health England, an executive agency of the UK Department of Health, collects hospital-level infection rates for *Staphylococcus aureus*, *C. difficile,* and other hospital-acquired infections.

The Health and Social Care Information Centre (HSCIC) functions as the central data clearing house for NHS and its contractors. The HSCIC administers the Hospital Episode Statistics, a routine data collection, during a patient's hospital stay. The HSCIC also generates quarterly the widely used Summary Hospital-level Mortality Indicator (SHMI), which is a risk-adjusted mortality rate at the NHS trust level. The SHMI is defined as death occurring in the hospital and up to 30 days after discharge, and it included 137 trusts in 2014. In-hospital and post-discharge administrative information are linked together through the NHS number, which is a unique patient ID and is required for all NHS services. The HSCIC also collects patient-reported outcome measures (PROM) for the treatment areas of hip and knee replacements, groin hernia, and varicose veins.[18]

The NJR covers hip, knee, ankle, elbow, and shoulder joint replacements and collects quality measures at the hospital and physician level. Indicators include case volume, clinical outcomes (eg, 90-day risk-adjusted mortality and readmission rates), PROMs such as Oxford Score and quality of life (eg, EQ-5D), and patient characteristics. The National Institute for Cardiovascular Outcomes Research operates

6 clinical cardiac audits, including cardiac surgery, percutaneous coronary intervention (PCI), and heart failure. Indicators include mortality for emergency operations, risk-adjusted coronary artery bypass grafting (CABG) and aortic valve replacement surgery, and patient population risk factors. While clinical registry participation is widespread, it varies between providers in scope and thoroughness. Moreover, the annual NHS Inpatient Survey, conducted by the Care Quality Commission, measures patient experience at each NHS service hospital across treatment areas and includes measures such as trust in clinicians, patient involvement, cleanliness, received respect and dignity, and an overall composite measure.

*Germany.* Along with the introduction of diagnosis-related groups (DRG) in the early 2000s, the self-governance system of payers and providers and its highest body, the Gemeinsamer Bundesausschuss (Federal Joint Committee; G-BA), propelled by correspondent legislation, introduced a statutory quality assurance system.[19] Annual, self-reported provider report cards compile structural, process, and outcome indicators at the hospital and medical department level for 30 tracer diagnoses and procedures, covering less than 30% of all inpatient cases across all 1,600 acute hospitals in the country.[20] More than 350 process and outcome indicators are collected, including evidence-based care compliance rates; readmission, infection, and mortality rates; and risk-adjusted readmission and mortality rates. Indicators are based on clinical data for the inpatient episode only. Further development and implementation of the quality assurance system has recently been centralized with the Institute for Quality and Transparency in Health Care (IQTIG).

Besides mandatory, national quality assurance, 3 notable measurement initiatives are Qualitätssicherung mit Routinedaten (QSR), which is operated by Allgemeine Ortskrankenkassen (AOK), the largest sickness fund, and Initiative Qualitätsmedizin (IQM) and Qualitätskliniken.de (4QD), which were both initiated by leading private hospital chains. QSR calculates risk-adjusted outcome indicators, such as 30-, 90-, and 365-day mortality rates; readmission and reoperation rates; and a composite indicator for 14 medical conditions and procedures such as acute myocardial infarction (AMI), PCI, stroke, appendicitis, prostate surgery, and hip and knee implants. The QSR initiative uses AOK administrative patient data and includes complications after hospital discharge.

IQM and 4QD, which both had around 300 member hospitals in 2015, pursue similar administrative data-based approaches to measurement. IQM calculates its indicators, primarily case volume and raw and risk-adjusted mortality rates, for 40 treatment areas such as stroke, chronic obstructive pulmonary disease (COPD), and spine surgery. It uses the German Inpatient Quality Indicator set, which is calculated with administrative inpatient data based on the US Agency for Healthcare Research and Quality (AHRQ) inpatient quality measures.[21] 4QD calculates process and outcome indicators such as blood infection postsurgery, pulmonary embolism postsurgery or colorectal cancer mortality rates, also based on administrative hospital data. IQM and 4QD, as well as the mandatory quality assurance system, strongly emphasize peer reviews to facilitate quality improvement through error identification and best-practice adoption.

Germany's registry infrastructure is currently limited, often voluntary, and primarily regionally based. Recently, however, some national registers, such as the German Joint Replacement Register and a nationwide standardized Cancer Registry, have been established. Furthermore, the patient experience questionnaire (PEQ) captures patient-reported experience measures (PREM) for all German hospitals at the medical department level. It consists of 15 questions in the categories of medical care, nursing care, hospital stay, and global hospital assessment. In Germany, no patient-reported outcome indicators are collected in a standardized, large-scale fashion.

*The Netherlands.* The Dutch Health Care Inspectorate (IGZ) supervises quality, safety, and access in Dutch medical care providers.[22] All hospitals submit a mandatory set of IGZ quality indicators annually to the Dutch Healthcare Authority (NZa). These include structural, process, outcome, and PROM indicators, such as surgery volumes, postoperative controls, reoperation rates, and postsurgery pain levels, with mortality and length of stay risk-adjusted. In 2014, 42 tracer treatment areas—such as cardiac and birth care, hip fracture, and cataract surgery—were covered. The Dutch Hospital Data (DHD) foundation, a consortium of the Dutch Hospital Association (NVZ) and the Federation of University Medical Centers, functions as a data clearing house for administrative data-based indicators and transfers mandatory data to the NZa. Hospital standardized mortality ratios (HSMR) are calculated by the Central Bureau of Statistics for 50 diagnosis groups based on administrative data from DHD. Based on the UK SHMI, the HSMR is a risk-adjusted outcome measure that compares observed death with expected death, adjusting for the underlying patient population.

Several provider initiatives exist as well. The Dutch Institute for Clinical Auditing (DICA) combines 13 registries for conditions such as colorectal and breast cancer, lung and thoracic surgery, and gynecological oncology. DICA covers structural, process, outcome, and PROM indicators. Hospital membership is not mandatory, but due to IGZ pressure and increased public awareness, most hospitals have joined. Meetbaar Beter (measurably better) is another provider-led voluntary registry initiative; it covers cardiovascular diseases and collects clinical outcome and PROM indicators. In 2015, it included 14 out of 16 heart centers. In addition, the voluntary health insurer initiative Consumer Quality Index (CQI) measures PREMs by distributing, collecting, and aggregating questionnaires for 18 medical conditions and care processes.

*Sweden.* Sweden's National Board of Health and Welfare (Socialstyrelsen) compiles administrative information in the National Patient Register for all hospital treatments. Data include a unique patient ID, patient demographic and geographical information, and clinical data. In the Cause of Death Register, Socialstyrelsen collects ICD-10 codes and patient data that can be linked through the patient's ID with Health Care Quality Register data. Traditionally, Swedish quality measurement is based on condition-specific registers. With the first registers established in the 1970s, 100 quality registers exist today and collect structural, process, and outcome indicators from hospitals on a voluntary and mandatory basis for relevant inpatient conditions. Registers are generally based on clinical data, but 75% of registers also collect some PROMs.[23] SWEDEHEART, one of the most prominent registers, also introduced partially risk-adjusted quality indices for its main treatment areas. For example, the TAVI Quality Index is a composite index reflecting structural factors, case volume, serious complications, and risk-adjusted mortality.[24] While SWEDEHEART has 100% coverage for all its cardiovascular conditions, other registers are less comprehensive. However, about 70% of registers cover more than 80% of their target population.[25,26]

The Swedish central and regional governments currently fund these registers with around €30 million annually.[27] To standardize outcome measurement across regions and medical conditions, Sweden consolidated its quality registers in 7 regional competence centers, with the stated purpose of enhancing IT infrastructure and sharing statistical expertise and methodologies across registers.[25] For example, the Uppsala Clinical Research Center operates 18 nationwide quality registers, including SWEDEHEART and SWEDVASC, the vascular surgery register.

*United States.* The National Quality Forum (NQF), a consensus-driven, all-stakeholder body, endorses most quality indicators used in public reporting and QRPs in the United States. NQF measures are preferred by federal programs as well as many state and private initiatives. Indicator development and measurement, however, is undertaken by numerous other organizations.

The AHRQ administers inpatient quality and patient safety indicators (IQI and PSI), which are calculated using Medicare hospital discharge data for 3,500 Medicare-registered, acute care hospitals. The 2015 IQI set includes 34 provider-level and regional indicators such as volume, utilization, and risk- and non-risk-adjusted mortality. With regards to mortality, the IQI indicator set covers esophageal and pancreatic resection, abdominal aortic aneurysm repair, CABG surgery, PCI, carotid endarterectomy procedures, AMI, heart failure, acute stroke, gastrointestinal hemorrhage, hip fracture, and pneumonia. The PSI indicator set includes 26 provider-level complication indicators, which are partially risk-adjusted and only cover complications treated in the same hospital as the initial care episode. Hospital-level PSI indicators cover areas such as pressure ulcer, bloodstream infections, sepsis, postoperative hip fracture, perioperative hemorrhage, postoperative deep vein thrombosis, and birth and obstetric traumas. About 75% of AHRQ provider-level indicators measure outcomes.

The Joint Commission annually collects quality information from its 3,300 member hospitals. The 2015 manual included 73 active indicators, with 56 process indicators in areas like surgical and emergency department care and 15 risk-adjusted outcome indicators from the Centers for Medicare & Medicaid Services (CMS) mortality and readmission measures.[28]

While most registries are regional and/or voluntary, the United States does have a few national mandatory registries. The Centers for Disease Control and Prevention (CDC) runs a mandatory national registry for assisted reproductive technology (ART) procedures. The Society of Thoracic Surgeons (STS) National Database collects voluntarily submitted quality process and outcome metrics for CABG, isolated aortic valve replacement, combined aortic valve replacement, and CABG at the surgeon and hospital level. It covers 90% of all adult cardiac surgery centers in the United States. New York's PCI and CABG registries collect data on

PCI and CABG procedures, including readmission rates, mortality rates, 40 patient risk factors, hospital and physician information, and patients' discharge status. With this data, the New York State Department of Health calculates risk-adjusted 30-day readmission and mortality rates. For CABG, similar statewide registry initiatives exist in Massachusetts, New Jersey, Pennsylvania, and California.

Furthermore, the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) is a standardized, national survey instrument for collecting PREMs. A total of 32 questions are administered to a random sample of adult patients across medical conditions and address aspects of patients' hospital stay, such as pain management, overall hospital rating, and hospital recommendation.

Table 1 (left column) summarizes the quality measurement results for each country; Figure 1 illustrates each country's scope of quality measurement vis-à-vis indicator types employed and underlying data source.

## *Reporting Quality*

Internal reporting provides performance feedback to hospital clinicians and administrators. Public reporting through provider websites or benchmarking portals allows patients and admitting physicians to assess specific hospitals for certain procedures. It also facilitates benchmarking and best-practice sharing among providers. But clinicians often resist public reporting as they doubt the accuracy and comprehensiveness of risk-adjustment methodologies, citing possible negative welfare effects such as patient selection and up-coding.[29] Yet, public reporting can refocus provider competition away from pure volume to a more patient-centric model that includes quality of care.[4] Some countries provide public access to reports (eg, in PDF format) that list the performance of different regions or hospitals along selected quality indicators. More generally, public reporting is often implemented via internet portals that allow specific hospital searches and at times benchmarking. The latest consumer information technology trends also include applications (eg, consumer reports, mobile apps) to make access to publicly available information more intuitive and faster. While most countries today have some public reporting, the information provided, the ability to benchmark hospitals, ease of use, and public and clinical acceptance levels differ widely.

**Table 1.** Summary Results Table for Measuring, Reporting, and Rewarding Quality[a]

| | Measuring Quality | Reporting Quality | Rewarding Quality |
|---|---|---|---|
| **England** | ➢ 41 mandatory clinical audits, registries (Quality Accounts)<br>➢ Volume, process, outcome, and PROMs indicators collected<br>➢ Data at hospital trust and physician levels<br>➢ National Patient Experience Survey reports PREMs | ➢ NHS Choices/MyNHS as main benchmarking portal<br>➢ Composite measures reported at trust level, not at hospital, medical condition levels<br>➢ Quality Accounts on hospital websites<br>➢ Private Dr Foster's Hospital Guide reports mortality rates | ➢ CQUIN with quality penalties at 2.5% of DRG payments<br>➢ Best practice tariffs tie reimbursement to process quality<br>➢ QRPs largely based on process performance |
| **Germany** | ➢ Mandatory report cards with structural and >350 process/outcome indicators for 30 conditions<br>➢ National private payer and provider initiatives<br>➢ PEQ patient experience survey reports PREMs<br>➢ Postdischarge period included in AOK QSR indicators | ➢ Main portal WeLi.de with condition- and hospital-based benchmarking<br>➢ ~275 indicators published, with 70% outcome/15% risk-adjusted<br>➢ AOK portal reporting QSR measures/selected registry participation<br>➢ Providers generally oppose public benchmarking | ➢ Bundled payment pilots for integrated care contracts<br>➢ Small scale, selective contracting pilots by regional payers link to quality<br>➢ Large scale, national bonus-malus payments planned for 2018 |

*Continued*

Table 1. *Continued*

| | Measuring Quality | Reporting Quality | Rewarding Quality |
|---|---|---|---|
| **Netherlands** | ➢ National Health Care Institute collects structural, process, outcomes measures in 42 diagnosis groups<br>➢ HSMRs centrally calculated for 50 diagnosis groups<br>➢ 18 registries collect diagnosis-specific metrics | ➢ Dutch Hospital Data and Dutch Institute for Clinical Auditing (registries) report primarily internally<br>➢ 2 online portals provide public provider-level information by disease/procedure, but benchmarking restricted | ➢ Dutch DBCs with some outpatient components and with selective contracting and free price negotiation<br>➢ Bundled payment pilots in chronic diseases (eg, COPD, diabetes) |
| **Sweden** | ➢ Partly mandatory >100 national quality registers<br>➢ National mortality registry records mortality & causes<br>➢ Volume, process, outcomes, PROMs collected<br>➢ Government professionalizing registries | ➢ Registries with some public reporting of hospital-level data<br>➢ Annual quality report focused on regions<br>➢ Web Guide to Care expanding public reporting; hospital benchmarking not possible yet | ➢ Stockholm region with orthopedics bundles including bonus-malus payments<br>➢ National SVEUS QRP initiative spreading QRP know-how to other regions |

*Continued*

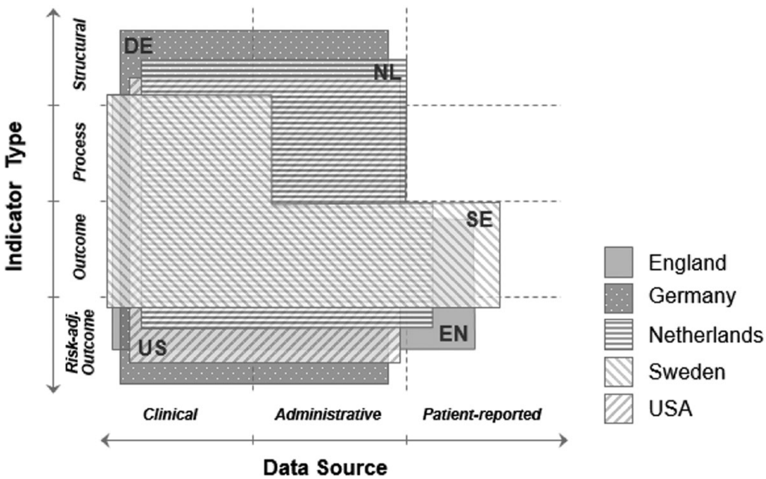
Table 1. *Continued*

| | Measuring Quality | Reporting Quality | Rewarding Quality |
|---|---|---|---|
| **United States** | ➢ All-stakeholder National Quality Forum endorses most measures<br>➢ AHRQ IQI and PSI based on administrative data<br>➢ Joint Commission with own quality measures<br>➢ Few national registries<br>➢ Much local, regional, and private innovation<br>➢ HCAHPS patient experience survey for PREMs | ➢ Hospital Compare reporting AHRQ indicators and PREM hospital star ratings<br>➢ Joint Commission Quality Check and Leapfrog Hospital Safety Score allow online benchmarking<br>➢ *US News*, Healthgrades, Consumer Reports provide benchmarking | ➢ Several Medicare QRPs (HVBP, HRRP, HACRP), partly outcomes-based<br>➢ BPCI initiative introducing bundled payments in existing FFS models<br>➢ Private providers and payers with much innovation in bundled payment and capitation programs |

*Abbreviations:* AOK, Allgemeine Ortskrankenkassen (general sickness funds); AHRQ, Agency for Healthcare Research and Quality; BPCI, Bundled Payments for Care Improvement, COPD, chronic obstructive pulmonary disease; CQUIN, Commissioning for Quality and Innovation; DRG, diagnosis-related group; DBC, diagnosis treatment combination; FFS, fee for service; HACRP, Hospital-Acquired Condition Reduction Program; HCAHPS, Hospital Consumer Assessment of Healthcare Providers and Systems; HRRP, Hospital Readmission Reduction Program; HSMR, hospital standardized mortality ratios; HVBP, Hospital Value-Based Purchasing; IQI, inpatient quality indicator; NHS, National Health Service; PEQ, patient experience questionnaire; PREM, patient-reported experience measures; PROM, patient-reported outcome measure; PSI, patient safety indicators; QRP, quality-related payment; QSR, Qualitätssicherung mit Routinedaten; SVEUS, a Swedish value-based health care knowledge development and sharing project; WeLi.de, Weisse Liste.de.

[a]Data from author analysis of research and expert interviews for each of the 5 countries.

**Figure 1.** Two-dimensional Overview for the Dimension Measuring Quality, Showing Which Type of Quality Indicators Are Collected and Which Data Type Is Used for Indicator Calculation




Source: Authors' analysis based on country research and expert interviews.

*England.* The English health care quality reporting system has expanded substantially and shifted from private, internal reporting to public reporting due to the open data policies between 2007 and 2010. Launched in 2007, NHS Choices integrates all NHS online services. For quality of care, the internet portal enables benchmarking of hospital trusts along a variety of indicators and by geography and service area. Similarly, MyNHS integrates publicly available data for patients to monitor their clinical commissioning group and NHS trust. However, relevant quality indicators are overall composite measure based and not condition based (eg, risk-adjusted hospital mortality rate for the NHS Trust, recommendations by own staff, and ratings of patient safety events). Trust benchmarking is possible, but only for the composite measures and not at the medical condition level. PREMs from the National Inpatient Experience Survey and Care Quality Commission inspection results are also reported on NHS Choices. Several other more focused NHS-affiliated websites report hospital quality information online. For example, Public Health England publishes spreadsheets on infection rates for all English trusts. In addition, the HQIP Quality Accounts are published online on a hospital's website.

Beyond NHS Choices, Dr Foster Intelligence publishes SHMI rates as well as its own standardized mortality rate HSMR, which captures in-patient mortality only. The Dr Foster Hospital Guide allows searches by NHS trust and downloads of league tables. Further, the portal includes benchmarking information on the quality of weekend care and commissioning care results for general practitioner groups. The latest report was published in 2013. HSCIC provides a public portal comparing PROMs for hip and knee replacement and groin hernia and varicose vein treatment. The hip and knee data is also published through a web portal run by the NJR, which allows hospital-specific searches and displays general volume information and clinical indicators such as 90-day, risk-adjusted mortality and revision rates, and PROMs. In general, several highly publicized reports (eg, the Francis report) of investigations into failings at NHS trusts[30] and strong political commitments to transparency[31] have strengthened support for public reporting of hospital quality.

*Germany.* The Weisse Liste.de (WeLi) portal, based on a public mandate, provides access to the searchable and more user-friendly hospital report cards from the national G-BA quality monitoring system. PEQ PREM results are also integrated in the WeLi benchmarking tool. Based on medical and geographical information, users can benchmark hospitals on structural data and quality indicators at a medical condition level. Out of 351 indicators, only around 233 indicators are published (2016/2017), with 30% being process, 55% simple outcome, and 15% risk-adjusted outcome indicators. Moreover, all major sickness funds run individual transparency portals, such as the TK Klinikführer (hospital guide), or adapted versions of the WeLi portal. All sickness fund portals publish data from the mandatory quality assurance system, with the AOK portal supplemented by QSR results for 8 conditions and selected registry hospital participation information.

4QD provides a web-based, public benchmarking portal for its member hospitals for 30 conditions based on the public quality assurance scheme and other conditions with indicators based on administrative data. 4QD lists a composite quality score that comprises subscores for medical quality, patient safety, patient experience, admitting physician satisfaction, and ethical concerns. In contrast, IQM only requires its member hospitals to publish their standardized IQM report cards on their individual websites.

Similar to the *US News & World Report* Best Hospitals rankings, the Focus Klinikliste reports rankings of hospitals for 15 medical indications and departments, including breast and colorectal cancer, cardiology, and orthopedics. Rankings are based mainly on reputation and volume and not on outcome data, and they are available only in magazine or in PDF format. Skepticism toward public reporting within the clinical community[5] is still rather high, but political pressure and nonprofit activism[32] in favor of public reporting and its integration into care pathways are increasing.

*The Netherlands.* Based on a public mandate, the IGZ monitors hospital quality and investigates complaints and accidents, alerting providers of identified shortcomings in their quality and provision of care. IGZ compiles hospital-level annual reports for all hospitals and reports these back privately. The website, ziekenhuizentransparant.nl ("transparent hospital"), operated by the DHD, publishes IGZ quality indicators; however, no hospital benchmarking is possible. Both DHD and DICA provide hospital internal-quality monitoring. DICA further aims to make its PROMs accessible to patients via a public DICA patient portal. Each hospital is also mandated to publish condition-based HSMR rates on its own website. In contrast, some disease-specific initiatives provide more thorough and consolidated public access to quality results. For example, Meetbaar Beter publishes an annual report providing mortality, readmission, reoperation, chest pain, and quality-of-life metrics for heart conditions—including coronary artery disease, atrial fibrillation, and aortic valve disease—at the hospital level.

Hosted by the National Health Care Institute, KiesBeter.nl ("choose better") is one of two transparency portals. On a voluntary basis, hospitals submit IGZ and registry data, CQI, and their own quality and patient experience information. An aggregated individual provider report card is generated, which includes data on case volumes, number of wound infections, number of hospital-acquired infections, and mortality rates. Likewise, NVZ administers a website called nvz-kwaliteitsvenster.nl ("quality window") that enables patients to review hospital quality along 10 dimensions across 114 hospitals and rehabilitation centers. Dimensions include patient experience, employee satisfaction, hospital infections, and the HSMR. The portal shows only the global HSMR, but provides a link for detailed HSMR reports by condition on respective provider homepages. The NVZ portal offers a breadth of information; however, actual quality information by condition and direct provider benchmarking is limited.

*Sweden.* No mandatory national initiative exists in Sweden, but registers have increased public reporting of provider-level quality indicators in past years. For example, since 2008, SWEDEHEART has published annual reports providing hospital-level process and outcomes data. Many of the register reports, however, are targeted at researchers and expert clinicians, with information that is difficult for patients to process and data that are often not consistently provided at the hospital level, but instead at the regional level. Socialstyrelsen publishes an annual report entitled "Quality and Efficiency in Swedish Health Care" to make public health care more transparent and accountable. The report primarily showcases comparisons of Swedish regions, but hospital comparisons are also presented. In the 2012 report, 169 different quality and efficiency indicators across treatment areas were presented, of which 50 are also shown at the hospital level.[33]

The Vårdguiden ("guide to care") website publishes quality and access information at the provider level, but no quality indicator benchmarking is provided. The web portals Vardenisiffror.se ("care in numbers") and Omvard.se ("about care") also publish quality information, but the information is limited to mortality and complication rates at the regional level and patient experience results, respectively. While public and clinical support for public reporting is strong,[34] limited availability of risk-adjusted outcome indicators makes benchmarking of different hospitals more difficult than it is in other countries.

*United States.* The web portal Hospital Compare reports AHRQ indicator-based complication, readmission, and mortality rates for all 3,500 Medicare-certified acute care hospitals based on Medicare patient data. It also provides mandatory hospital-level patient survey data and structural information. Hospitals are benchmarked along the above criteria, with some verbal statements of whether a hospital is above, at, or below the national average. Further details, such as numeric rates and patient volumes, can be found by looking deeper into the data. In 2015, CMS for the first time published Hospital Compare Star Ratings based on patient experience as measured by the HCAHPS survey. For example, CMS awarded the top 5-star rating to 251 (7%) of the 3,500 hospitals assessed.

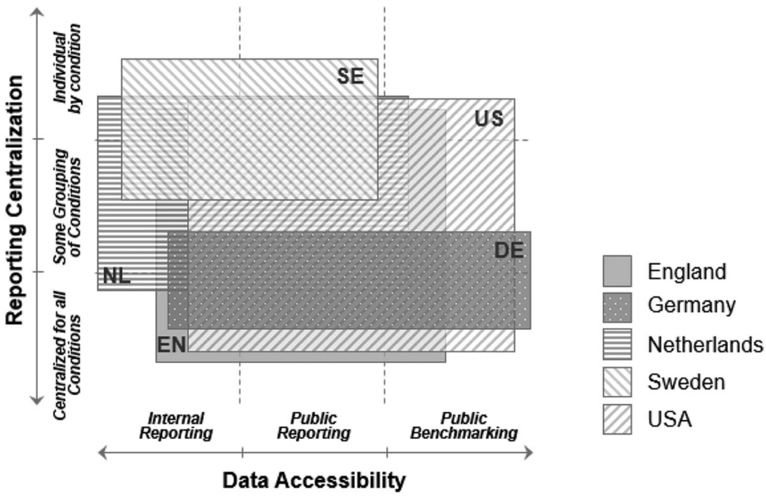
*US News* compiles the Best Hospitals national and regional rankings for 17 overarching medical specialties such as cancer, cardiology and heart surgery, gastroenterology, gynecology, and orthopedics. Among others, criteria include specialists' reputation over a 3-year period, 30-day risk-adjusted survival rates, 7 dimensions of patient safety, patient volume, and nursing intensity. All dimensions are aggregated in an overall ranking score. Further, the *US News* website allows a geographic search and report card–based benchmarking of hospitals. Healthgrades and Consumer Reports provide subscriber-based access to hospital report cards with composite indicators on mortality, complication, safety, infections, readmission, and adverse surgery events.

The Joint Commission Quality Check allows patients and physicians to search hospitals based on geographic information and accreditation and certification program, but not by medical condition. The portal lists different accredited and certified service lines and Joint Commission quality awards and gives access to accreditation and certification reports. Users can also download Joint Commission quality indicators in Excel format for all certified hospitals. The Leapfrog Group, based on its Hospital Survey results, assigns a Hospital Safety Score to more than 2,500 US hospitals annually. The score aggregates 28 publicly available, primarily process indicators.

Registries publish their results on single-condition, registry-specific websites or in annual reports. The CDC publishes annual ART Success Rates reports and allows specific clinic searches. The regional New York State and other state CABG and PCI registries and departments of health publish annual reports (in PDF format) based on the provider-level report card results. In contrast, the STS provides primarily private reporting to its member hospitals, with a voluntary online public reporting option. In addition, the STS teamed up with Consumer Reports to publish voluntary heart surgery report cards, including overall performance, survival, and complications. Public and professional attitudes toward public reporting are mixed, with health consumers and employers demanding hospital quality information and benchmarking to inform their provider choice, but there is continued provider skepticism.[35-37]

Table 1 (middle column) summarizes the quality reporting results and Figure 2 illustrates the different reporting strategies, with regards to degree of reporting centralization and data accessibility.

**Figure 2.** Two-dimensional Overview for the Framework Element Reporting Quality, Showing the Degree of Reporting Centralization and Data Accessibility




Source: Authors' analysis based on country research and expert interviews.

## *Rewarding Quality*

Measuring and reporting quality provides the basis for the introduction of QRPs, also termed pay-for-performance (P4P) or value-based payments (VBP). QRP programs generally reward providers for superior quality or penalize them for inferior quality. QRP programs can be implemented at a national or regional level and take mandatory or voluntary forms. They can comprise the entire reimbursement package (eg, bundled payments) or a share of reimbursement and base performance on structural, process, or outcome indicators.[38,39] They can focus on a single medical condition or aim to improve inpatient quality of care more broadly across conditions.[40] The QRP timing can range from immediate payment reduction to reduction only after continued, multiyear subpar performance or continued indirect financial pressure through bundled payments.

Different QRP models can play different roles depending on type of medical condition and procedure. Chronic diseases are more likely to see capitation payments with focus on outpatient care, possibly with bonus-malus payments for avoidance of hospitalization. In contrast, acute care consists of more discrete interventions and is thus more fitting for bundled payments or bonus-malus payments, which reward providers with a bonus per patient if quality performance is above average or meets predefined thresholds. If complications, readmissions, and outpatient care are covered in the bundled payment, providers also have financial incentives to coordinate care, improve outcomes, and avoid underprovision of care.[41] Furthermore, bundled payments are sometimes coupled with a bonus-malus component or close monitoring of quality standards to avoid negative effects on quality. Based on differences in measuring and reporting quality and varied degrees of P4P acceptance, the examined countries show a wide variation in scope and type of QRP application.

*England.* NHS England has implemented several QRPs, both in hospital care and in ambulatory general practitioner care (not discussed here). CQUIN allows NHS health commissioners to hold back 2.5% of the hospital payments contingent on quality; one-fifth is assessed according to 4 national metrics, with four-fifths assessed according to locally defined metrics. At the national level, CQUIN's clinical focus is on sepsis reduction, antimicrobial resistance, and medical care for the mentally ill. Indicators are mostly process indicators. Local measures focus primarily on overarching or population health areas and indicators, such as mental health or reduction in inappropriate emergency department visits and transfers.

Best practice tariffs (BPTs) pay hospitals higher-than-average fees for certain high-volume medical conditions depending on process quality achieved. Introduced in 2010, BPTs cover stroke, fragility hip fracture, cataract, knee fracture, interventional radiology, and day-case cholecystectomy. Each BPT consists of a base payment and a premium if predefined, condition-specific best practices are met. For example, for stroke care, the 2 process criteria are treatment in a stroke unit and immediate brain imaging. Participation in the BPT program is voluntary for hospitals, but clinicians often prefer BPT to CQUIN due to its clarity and standardization.[42] Early analysis of BPT-associated quality indicates mixed results. For example, process and outcome indicators improved for hip fracture, but no improvements were seen in the more complex treatment of stroke patients.[43]

Lastly, the voluntary program Advancing Quality in Northwest England provides incentives for process improvement in the care of 13 conditions—such as heart attack, dementia, and diabetes—to reduce mortality and hospital cost. Up to 4% of the hospital budget is withheld and only paid out if certain process goals are achieved.[40,44]

*Germany.* QRP components are present today in a few local integrated care and selective contracting programs. In Germany's integrated care pilots, sickness funds pay a cross-sectoral group of providers (eg, one or more hospitals, one or more specialists or general practitioners in the ambulatory sector, and/or a rehabilitation provider) a fixed fee per patient. The integrated care models have been primarily applied for chronic diseases and in 2011 included 6,400 contracts, 1.9 million patients, and €1.4 billion in reimbursements.[45] Also, AOK Hessen's stroke integrated care contract includes bonus-malus payments depending on 30-day mortality or 1-year readmission rates. In addition, AOK Hessen's selective hip and knee implantation contracts, for example, require baseline standards on QSR quality indicators.

More broadly, recent hospital legislation stipulates that from 2018 onward, hospitals will receive bonuses or penalties depending on their performance above or below national averages on newly developed quality indicators. Likewise, quality will be integrated as a key dimension in sickness funds, inpatient medical services, annual contracting, and state hospital capacity planning processes, with the potential to exclude hospitals from medical service lines or to close hospitals if quality performance is continuously substandard.[46]

*The Netherlands.* QRPs in the Netherlands are in their infancy. The Dutch DRG system—diagnosis treatment combinations (DBC)—covers inpatient care, with some minor components of outpatient care and some aspects of a bundled payment. After a significant reduction, the DBC system today consists of 4,400 multidisciplinary DBCs.[47] Health insurers are free to selectively contract with providers. For 70% of DBCs, they can negotiate prices. While competition is still mainly price-based, quality is increasingly taken into consideration.[48]

Bundled payments were introduced in 2010 for chronic disease areas, including diabetes, COPD, and vascular risk management.[49] A principal contracting agent, most often a general practitioner within a larger care group, provides the medical care in-house or subcontracts with provider partners, which include hospitals for the needed inpatient care. The principal agent has responsibility for care organization, delivery, and quality. Prices for the care group bundles are freely negotiated between care group leaders and insurers. Likewise, capitation payment pilots have been introduced in treatment of Parkinson's disease.

*Sweden.* In Sweden, QRPs are focused on bundled payments, which also have a bonus-malus aspect. In 2009, the Stockholm County Council (SCC) introduced payment bundles for cataract surgery and hip and knee replacements. The hip and knee OrthoChoice bundles cover treatment steps, diagnostics, surgery with postoperative care, implant, and follow-up checks. They include only patients from the lowest risk categories without comorbidities (ASA 1 and 2). Providers are responsible for nonacute complications of up to 5 years after surgery. A total of 3% of the €6,300 fee is paid depending on achievement of predefined outcome quality goals (eg, patient pain assessments). Patients are free to choose providers based on quality information provided on the SCC website.[50] Initial results have shown a positive effect for both outcomes and cost within the OrthoChoice program. Complication rates fell by almost 40% over the 2 years after OrthoChoice introduction and per-patient costs were reduced by 17% due to lower payments to providers.[51] At the same time, SCC analysis based on data from the national case-costing database also indicates that, between 2008 and 2011, cost per procedure for hip and knee replacements for SCC acute-care hospitals increased by 19% due to higher fixed costs per procedure. New, private operators perform the less complicated procedures while acute hospitals provide care for the more complex procedures.[52]

Based on the OrthoChoice experience, Stockholm County introduced additional, more advanced bundles. The spine bundle, for example, covers the entire care chain, payments are risk- and outcome-adjusted, and the provider covers cost of complications up to 2 years after treatment.[53] Other Swedish counties, guided by the joint value-based health care knowledge development and sharing project SVEUS, have introduced or are planning to introduce QRPs (eg, Skåne County in cataract and bypass surgery).

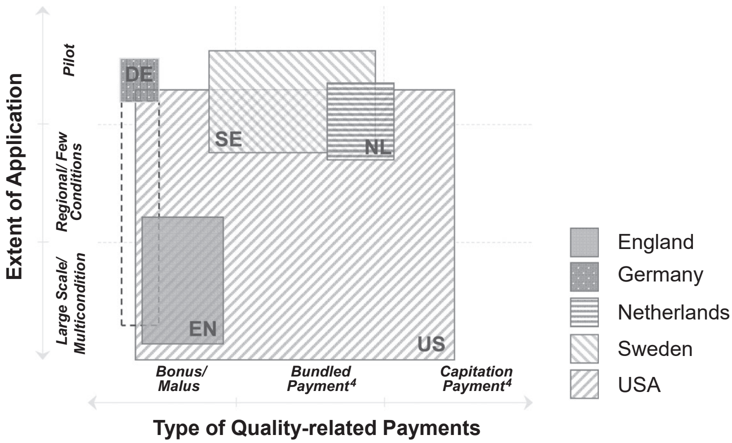
*United States.* Out of the 5 health systems examined, the United States has the most comprehensive QRP programs. On the one hand, Medicare has initiated large value-based payment programs including bonus-malus payments for good and bad quality and bundled payment elements. On the other hand, many large integrated-care delivery systems, compelled through the accountable care organization stipulations of the Affordable Care Act, introduced bundled payment programs or, in conjunction with their own health insurance plan, introduced population-based, capitated payment programs.

Medicare pays for value primarily through 3 programs, the Hospital Value-Based Purchasing (HVBP) program, the Hospital Readmission Reduction Program (HRRP), and the Hospital-Acquired Condition Reduction Program (HACRP). In fiscal year 2015-2016, the HVBP program adjusted 1.75% of Medicare DRG payments, or about $1.6 billion, according to whether hospitals performed below or above national average.[54,55] The HRRP reduces all condition-specific DRG payments for AMI, heart failure, pneumonia, COPD, and total hip and knee arthroplasty by an amount determined by the condition-specific, risk-adjusted, 30-day readmission rate. If the hospital performs in line with or better than expected based on its patient population, no penalty payment is deducted. The HACRP adjusts payments according to a hospital's performance with regards to hospital-acquired infections based on the AHRQ PSI #90 Composite Indicator. At the end of 2015, CMS also started to publish the HACRP indicators on Hospital Compare, linking its quality reporting and QRP components.[56] While all these measures are applied to a broad number of hospitals and a large number of patient cases, the actual value at risk for inferior performance is marginal.

In 2012, CMS commenced the Bundled Payments for Care Improvement (BPCI) initiative, which includes 4 new bundled payment models that each cover different elements of the care cycle. The most advanced bundled care model includes the acute period plus 30-day readmission for 48 pilot conditions. A BPCI assessment is difficult due to the short time horizon since implementation and the small number of participating hospitals. Some early results indicate spending reduction for postacute care, signaling a possible alignment of incentives across different providers, but inpatient episode spending between BPCI hospitals and non-BPCI hospitals remained equivalent.[57]

QRP innovation by large, private provider organizations is substantial. Geisinger Health System introduced a bundled payment model for elective coronary artery bypass graft surgery, which included preoperative, hospital, and postoperative care and resulted in significant savings for the health plan and increased profitability for Geisinger.[58] Likewise, several providers—including Cleveland Clinic, Mayo Clinic, and Virginia Mason—have engaged in bundled payment contracts with larger US employers, such as Boeing, Wal-Mart, and Lowes, for complex, elective care such as cardiac surgery.[59]

**Figure 3.** Two-dimensional Overview for the Framework Element Rewarding Quality, Showing the Extent of QRP Application and the Type of QRPs Employed




Source: Authors' analysis based on country research and expert interviews.

Accountable care organization agreements between insurers and providers, where providers care for certain patient populations within a shared risk payment model, have become more prevalent, with the Medicare Shared Savings Program the most widespread. Key characteristics are bonus payment for providers if cost objectives are met, quality indicators determining size of bonus payments, and free provider choice for patients.[60] Likewise, the Alternative Quality Contract between Blue Cross Blue Shield of Massachusetts and 11 provider groups includes global, capitated payments per patient and a time period to providers and P4P premiums if quality targets are met. The explicit control of and incentives for quality, with up to 10% of the total monthly member payment tied to quality goals, address some of the historic concern with global payment models.[61]

Table 1 (right column) summarizes the rewarding quality results; Figure 3 illustrates which type of QRPs have been employed and their degree of application.

## Discussion

This section summarizes results across countries and identifies similarities, differences, and unique features of country strategies for quality measurement, public reporting, and quality-related payments. Based on country results and observation of best practices and commonly faced challenges, 5 general policy levers to enhance quality accountability were identified. These are illustrated with examples from the examined countries.

### *Measuring Quality*

Approaches in quality measurement exhibit the most similarities across countries. The countries examined in this study use an overlapping set of structural, process, and outcome measures. Germany and the Netherlands cover the widest set of structural indicators, including physician staff levels and medical equipment. All countries measure case volume metrics, the structural indicator with the strongest link to quality,[62] with each debating the trade-offs between clinical and administrative data, since more detailed clinical data enable t-discharge events in data coverage, as a way to reduce the burden of measurement on providers. National governments play an essential role, mandating some quality measurement in all countries.

Yet differences in measurements have emerged. Quality indicators commonly cover only the inpatient hospital stay, but QSR indicators in Germany and the SHMI in England cover periods of 30 days or a longer time frame post-discharge. While PROMs are increasingly being collected in England, Sweden, and the Netherlands, this occurs only in a very limited fashion in Germany and the United States. Other unique features to country approaches are noteworthy: Germany has the most comprehensive set of risk-adjusted outcome indicators, based on both clinical and administrative data, while Sweden and England benefit from unique patient IDs that allow patient data integration across regions and systems. Sweden is also the only country with a comprehensive registry infrastructure.

### *Reporting Quality*

While quality reporting varies across countries, substantial overlap exists. Each country publishes at least some hospital-level quality information; however, analysts and patient groups have expressed some concern about differing results for hospitals across initiatives and the quality of data representation.[6] Some countries have web portals that allow comprehensive benchmarking of hospitals across a spectrum of quality indicators and treatment areas (eg, Germany and the United States); others are more hesitant on public hospital-level benchmarking (eg, the Netherlands and Sweden). Composite quality measures are also used differently across countries. England and the United States, for example, often focus on these while Germany, the Netherlands, and Sweden use fewer composite measures. Centralization and representations of reporting vary across countries. In Germany, one portal centrally reports data from mandatory measuring initiatives while in Sweden and the Netherlands reporting is diffused across registry platforms or annual reports. Due to the fragmented US health care system, several quality reporting (as well as measurement) initiatives have been introduced in parallel on national, state, or regional levels, covering public and/or private insurance programs and suffering from a lack of alignment. With Dr Foster Intelligence, England attempts to improve the visual representation of data for patients. At the same time, England has a multitude of different reporting systems (eg, NHS Choices and MyNHS) due to its open data policy; however, the plethora of platforms and information download options can potentially confuse patients. Importantly, physician and provider support for public reporting is strongest in Sweden.

### *Rewarding Quality*

For the 5 study countries, QRP strategies exhibit the greatest diversity. Similarities are limited to an awareness of QRP relevance and pilots for selected medical conditions. But countries differ in their reliance on process and/or outcomes measures. Process measures are primarily used, but indicator application differs across scope and type of QRPs employed. The United States introduced outcome measures into selected QRP programs; England still relies exclusively on process measures. Further, countries focus on different QRP elements. Sweden, for example, emphasizes bundled payments, while England employs large-scale bonus-malus payments, but has not implemented bundled payments. The United States focuses equally on bonus-malus and bundled payments. Germany operates integrated-care bundle pilots and plans to introduce a large-scale bonus-malus system by the end of 2018. The United States has the most advanced and comprehensive strategy vis-à-vis number of QRP elements employed, breadth of medical condition coverage, and financial impact on providers. Indeed, the United States is creating a market for quality information, where measurement, reporting, and quality improvements result in higher payments for providers. In response, private hospital chains (eg, Cleveland Clinic and Geisinger) are producing innovative measurement, reporting, and QRP approaches.

This assessment reveals diverse progress levels and policy approaches concerning hospital quality accountability and its 3 critical components: measurement, reporting, and incentivization of quality. Gathering distinct quality accountability approaches within a systematic framework facilitates international comparisons and allows the identification of general cross-country policy points. Countries less advanced in some aspects of quality accountability can adopt policies and best-practice approaches that have been found to be successful in other countries. Furthermore, the international discourse on quality of care measurement, reporting, and incentivization[7,16,40,63] provides the latest insights and the backdrop on which the country results are assessed and more general policy lessons are developed.

## *Policy Levers*

Based on the observed best practices and challenges and the widespread international debate around these issues, we can glean 5 policy levers that enhance breadth, depth, and value of quality accountability systems internationally. They are as follows:

1. Government support and standard setting for quality accountability
2. Balancing system centralization and decentralization
3. Provider involvement and support for quality accountability
4. Focusing on outcome indicators over process indicators
5. Quality reporting at hospital and disease levels

*Government Support and Standard Setting for Quality Accountability.* Health care is a public good. Combined with the complexity of health care markets in general and market inefficiencies in quality of care in particular, governments must assume responsibility for promoting quality accountability and implementing the necessary measures for its fulfillment. Similarly, health systems struggle under the diverging interests of multiple stakeholders, and governments should facilitate interest convergence on quality accountability and quality improvement objectives.[4]

In the United States, for example, the Affordable Care Act substantially accelerated the shift toward quality accountability, compelling CMS and AHRQ to develop condition-specific, patient-relevant, and provider-level indicators and requiring CMS to expand its QRP systems to create the first-ever market for quality information. With the Medicare and Medicaid programs, CMS has both strong bargaining power and a standard-setting role as the largest single payer of hospitals' annual budgets in the country. In the Netherlands, quality accountability is a political priority for the government, which named 2015 "The Year of Transparency for Dutch Health Care." Significant resources have been invested to develop and implement quality indicators for provider-level measurement of quality.[64] Based on consensus among health care stakeholders, the Netherlands Transparency Calendar delineates a mandatory phase-in of provider-level quality indicators across conditions. In Sweden, the previous central government assumed an important role in consolidating a quality accountability system. The national and regional governments have provided registers and other system elements with significant financial support, direction, and standardization pressure.[27]

Of all government quality accountability tasks, mandating and facilitating an integrated health system IT infrastructure and common information standards are particularly important. Virtually all aspects of health care delivery and payment is information intensive.[4] Collecting, transferring, reporting, and assessing quality data across the entire health care system places a large burden on health IT infrastructure and human resources, and it requires hardware and software interoperability and common standards for indicators, data usage and storage, and algorithms.

In the United States, the Health Information Technology for Economic and Clinical Health Act's incentive program has led to a substantial increase in electronic health record (EHR) adoption. In 2014, 75% of hospitals had an EHR system implemented, with almost 50% being comprehensive EHR systems.[65] While system interoperability between different providers is still challenging, widespread adoption of EHR systems has substantially facilitated electronic documentation and sharing of quality data.

In all 5 countries, government mandates, incentives, and standards have led to more quality accountability and spurred private sector initiatives. Government policy plays a central role in increasing quality accountability by setting software, information, and data system standards and by regulating and incentivizing public reporting and quality reimbursement. The degree of governmental standard setting varies across countries and even within a country's different quality accountability initiatives and health system components. However, the evidence shows that thorough, comparable, and lasting quality accountability requires a more comprehensive and potent government standard-setting role.

*Balancing System Centralization and Decentralization.* Health systems have substantially different degrees of centralization when it comes to quality accountability and factors such as decision making, technical infrastructure, and information standards. A more centralized approach facilitates national comparability and increases scope for provider quality competition and best-practice adoption. A less centralized, pluralistic approach allows more innovation and adoption at the local level and facilitates support from local providers.

A comparison between US and English quality accountability systems revealed the trade-offs between centralization and decentralization.[66] The United States has a multitude of local, regional, national, private, and public quality accountability systems that are often not integrated and have limited comparability. In England, in contrast, the NHS and associated regulatory agencies have substantial mandatory power and endow statutory powers to the HSCIC to ensure IT system integration and data comparability across the entire NHS system. Similarly, centralization of management structures in the US Veterans Health Administration was identified as one of the key enabling factors to quality improvement since the mid-1990s, especially in contrast to the more decentralized Medicare system.[67]

Quality measurement and reporting are currently undertaken by a range of organizations in each country. These organizations can focus on specific medical conditions or procedures such as registries or cover multiple medical conditions. A condition-based system provides specific insights for particular medical and patient requirements in different conditions and procedures; however, it also creates information silos and does not allow direct benchmarking of providers across diseases and procedures. Having numerous parallel initiatives that cover multiple or all conditions and procedures, such as the mandatory quality assurance system and the QSR initiative in Germany or Hospital Compare and the Joint Commission Quality Check in the United States, generates system competition and associated positive effects including system innovation, checks and balances, and more responsiveness. At the same time, this may lead to contradictory results for the same hospital and condition, [6,68] which conflicts with patients', physicians', and regulators' interest in one accepted information source. Moreover, having multiple, overlapping public reporting outlets, such as NHS Choices, MyNHS, registry websites, and data.gov.uk, creates a reporting plethora of sources that can be hard for patients to navigate.

Balancing the advantages of a consolidated and streamlined quality accountability system with the innovative power and checks and balances associated with a more decentralized system are crucial to establishing an accountability system that generates maximum transparency for patients, providers, and payers and consistently improves its quality measuring, reporting, and rewarding strategies.

*Provider Involvement in and Support for Quality Accountability.* Quality accountability benefits immensely from provider support and involvement.[4] Providers can use quality data to improve their services by identifying and implementing best practices. Physicians can explain reported data to their patients and can encourage them to use quality indicators to inform their health consumer decisions[7]; recent evidence showed that patients often require not just quality data but also advice and interpretation to make sense of the data.[69] Hospital management can institutionalize quality accountability by providing appropriate incentives and governance structures in their organizations.

Private provider groups—such as Helios Group in Germany or Cleveland Clinic, Partners Healthcare, and Geisinger in the United States—have started to emphasize quality accountability as a key business objective. Helios pursued a quality-focused expansion strategy in which they stressed the benefits of their quality measurement and reporting system for newly acquired hospitals and their patients.[70] The Helios strategy weakened the more quality accountability-skeptical position of the German Hospital Federation.

In Sweden, registers have mostly been founded by clinicians and remain independent and run by medical experts. A large majority of medical department heads support quality registers and are interested in expanding scope and usage.[34] Similarly, clinicians credit quality accountability with a positive influence on outcomes and have been in favor of expanding the range of indicators and treatment areas covered in public reports. Thus, physician support for quality accountability has facilitated growth and a more central role of quality registers in Swedish health care provision.

While the benefits from provider involvement are clear, in practice providers are often still hesitant about quality accountability. Physicians often claim that quality cannot be compared across providers and, further, that competition on quality is not in line with the traditional, collaborative mind-set of medicine.[4,5] As such, medical societies often constrain enhanced quality measurement and reporting. In 2013 and 2014, for example, German hospitals and regional hospital federations sued the AOK sickness fund to stop public reporting of QSR indicators. Providers criticized the lack of methodological clarity. The German provider initiatives IQM and 4QD recently combined under one umbrella organization, with one of its aims to oppose public hospital quality benchmarking portals.

Similarly, a recent survey of US hospital leaders' opinions on quality accountability showed a high degree of opposition, stating major concerns about validity and utility of quality measures and problems associated with public reporting.[36] A recent survey of US quality improvement experts also indicated provider opposition to public reporting as a major obstacle for quality accountability.[37] Physicians often exhibit a high degree of skepticism with regard to reliability and accuracy of measuring techniques and insufficient recognition of patients' rights to access and use quality performance data. Furthermore, the American Medical Association in June 2015 called for a slower timetable on introduction of HVBP schemes.

The aforementioned country examples demonstrate both the problem of continued provider opposition to and the power of provider endorsement of quality accountability. Getting providers and physicians to support and endorse quality accountability is a crucial step in encouraging patients to include publicly available quality benchmarks in their treatment decisions and provider choice and to focus competition among providers on quality.

*Focusing on Outcome Indicators Over Process Indicators.* As the results section demonstrates, in most countries, measuring, reporting, and rewarding quality still focuses on process indicators rather than on outcomes. While adherence to treatment protocols can contribute to positive outcomes, outcomes achieved by following the same protocol may still differ substantially[4] due to factors outside the protocol. Care may be exceptional in most process parts, but overall disastrously inadequate due to a vital error in one part.[71] The scientific evidence for the positive relationship of process indicators to health outcomes is limited.[72] In contrast, outcome indicators are of more intrinsic interest since they measure endpoints that directly matter to patients and providers, such as complication, reoperation, or mortality rates, or self-reported effects on mobility and pain levels. Outcome indicators, however, require sophisticated risk adjustment to ensure fair comparison between hospitals treating different patients and to avoid patient selection by providers.

Despite the relevance of outcome indicators and advances in risk adjustment, comprehensive reporting of risk-adjusted outcomes is scarce. The US health care system is heavily focused on process and patient experience measures. From the approximately 450 AHRQ, AMA, and Joint Commission indicators reviewed, more than 80% are process measures. Likewise, more than 80% of the 1,958 measures in the broader National Quality Measures Clearinghouse are process or patient experience measures and only 7% are actual outcome indicators.[73] The consensus-driven approval process of the NQF puts a premium on process measures favored by providers,[4] in part because process comparability between hospitals is more easily achieved.

The 2015 introduction of star ratings on Hospital Compare is based only on HCAHPS (patient experience survey) results and not, for example, on AHRQ outcome indicators. In 2015, the Joint Commission suspended the collection of outcome indicators in several conditions and relaxed its outcome reporting requirements.[28] Moreover, the *US News* Best Hospitals rankings are almost entirely based on reputation and not, as stated, on a mix of indicators including outcomes.[74]

Similarly, the Dutch quality accountability system focuses on structural and process indicators. The Court of Audits has criticized Dutch quality indicators, developed through the Zichtbare Zorg program, for their lack of meaningful, "practical value."[64] Similarly problematic, Swedish registers have so far developed only few risk-adjusted outcome indicators, which are important for direct and fair provider-level benchmarks. In England, QRP systems CQUIN and BPT are exclusively focused on process measures.

While process measures and patient experience measures are often easier to agree upon in a consensus-based model, success in health care delivery should be measured in better patient results and not compliance with predefined process standards or best amenities. Health outcomes are the ultimate validators of quality of hospital care.[71]

*Quality Reporting at Hospital and Disease Levels.* With regards to breadth, quality data can be reported at the regional, hospital group or trust, individual hospital, or physician level. With regards to depth, reporting can aggregate all diseases and/or procedures in a hospital, group diseases or procedures (eg, cardiovascular), report individual medical conditions or procedures (eg, CABG), or report outcomes for individual physicians.

Quality reporting needs to be meaningful and actionable for patients, admitting physicians, and treating physicians. *Meaningful* implies differentiation between diseases and procedures with different complications and risk profiles. *Actionable* implies an ability to differentiate between the relevant care units so patients and physicians can choose the best providers at the disease or medical department level, and so providers can assess their own performance relative to national and competitor benchmarks. Composite measures aggregate individual quality indicators into summary measures and might provide an ostensibly clearer picture of overall performance. However, they also average out performance across departments, reduce visibility for patients mostly interested in a single procedure, and assume a weighting of underlying indicators, which conflicts with the many different preferences across patients[75] and admitting physicians.

Furthermore, the reporting level needs to take into account statistical constraints of indicators, mainly the risk-adjustment capability and the need for large case volumes. Both are better served at the hospital as opposed to the physician level. Quality measures need to be comparable between hospitals to ensure an apples-to-apples comparison. This implies risk adjustment, but also demands standardized and validated indicators and data inputs.

In England, many of the indicators collected and reported indicate quality at the hospital trust level. While most trusts are composed of only one large hospital, trust consolidation is grouping hospitals together. Thus, trust-level reporting averages different hospital-level quality and obscures outcomes at the relevant hospital unit of analysis. NHS England is also facing sectionalism among NHS trusts and Clinical Commissioning Groups, which often compose and report their own local quality indicators on top of national ones.

While the current Swedish outcome accountability system facilitates regional health system competition around outcomes, reporting of outcomes at a hospital level occurs only in isolated circumstances. This limits both hospital-level competition on outcomes and patients' ability to source hospital quality information in an effortless and standardized fashion. The relevant level of quality-of-care competition is not the county, but the provider level.

Balancing the need for meaningful indicators against the statistical and reporting requirements of sound indicators is best achieved at the hospital level for diseases and procedures. Quality performance has to be displayed at the relevant unit of production, which is the disease area or procedure within a hospital.

## Conclusion

This study shows that measuring, reporting, and rewarding quality has evolved substantially. It also identifies considerable differences and varied progress levels in the quality accountability strategies of the countries examined. Each country pursues its own policies, often leading to increased efforts and costs, uncertainty, and worse policy outcomes than would happen with more cross-country comparison and learning.

Today's quality accountability systems help to identify substantial quality variation between health providers and across medical conditions. However, they are often insufficient to truly facilitate provider competition around quality. They report quality not at the hospital but at a more aggregated level, focus only on few conditions, and often use process indicators that are less meaningful to patients. Furthermore, due to provider resistance, easily accessible hospital quality benchmarking is not always available. Advancing the 5 aforementioned policy levers will help to facilitate more comprehensive quality accountability and quality competition among providers.

This article aims to enhance best-practice learning and data comparability across countries. Similar objectives are pursued by international institutions like the Organisation for Economic Co-operation and Development (OECD) and the International Consortium for Health Outcomes Measurement (ICHOM). Through the Health Care Quality Indicators project, the OECD measures and compares the quality of service provision at a system level across countries and fosters a harmonization of outcomes data across member states. The OECD strives to stimulate cross-national learning, in particular with regards to the effects of certain health system factors on quality of care.[76] ICHOM develops international outcome indicators across medical conditions to facilitate best-practice identification and adoption across geographies.[77] Research institutes such as the Commonwealth Fund, the Rand Corporation, or the European Observatory on Health Systems and Policies support evidence-based health policymaking through comprehensive analysis and comparison of health system dynamics across countries.

By learning from strategies and experiences in other countries, health care stakeholders can apply best practices and avoid others' mistakes. The government takes a central role in establishing standards and incentives for quality accountability and health IT system integration. The benefits and costs of centralization and decentralization need to be balanced to ensure both national comparability and continued innovation. Health systems need to shift toward quality measuring, reporting, and rewarding based more on outcomes and less on process measures. Providers need to become proponents of quality accountability, especially within their organizations and to their patients. Quality measurement and reporting should focus on the hospital and disease or procedure level to enable meaningful and actionable choices for patients. Operating these policy levers can enhance quality accountability within each of the 5 examined countries and facilitate integration, comparability, and best-practice sharing among countries to achieve higher performing health systems.

## References

1. Chung SC, Sundström J, Gale CP, et al. Comparison of hospital variation in acute myocardial infarction care and outcome between Sweden and United Kingdom: population based cohort study using nationwide clinical registries. *BMJ*. 2015;351:h3913. doi:10.1136/bmj.h3913.
2. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med*. 2009;361(14):1368-1375. doi:10.1056/NEJMsa0903048.
3. Wouters MW, Siesling S, Jansen-Landheer ML, et al. Variation in treatment and outcome in patients with non–small cell lung cancer by region, hospital type and volume in the Netherlands. *Eur*

*J Surg Oncol*. 2010;36 (Suppl 1):S83-S92. doi:10.1016/j.ejso.2010.06.020.

4. Porter ME, Teisberg EO. *Redefining Health Care: Creating Value-Based Competition on Results*. Boston, MA: Harvard Business Review Press; 2006.
5. Costa SD. Qualitätsmanagement im Krankenhaus—Nicht zum Nutzen der Patienten: Qualitätsmanagement ist für die Medizin ähnlich nützlich wie Ornithologie für die Vögel. *Dtsch Ärztebl*. 2014;111(38):1344-1345.
6. Austin JM, Jha AK, Romano PS, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Aff (Millwood)*. 2015;34(3):423-430. doi:10.1377/hlthaff.2014.0201.
7. Kumpunen S, Trigg L, Rodrigues R. *Public Reporting in Health and Long-term Care to Facilitate Provider Choice*; 2014. Policy Summary 13. Geneva, Switzerland: World Health Organization; 2014. http://www.euro.who.int/en/about-us/partners/observatory/publications/policy-briefs-and-summaries/public-reporting-in-health-and-long-term-care-to-facilitate-provider-choice. Accessed December 10, 2016.
8. Cacace M, Ettelt S, Brereton L, Pedersen JS, Nolte E. *How Health Systems Make Available Information on Service Providers: Experience in Seven Countries*. Santa Monica, CA: Rand Corporation; 2011. http://www.rand.org/pubs/technical_reports/TR887.html. Accessed December 10, 2016.
9. Himmelstein DU, Jun M, Busse R, et al. A comparison of hospital administrative costs in eight nations: US costs exceed all others by far. *Health Aff (Millwood)*. 2014;33(9):1586-1594. doi:10.1377/hlthaff.2013.1327.
10. Mossialos E, Wenzl M, Osborn R, Sarnak D, eds. *2015 International Profiles of Health Care Systems*. New York, NY: Commonwealth Fund; 2016. http://www.commonwealthfund.org/~/media/files/publications/fund-report/2016/jan/1857_mossialos_intl_profiles_2015_v7.pdf. Accessed December 10, 2016.
11. Shaw C, Bruneau C, Kutryba B, de Jongh G, Suñol R. Towards hospital standardization in Europe. *Int J Qual Health Care*. 2010;22(4):244-249. doi:10.1093/intqhc/mzq030.
12. Davis K. Slowing the growth of health care costs—learning from international experience. *N Engl J Med*. 2008;359(17):1751-1755.
13. Cordasev H, Björnberg A, Hjertqvist O. *Cross Border Care EU—How to Choose the Best Hospital? A Study of Hospital Information Portals in Five European Countries*. Täby, Sweden: Health Consumer Powerhouse; 2010. http://www.healthpowerhouse.com/files/

HCP-EP-seminar_report%20HC-2.pdf. Accessed December 10, 2016.

14. Abrams M, Nuzum R, Zezza M, Ryan J, Kiszla J, Guterman S. The Affordable Care Act's Payment and Delivery System Reforms: A Progress Report at Five Years. http://www.commonwealthfund.org/publications/issue-briefs/2015/may/aca-payment-and-delivery-system-reforms-at-5-years. Published May 7, 2015. Accessed December 10, 2016.
15. Schreyögg J, Bäuml M, Krämer J, Dette T, Busse R, Geissler A. Forschungsauftrag zur Mengenentwicklung nach § 17b Abs. 9 KHG: Endbericht; 2014. Cited by: "Forschungsbericht zur Mengenentwicklung im Krankenhaus" veröffentlicht [press release]. Berlin, Germany: GKV-Spitzenverband; July 10, 2014. https://www.gkv-spitzenverband.de/presse/pressemitteilungen_und_statements/pressemitteilung_151872.jsp. Accessed December 10, 2016.
16. Shwartz M, Restuccia JD, Rosen AK. Composite measures of health care provider performance: a description of approaches. *Milbank Q*. 2015;93(4):788-825. doi:10.1111/1468-0009.12165.
17. Figueroa JF, Tsugawa Y, Zheng J, Orav EJ, Jha AK. Association between the value-based purchasing pay for performance program and patient mortality in US hospitals: observational study. *BMJ*. 2016;353:i2214. doi:10.1136/bmj.i2214.
18. Provisional Monthly Patient Reported Outcome Measures (PROMs) in England: April 2014 to March 2015. August 2015 release. NHS Digital website. http://content.digital.nhs.uk/catalogue/PUB17877. Published August 13, 2015. Accessed December 10, 2016.
19. Busse R, Nimptsch U, Mansky T. Measuring, monitoring, and managing quality in Germany's hospitals. *Health Aff (Millwood)*. 2009;28(2):w294-w304. doi:10.1377/hlthaff.28.2.w294.
20. AQUA-Institut GmbH. *Qualitätsreport 2013*. Göttingen, Germany: AQUA-Institut GmbH; 2014. https://sqg.de/front_content.php?idart=1112. Accessed December 10, 2016.
21. Mansky T, Nimptsch U, Cools A, Hellerhoff F. *G-IQI—German Inpatient Quality Indicators: Version 4.0*. 4th ed. Berlin, Germany: Universitätsverlag der TU Berlin; 2013. https://www.seqmgw.tu-berlin.de/menue/inpatient_quality_indicators/g_iqi_definitionshandbuecher. Accessed December 10, 2016.
22. Schafer W, Kroneman M, Boerma W, et al. The Netherlands: health system review. *Health Syst Transit*. 2010;12(1):v-xxvii, 1-228.

23. OECD. *OECD Reviews of Health Care Quality: Sweden 2013*. Paris, France: OECD Publishing; 2013. doi:10.1787/9789264204799-en. Accessed December 10, 2016.
24. Tillberg J, ed. *SWEDEHEART Annual Report 2014*. Huddinge, Sweden: Karolinska University Hospital; 2015. https://www.researchgate.net/publication/285400347_SWEDEHEART_Annual_Report_2014. Accessed December 10, 2016.
25. Emilsson L, Lindahl B, Köster M, Lambe M, Ludvigsson JF. Review of 103 Swedish healthcare quality registries. *J Intern Med*. 2015;277(1):94-136. doi:10.1111/joim.12303.
26. Nationella Kvalitetsregister. *Nulägesrapport Våren 2015*. Stockholm, Sweden: Nationella Kvalitetsregister; 2015. http://www.kvalitetsregister.se/download/18.208f26d6152bccdb86d73677/1455963126410/Nulagesrapport-Nationella-Kvalitetsregister-2015.pdf. Accessed December 10, 2016.
27. Nationella Kvalitetsregister. Överenskommelse om satsning på Nationella Kvalitetsregister, 2012-2016. http://kvalitetsregister.se/download/18.208f26d6152bccdb86d722fb/1455952754031/Overenskommelse-SKL-staten-om-satsning-pa-Nationella-Kvalitetsregister-2012-2016.pdf. Accesed December 20, 2016.
28. Joint Commission. Joint Commission Measure Sets Effective January 1, 2015. http://www.jointcommission.org/assets/1/6/TJC_Measures_2015__11_15.pdf. Updated January 9, 2015. Accessed December 10, 2016.
29. Dranove D, Kessler D, McClellan M, Satterthwaite M. Is more information better? The effects of "report cards" on health care providers. *J Political Econ*. 2003;111(3):555-588.
30. Francis R. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Summary*. HC 947. London, England: Stationery Office; 2013.
31. Blitz R. Hunt calls for more transparency in NHS: Minister wants to reverse "targets and performance" culture. *Financial Times*. https://www.ft.com/content/52e58c82-5818-11e2-90c6-00144feab49a. Updated January 6, 2013. Accessed December 10, 2016.
32. Schwenk U, Schmidt-Kaehler S. Public Reporting: Transparenz über Gesundheitsanbieter erhöht Qualität der Versorgung. *Spotlight Gesundheit*. 2016;1. https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/spotlight-gesundheit-012016. Accessed December 10, 2016.
33. Glenngard AH. The Swedish health care system, 2014. In: Mossialos E, Wenzl M, Osborn R, Sarnak D, eds. *2015 International Profiles of Health Care Systems*. New York, NY: Commonwealth Fund; 2016. http://www.commonwealthfund.

org/~/media/files/publications/fund-report/2016/jan/1857_mossialos_intl_profiles_2015_v7.pdf. Accessed December 12, 2016.

34. Åsgård S. Verksamhetschefer positiva till kvalitetsregister i vården. *Dagens Medicin*. July 6, 2011. http://itivarden.idg.se/2.2898/1.394747/verksamhetschefer-positiva-till-kvalitetsregister-i-varden. Accessed December 12, 2016.
35. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. *JAMA*. 2000;283(14):1866-1874.
36. Goff SL, Lagu T, Pekow PS, et al. A qualitative analysis of hospital leaders' opinions about publicly reported measures of health care quality. *Jt Comm J Qual Patient Saf*. 2015;41(4):169-176.
37. Sinaiko AD, Eastman D, Rosenthal MB. How report cards on physicians, physician groups, and hospitals can have greater impact on consumer choices. *Health Aff (Millwood)*. 2012;31(3):602-611. doi:10.1377/hlthaff.2011.1197.
38. Miller HD. From volume to value: better ways to pay for health care. *Health Aff (Millwood)*. 2009;28(5):1418-1428. doi:10.1377/hlthaff.28.5.1418.
39. Ogundeji YK, Bland JM, Sheldon TA. The effectiveness of payment for performance in health care: a meta-analysis and exploration of variation in outcomes. *Health Policy*. 2016;120(10):1141-1150. doi:10.1016/j.healthpol.2016.09.002.
40. Milstein R, Schreyoegg J. Pay for performance in the inpatient sector: a review of 34 P4P programs in 14 OECD countries. *Health Policy*. 2016;120(10):1125-1140. doi:10.1016/j.healthpol.2016.08.009.
41. Shih T, Chen LM, Nallamothu BK. Will bundled payments change health care? Examining the evidence thus far in cardiovascular care. *Circulation*. 2015;131(24):2151-2158. doi:10.1161/CIRCULATIONAHA.114.010393.
42. Marshall L, Charlesworth A, Hurst J. *The NHS Payment System: Evolving Policy and Emerging Evidence*. London, England: Nuffield Trust; 2014. http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/140220_nhs_payment_research_report.pdf. Accessed December 12, 2016.
43. McDonald R, Zaidi S, Todd S, et al. A qualitative and quantitative evaluation of the introduction of Best Practice Tariffs: An evaluation report commissioned by the Department of Health. 2012.
44. McDonald R, Boaden R, Roland M, et al. A qualitative and quantitative evaluation of the Advancing Quality pay-for-performance

programme in the NHS North West. *Health Serv Delivery Res.* 2015;23(3). doi:10.3310/hsdr03230.
45. Busse R, Blümel M. Germany: health system review. *Health Syst Transit.* 2014;16(2):1-296, xxi.
46. Gesetz zur Reform der Strukturen der Krankenhausversorgung (Krankenhausstrukturgesetz—KHSG). *Bundesgesetzblatt.* 2015;I(51):2229-2253.
47. Wammes J, Jeurissen P, Westert G. *The Dutch Health System, 2015*. Utrecht, The Netherlands: NVAG; 2016. http://www.nvag.nl/afbeeldingen/2015/nscholing/Netherlands%20Health%20Care%20System%202014%20(PDF).pdf. Accessed December 12, 2016.
48. Busse R, Geissler A, Quentin W, Wiley M, eds. *Diagnosis-Related Groups in Europe: Moving Towards Transparency, Efficiency and Quality in Hospitals*. Maidenhead: Open University Press; 2011. European Observatory on Health Systems and Policies Series.
49. Struijs JN, Baan CA. Integrating care through bundled payments–lessons from the Netherlands. *N Engl J Med.* 2011;364(11):990-991. doi:10.1056/NEJMp1011849.
50. Porter ME, Marks CM, Landman ZC. OrthoChoice: bundled payments in the county of Stockholm (A). Case Study 9-714-514. *Harvard Business School*. Published June 4, 2014. Accessed December 12, 2016.
51. Porter ME, Marks CM, Landman ZC. Ortho Choice: bundled payments in the county of Stockholm (B). Case Study 9-714-515. *Harvard Business School*. Published June 4, 2014.
52. Hagström S, Karlsson M. Vårdval—hurska öppenvården beskrivas och ersättas i framtiden? 2013. KPP-DRG, Stockholms Läns Landsting. http://www.vardgivarguiden.se/avtaluppdrag/vardinformatik/statistik-och-rapporter/diagnosgranskningar/vardval—hur-ska-oppenvarden-beskrivas-och-ersattas-i-framtiden/. Page 62. Accessed December 20, 2016
53. Hälso- och sjukvardsförvaltningen. *Ryggkirurgi: Förfrågningsunderlag Enligt lov Vårdval*. Stockholm, Sweden: Hälso- och sjukvardsförvaltningen; January 9, 2013. http://www.sll.se/Global/Politik/Politiska-organ/Halso-och-sjukvardsnamnden/2013/2013-03-18/bilaga-ryggkirurgi-reviderad130311.pdf. Accessed December 12, 2016.
54. Hospital value-based purchasing: overview. Centers for Medicare & Medicaid Services website. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html?redirect=/hospital-value-based-purchasing. Updated October 20, 2015. Accessed December 12, 2016.

55. National provider call: hospital value-based purchasing (VBP) program. Centers for Medicare & Medicaid Services website. http://docplayer.net/4565882-National-provider-call-hospital-value-based-purchasing-vbp-program.html. Published April 29, 2014. Accessed December 12, 2016.
56. Overview: hospital-acquired condition (HAC) reduction program. QualityNet website. https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228774189166. Accessed December 12, 2016.
57. Tsai TC, Joynt KE, Wild RC, Orav EJ, Jha AK. Medicare's bundled payment initiative: most hospitals are focused on a few high-volume conditions. *Health Aff (Millwood)*. 2015;34(3):371-380. doi:10.1377/hlthaff.2014.0900.
58. Brennan KF. Successful case studies in accountable care. Paper presented at: ACO Congress; October 25-27, 2010; Los Angeles, CA.
59. Porter ME, Kaplan RS. How to pay for health care. *Harvard Bus Rev*. 2016:88-100. https://hbr.org/2016/07/how-to-pay-for-health-care. Accessed December 12, 2016.
60. Berenson RA, Burton RA. Next steps for ACOs: will this new approach to health care delivery live up to the dual promises of reducing costs and improving quality of care? Health Affairs/Robert Wood Johnson Foundation: Health Policy Brief. Published January 31, 2012.
61. Chernew ME, Mechanic RE, Landon BE, Safran DG. Private-payer innovation in Massachusetts: the "alternative quality contract." *Health Aff (Millwood)*. 2011;30(1):51-61.
62. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346(15):1128-1137.
63. Porter ME. What is value in health care? *N Engl J Med*. 2010;363(26):2477-2481.
64. Oldenburg C, Van den Berg J, Leistikow I. All aboard, getting nationwide indicators on the rails: collaborative governance as a strategy for developing effective national quality indicators for hospital care. *Br Med J Outcomes*. 2015.
65. Adler-Milstein J, DesRoches CM, Kralovec P, et al. Electronic health record adoption in US hospitals: progress continues, but challenges persist. *Health Aff (Millwood)*. 2015;34(12):2174-2180. doi:10.1377/hlthaff.2015.0992.
66. Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. *Milbank Q*. 2001;79(2):281-315.

67. Jha AK, Perlin JB, Kizer KW, Dudley RA. Effect of the transformation of the Veterans Affairs health care system on the quality of care. *N Engl J Med*. 2003;348(22):2218-2227. doi:10.1056/NEJMsa021899.
68. Rothberg MB, Morsi E, Benjamin EM, Pekow PS, Lindenauer PK. Choosing the best hospital: the limitations of public quality reporting. *Health Aff (Millwood)*. 2008;27(6):1680-1687. doi:10.1377/hlthaff.27.6.1680.
69. Boyle D. Barriers to choice: a review of public services and the government's response. GOV.UK website. https://www.gov.uk/government/publications/barriers-to-choice-public-services-review. Published January 24, 2013. Accessed December 12, 2016.
70. Richter-Kuhlmann E, Flintrop J. Das Gespräch mit Prof. Dr. med. Ralf Kuhlen, medizinischer Konzerngeschäftsführer der Helios Kliniken GmbH, Berlin: Der Skepsis Qualität entgegensetzen. *Dtsch Ärztebl*. 2014;111:35-36.
71. Donabedian A. Evaluating the quality of medical care, 1966. *Milbank Q*. 2005;83(4):691-729. doi:10.1111/j.1468-0009.2005.00397.x.
72. Copnell B, Hagger V, Wilson SG, Evans SM, Sprivulis PC, Cameron PA. Measuring the quality of hospital care: an inventory of indicators. *Intern Med J*. 2009;39(6):352-360. doi:10.1111/j.1445-5994.2009.01961.x.
73. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med*. 2016;374(6):504-506. doi:10.1056/NEJMp1511701.
74. Sehgal AR. The role of reputation in *US News & World Report's* rankings of the top 50 American hospitals. *Ann Intern Med*. 2010;152(8):521-525. doi:10.7326/0003-4819-152-8-201004200-00009.
75. Dixon A, Robertson R, Appleby J, Burge P, Devlin N, Magee H. *Patient Choice: How Patients Choose and How Providers Respond*. London, England: The King's Fund; 2010.
76. Kelley E, Hurst J. Health Care Quality Indicators Project Conceptual Framework Paper. OECD Health Working Papers. Paris, France: OECD Publication Service; March, 9 2006. http://www.oecd.org/health/health-systems/36262363.pdf. Accessed December 12, 2016.
77. Porter M, Larsson S, Ingvar M. A new initiative to put outcomes measurement at the center of health reform. Health Affairs blog. October 31, 2012. http://healthaffairs.org/blog/2012/10/31/a-new-initiative-to-put-outcomes-measurement-at-the-center-of-health-reform. Accessed December 12, 2016.

*Funding/Support:* Christoph Pross is supported by a general PhD Scholarship from the Konrad Adenauer Foundation.

*Conflict of Interest Disclosures*: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. No disclosures were reported.

*Acknowledgments*: We would like to thank our country experts for their insights and feedback: David Himmelstein (City University of New York), Kevin Boziac (Dell Medical School), Neil Shah (University Medical Center Brackenridge), Natasha Taylor (BCG), Caleb Stowell (ICHOM), and Tom Rice (University of California, Los Angeles) in the USA; Irini Papanicolas (LSE Health), Vaeena Raleigh (Kings Fund), Roger Taylor (Dr Foster Intelligence), Thomas Kelley (ICHOM), and Mark Ferreira (Alliance Medical) in England; Johan van Manen (Dutch Health Authority), Pieter de Bey (BCG), Evout Van Ginneken (European Observatory on Health Systems and Policies), and Lisa Haverkorn (Dutch Institute for Clinical Auditing) in the Netherlands; Lisbeth Serdén, Birgitta Lindelius, and Anna Dörvan (National Board of Health and Welfare), Hanna Sjöberg (Swedish Agency for Health and Care Services Analysis), Charlotta Gyland (Skane University Hospital), and Gabriel Osterdahl (BCG) in Sweden; and Jens Deerberg-Wittram (BCG) in Germany. Final results and discussion are work of the authors only.

*Address correspondence to:* Christoph Pross, Berlin University of Technology, Dept. of Health Care Management, Administrative office H80, Str. des 17 Juni 135, 10623 Berlin, Germany (email: christoph.pross@campus.tu-berlin.de).

# Appendices

**Table A1.** List of Experts Interviewed by Country (Name, Position, and Institution)

| Country | Name | Position | Institution |
|---|---|---|---|
| US | David Himmelstein | Professor | CUNY School of Public Health at Hunter College |
| US | Tom Rice | Distinguished professor | Department of Health Policy and Management, University of California Los Angeles |
| US | Kevin Boziac | Chair of surgery & perioperative care | Dell Medical School, University of Texas at Austin |
| US | Caleb Stowell | VP of standardization and business development and partnerships | International Consortium for Health Outcomes Measurement (ICHOM) |
| US | Natasha Taylor | Principal | The Boston Consulting Group (BCG) |
| US | Neil Shah | Attending physician | University Medical Center Brackenridge |
| EN | Irini Papanicolas | Assistant professor | LSE Health, London School of Economics |
| EN | Vaeena Raleigh | Senior fellow policy | The King´s Fund |
| EN | Roger Taylor | Strategy director, co-founder | Dr Foster |
| EN | Thomas Kelley | VP of business development and partnerships, EMEA | ICHOM |
| EN | Mark Ferreira | Medical director | Alliance Medical |
| NL | Johan van Manen | Senior policy adviser | Dutch Health Authority |
| NL | Evout van Ginneken | Senior researcher | European Observatory on Health Systems and Policies, Berlin University of Technology |

*Continued*

**Table A1.** *Continued*

| Country | Name | Position | Institution |
|---|---|---|---|
| NL | Pieter de Bey | Principal | BCG |
| NL | Lisa Haverkorn | Researcher | Dutch Institute for Clinical Auditing |
| SE | Lisbeth Serdén | Project manager | Swedish National Board of Health and Welfare |
| SE | Birgitta Lindelius | Researcher | Swedish National Board of Health and Welfare |
| SE | Anna Dörvan | Researcher | Swedish National Board of Health and Welfare |
| SE | Hanna Sjöberg | Program director | Swedish Agency for Health and Care Services Analysis |
| SE | Charlotta Gyland | Activity controller VBHC | Skane University Hospital |
| SE | Gabriel Osterdahl | Senior knowledge expert VBHC | BCG |
| DE | Jens Deerberg-Wittram | Executive director | BCG |

*Abbreviations:* DE, Germany; EN, England; NL, the Netherlands; SE, Sweden; US, United States.

**Table A2.** List of Abbreviations

| | |
|---|---|
| AMI | Acute myocardial infarction |
| CABG | Coronary artery bypass graft |
| COPD | Chronic obstructive pulmonary disease |
| DRG | Diagnosis-related group |
| EQ-5D | EuroQol five dimensions questionnaire |
| ICHOM | International Consortium for Health Outcomes Measurement |
| QRP | Quality-related payment |
| PCI | Percutaneous coronary intervention |
| PREM | Patient-reported experience measures |
| PROM | Patient-reported outcome measure |
| **England** | |
| BPT | Best practice tariff |
| CQUIN | Commissioning for Quality and Innovation |
| HQIP | Healthcare Quality Improvement Partnership |
| HSCIC | Health and Social Care Information Centre |
| NJR | National Joint Registry |
| SHMI | Summary Hospital-level Mortality Indicator |
| **Germany** | |
| AOK | General Health Insurance (Allgemeine Ortskrankenkassen) |
| G-BA | Federal Joint Committee (Gemeinsamer Bundesausschuss) |
| IQM | Initiative Qualitätsmedizin |
| IQTIG | Institute for Quality and Transparency in Health Care |
| QSR | Qualitätssicherung mit Routinedaten |
| PEQ | Patient experience questionaire |
| 4QD | Qualitätskliniken.de |
| WeLi | Weisse Liste.de |
| **Netherlands** | |
| CQI | Consumer Quality Index |
| DBC | Diagnosis Treatment Combination |
| DICA | Dutch Institute for Clinical Auditing |
| DHD | Dutch Hospital Data foundation |
| HSMR | Hospital standardized mortality ratio |
| IGZ | Health Care Inspectorate |
| NVZ | Dutch Hospital Association |

*Continued*

Table A2. *Continued*

| | |
|---|---|
| NZa | Dutch Healthcare Authority |
| **Sweden** | |
| SCC | Stockholm County Council |
| SVEUS | Swedish national collaboration for value-based reimbursement and monitoring of health care |
| SWEDEHEART | Swedish Web-system for Enhancement and Development ofEvidence-based care in Heart disease Evaluated According to Recommended Therapies |
| SWEDVASC | National Vascular Surgery Register |
| **United States** | |
| ACA | Patient Protection and Affordable Care Act |
| ACO | Accountable Care Organization |
| AHRQ | Agency for Healthcare Research and Quality |
| AMA | American Medical Association |
| ART | Assisted reproductive technology |
| BPCI | Bundled Payment for Care Improvement |
| CDC | Centers for Disease Control and Prevention |
| CMS | Centers for Medicare & Medicaid Services |
| EHR | Electronic Health Record |
| HACRP | Hospital-Acquired Condition Reduction Program |
| HCAHPS | Hospital Consumer Assessment of Healthcare Providers and Systems survey |
| HRRP | Hospital Readmission Reduction Program |
| HVBP | Hospital Value-Based Purchasing |
| IQI | Inpatient quality indicator |
| NQF | National Quality Forum |
| PSI | Patient safety indicator |
| STS | Society for Thoracic Surgeons |

# Kapitel 3: Qualitätsunterschiede zwischen deutschen Krankenhäusern

Christoph Pross, Reinhard Busse, Alexander Geissler

Postprint, erschienen im August 2017 als: Pross C, Busse R, Geissler A. Hospital quality variation matters: a time-trend and cross-section analysis outcomes in German hospitals from 2006-2014. *Health Policy*. 2017;121(8):842-852. https://doi.org/10.1016/j.healthpol.2017.06.009

**Hintergrund**: Seit 2006 gibt es in Deutschland im Rahmen der externen Qualitätssicherung Qualitätsberichte für Krankenhäuser. Obwohl Qualitätsunterschiede zwischen Leistungserbringern international in mehreren Ländern aufgezeigt wurden, fanden Qualitätsunterschiede zwischen Krankenhäusern in Deutschland bisher nur geringe Beachtung. Gleichzeitig wurden die verfügbaren Qualitätsdaten der externen Qualitätssicherung oder der Initiative Qualitätssicherung mit Routinedaten bisher nur eingeschränkt im Zeitverlauf betrachtet.

**Ziele**: Für die elektiven Leistungsbereiche Cholezystektomie, Herzschrittmacher-Implantation, Hüft-Endoprothesen-Implantation und perkutane koronare Intervention (PCI) sowie die Notfalldiagnosen Herzinfarkt und Schlaganfall werden allgemeine Qualitätstrends und Qualitätsunterschiede zwischen Krankenhäusern dargestellt. Auf Basis der Qualitätsunterschiede werden die Relevanz einer sorgfältigen Krankenhausauswahl für Patienten und einweisende Ärzte verdeutlicht und regulative Maßnahmen zur Steigerung der Krankenhausqualität diskutiert.

**Methoden**: Wir verwenden risikoadjustierte und einfache Ergebnisindikatordaten der G-BA Qualitätsberichte und der AOK QSR Initiative für deutsche Krankenhäuser und ausgewählte

Leistungsbereiche für die Jahre 2006 bis 2014. Die Analyse der Entwicklung von allgemeiner Behandlungsqualität und Qualitätsvarianz erfolgt durch Box-Whisker Plots und einfache lineare Regressionen jeweils für die ausgewählten Indikatoren. Für die Analyse der Qualitätsunterschiede werden die Werte der risiko-adjustierten Qualitätsindikatoren in Quintile eingeteilt und die Durchschnitte der Quintile gebildet. Darauf basierend wird der Outcome-Unterschied zwischen dem zweitbesten und dem schlechtesten Quintile gebildet, welches gleichzeitig die Steigerung des Mortalitäts- oder Reoperationsrisikos darstellt.

**Ergebnisse**: Die Qualitätstrends variieren je nach Leistungsbereich und Indikator. Cholezystektomie, Herzschrittmacher-Implantation und Hüft-Endoprothesen-Implantation zeigen positive Trends. Hingegen hat sich die Krankenhaussterblichkeit bei PCI erhöht. Die 30-Tage-Sterblichkeit bei Schlaganfall und Herzinfarkt hat sich verbessert, aber die 90-Tage Wiedereinlieferungsrate hat sich verschlechtert. Risiko-adjustierte Qualitätsunterschiede sind sowohl auf nationalem als auch regionalem Level weiterhin hoch. Für Cholezystektomie und Hüft-Endoprothesen-Implantation erhöht sich das Risiko für eine Re-Operation aufgrund von Komplikationen um das 4,6- und 11,5-fache bei einer Behandlung in einem Krankenhaus aus dem fünften und schlechtesten Quintile der Krankenhäuser relativ zu einem Krankenhaus aus dem zweitbesten Quintile. Für PCI erhöht sich das Mortalitätsrisiko um das 6,9-fache. Bei Schlaganfall und Herzinfarkt ist die 30-Tage Sterblichkeit um das 3,1- und 3,2-fache erhöht. Für Herzschrittmacher-Implantation steigt das durchschnittliche Mortalitätsrisiko um das 7,4-fache von einem Krankenhaus im dritten zu einem Krankenhaus im fünften Quintile. Die Qualitätsunterschiede bleiben auch für kleinere regionale Märkte wie Berlin, Hamburg und München bestehen. Public Reporting bedarf einer Aufbereitung und prägnanten Präsentation der hohen Qualitätsunterschiede, um die Brisanz und den Nutzen von qualitätsbasierter Krankenhauswahl zu verdeutlichen. Eine Zentralisierung von Behandlungen sollte Krankenhäuser mit guter Ergebnisqualität stärken und Krankenhäuser mit wiederholt schlechter Qualität sollten von der Behandlung ausgeschlossen werden.

Health Policy 121 (2017) 842–852

Contents lists available at ScienceDirect

Health Policy

journal homepage: www.elsevier.com/locate/healthpol

# Hospital quality variation matters – A time-trend and cross-section analysis of outcomes in German hospitals from 2006 to 2014

Christoph Pross[a], Reinhard Busse[a,b], Alexander Geissler[a,*]

[a] Berlin University of Technology, Department of Health Care Management, Germany
[b] European Observatory on Health Systems and Policies, Berlin Centre of Health Economics Research, Germany

ARTICLE INFO

*Article history:*
Received 16 February 2017
Received in revised form 19 June 2017
Accepted 25 June 2017

*Keywords:*
Hospital quality of care
Outcome variation
Quality assessment
AMI
Stroke
PCI
Cholecystectomy
Hip replacement
Pacemaker implantation
Germany

ABSTRACT

Awareness of care variation and associated differences in outcome quality is important for patients to recognize and leverage the benefits of hospital choice and for policy makers, providers, and suppliers to adapt initiatives to improve hospital quality of care. We examine panel data on outcome quality in German hospitals between 2006 and 2014 for cholecystectomy, pacemaker implantation, hip replacement, percutaneous coronary intervention (PCI), stroke, and acute myocardial infarction (AMI). We use risk-adjusted and unadjusted outcomes based on 16 indicators. Median outcome and outcome variation trends are examined via box plots, simple linear regressions and quintile differences. Outcome trends differ across treatment areas and indicators. We found positive quality trends for hip replacement surgery, stroke and AMI 30-day mortality, and negative quality trends for 90-day stroke and AMI readmissions and PCI inpatient mortality. Variation of risk-adjusted outcomes ranges by a factor of 3–12 between the 2nd and 5th quintile of hospitals, both at the national and regional level. Our results show that simply measuring and reporting hospital outcomes without clear incentives or regulation – "carrots and sticks" – to improve performance and to centralize care in high performing hospitals has not led to broad quality improvements. More substantial efforts must be undertaken to narrow the outcome spread between high- and low-quality hospitals.



## 1. Introduction

In recent years, substantial variation in the quality of hospital process and outcome indicators has been observed in and across health systems [1–5]. This variation is reported for emergency conditions, such as acute myocardial infarction (AMI) or stroke [4,7], as well as for elective procedures such as joint replacements and pacemaker implantation [8].

Across all conditions, outcome variation is often caused by medical errors and adverse events, which can be attributed to structural, process, and/or training deficiencies [9,10]. Data from European Union member states show that medical errors and adverse events occur in 8% to 12% of hospitalizations [11]. Studies analyzing medical errors show that 50–70% of these are preventable through comprehensive patient safety measures, better communication, and quality management systems [12].

Reducing outcome variation and improving outcomes overall are priorities for both policy makers and clinical leaders. In many countries, significant financial and time resources are currently being invested to build or enhance quality measurement and reporting systems, and to operate these systems in clinical practice [13,14]. Some of these systems were first implemented more than a decade ago. Examining recent time trends of hospital quality and inter-hospital variation is critical to evaluate and improve these measurement and reporting programs since their implementation.

Studies examining United States (US) hospital outcome data have reported mixed outcome quality trends. A study analyzing administrative data from 2001 to 2005 and multiple databases found improved risk-adjusted mortality, but mixed results for risk-adjusted complications and morbidity results [15]. Likewise, an analysis using Hospital Compare data from 2003 to 2009 for hospitals in 13 US states deduced no significant improvements in risk-adjusted mortality or complication ratios in Coronary Artery Bypass Graft (CABG) surgery or hip and knee replacements [16]. Similarly, no improvements in risk-adjusted death or re-hospitalization for ischemic stroke between 2003 and 2006 were identified [7]. In contrast, Cohen et al. found significant improvements in adverse, risk-adjusted surgical outcomes such as

* Corresponding author at: Berlin University of Technology, Dept. of Health Care Management, Administrative Office H80, Str. des 17. Juni 135, 10623, Berlin, Germany.
*E-mail address:* a.geissler@tu-berlin.de (A. Geissler).

http://dx.doi.org/10.1016/j.healthpol.2017.06.009
0168-8510/

mortality, morbidity, and infection between 2006 and 2013 in hospitals participating in the American College of Surgeons National Surgical Quality Improvement Program [17]. And a significantly improved overall standardized in-hospital AMI mortality was identified in data from the US Nationwide Inpatient Sample between 2001 and 2011 [18]. For a shorter timeframe, Werner and Bradlow observed improved hospital performance for process and outcome measures for AMI, heart failure, and pneumonia in CMS Hospital Compare data between 2004 and 2006 [19]. Results in European countries are similarly varied [20–24].

Since 2004, all German acute care hospitals are obliged to report structural, process, and outcome indicators as part of a national quality monitoring program [25]. The program collects clinical data for more than 350 indicators in 30 treatment areas, and as such places substantial administrative and resource-intensive burdens on hospitals [26]. To expand and improve national quality monitoring, the German government has recently created the Institute for Quality Assurance and Transparency in Healthcare (IQTIG). Annual national hospital quality reports summarize quality measures at a national average and relate them to those from the year prior for comparability and benchmarking purposes [27]. A study of the underlying hospital-level data for the 2004–2008 period found an overall improvement of 50% based on 204 analyzed indicators; however, this was primarily shaped by process indicators and many of the outcome indicators showed deterioration [28]. An investigation of national inpatient diagnosis-related group (DRG) statistics found AMI mortality rates to remain constant between 2005 and 2009 [29]. In a more recent analysis, Kraska et al. examined two indication appropriateness, three process, and one unadjusted outcome indicator from the mandatory hospital reporting from 2006 to 2012 and found significant quality improvements across all six measures. The data was divided into two-year periods, and he main improvement was observed to have occurred within the first measurement interval from 2006 to 2008 [30].

Using data for more substantial time periods (i.e. >5 years) and a greater number of hospitals (e.g. all acute hospitals for one country) has only recently become possible as many outcome measurement systems worldwide were only initiated in the last ten years [31]. Yet the evidence of improved (or deteriorated) outcomes, especially risk-adjusted outcomes, is mixed in Germany and other countries. Stroke and AMI outcome indicators in Germany, for example, have not yet been analyzed comprehensively, nor over time. In the international context, the analysis of inter-hospital variation over time is also scarce.

To address some of these research gaps, we analyze outcomes and their variation between 2006 and 2014 in German hospitals for the inpatient procedures cholecystectomy, pacemaker implantation, hip joint replacement, and percutaneous coronary intervention (PCI), and two emergency conditions, stroke and AMI. We examine three dimensions: 1) whether raw andy, pacemaker implantation, hip joint replacement, and percutaneous coronary intervention (PCI), and two emergency conditions, stroke and AMI. We examine three dimensions: 1) whether raw and risk-adjusted outcomes have improved, 2) whether inter-hospital variation has declined, and 3) the current magnitude of outcome variation between hospitals.

## 2. Methods

### 2.1. Dataset

We analyze a hospital-level panel dataset with repeated observations, and integrate 16 outcome indicators for six treatment areas from the two premier provider quality reporting systems in Germany. These are the national mandatory quality monitoring system of the Federal Joint Committee (G-BA) and the Quality Assurance with Administrative Data (QSR) from Germany's largest sickness fund, the AOK (25.3 Mio insured persons, ∼35% within the statutory health insurance scheme in 2016 [32]). We focus on outcome indicators as they are relevant end-points for patients. Further, the scientific evidence for the direct relationship between process and outcome indicators is unclear [33,34].

Indicators from the national quality monitoring system of G-BA are self-reported by hospitals, documented at the patient level, and publicly available for research and reporting at an aggregated hospital level. They comprise indication appropriateness, process, and outcome indicators for 30 treatment areas, from which we have selected a subset of ten (see Table 1). The selection considers data restrictions both in terms of indicator comparability across years and size of hospital samples.

The ten selected G-BA outcome indicators cover the four inpatient procedures only. They include six unadjusted outcome indicators such as mortality and re-intervention rates and four risk-adjusted outcome indicators such as risk-adjusted mortality and risk-adjusted re-intervention ratios. Risk-adjusted indicators compare the number of observed events (e.g. mortalities) with the number of expected events, with the latter calculated through a logistic regression that accounts for patient risk-factors such as age, comorbidities, gender and patient volume [35]. Annual risk-adjustment is undertaken centrally and ensures comparability of outcomes across hospitals and their respective patient samples.

The remaining six outcome indicators are from QSR and apply to the two emergency medical conditions stroke and AMI. They are centrally calculated by the WIdO Research Institute (a scientific body of the AOK), based on administrative data of AOK-insured patients. Stroke includes ICD diagnoses (i) intracerebral hemorrhage (ICD Code I61), (ii) ischemic stroke (I63), and (iii) stroke not specified as hemorrhage or ischemic (I64). AMI includes the diagnoses ST and non-ST elevation myocardial infarction (STEMI and NSTEMI) (I21) and subsequent STEMI and NSTEMI myocardial infarction (I22). The 30-day standardized mortality ratio (SMR) includes events up to 30 days after hospital discharge, comparing number of observed with number of expected events. Expected events for all data years are calculated based on the latest 2014 risk-adjustment model, via a logistic regression based on patients treated in each year and their respective risk-factors such as age, gender and comorbidities [36]. 30-day mortality and 90-day readmission rates are unadjusted. The QSR stroke and AMI data is not available publicly, access was granted only for this study and outcomes were provided at an aggregated level for each hospital. We link QSR indicators with the national quality monitoring data through standardized hospital IDs and hospital address data.

### 2.2. Statistical analysis

To examine outcome trends over time, we use box and whisker plots. For each indicator, respective patient volumes are included as analytical weights. The median illustrates the outcome trend. The 75th and 25th percentiles (box) specifies the interquartile range (IQR), the upper and lower adjacent values (whisker) are the most extreme values within 1.5 × IQR of the respective quartile. Both methods capture outcome variation.

To validate graphically identified time trends, we model indicator time trends using linear, longitudinal regressions. The log of the respective outcome indicator $QI_{it}$ for hospital $i$ in year $t$ is the dependent variable and $YEAR_t$ and a $log\_dummy_{it}$ are predictors. The time variable is an exogenous variable and the year coefficient summarizes the outcome time trend. Outcomes are log transformed to reduce skew and approximate normality [37]. We correct for zero values in the dependent variable, i.e. the optimal level of realized quality – a zero standardized mortality ratio (SMR) or zero non-

**Table 1**
Overview on data data sources, quality indicators and available data years.

| | Treatment Area | Indicator Name | Indicator Description | Data Years | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 6 | 8 | 10 | 12 | 13 | 14 |
| Mandatory National Quality Monitoring; admission-based data with observations limited to hospital stay; self-reported by hospitals aggregated on hospital-level; publicly available | Cholecystectomy (CHOL) | Risk-adjusted ratio of re-interventions due to complications | Ratio of observed vs. expecteted cases within hospital stay, expected cases being calculated via logic regression based on patient volume and patient risk profiles | | | | * | * | * |
| | | Re-interventions due to complications after laparoscopic surgery | Percentage of cases within hospital stay, unadjusted | * | * | * | * | * | * |
| | Pacemaker Implantation (PACE) | Risk-adjusted inpatient mortality ratio (SMR) | Standardized Mortality Ratio of observed vs. expecteted cases within hospital stay, expected cases being calculated via logic regression based on patient volume and patient risk profiles | | | | * | * | * |
| | | Inpatient mortality | Percentage of cases within hospital stay, unadjusted | | | | * | * | * |
| | | Surgical complications | Percentage of cases within hospital stay, unadjusted | * | * | * | * | * | * |
| | Hip replacement (HIP) | Risk-adjusted ratio of re-inter-ventions due to complications | Ratio of observed vs. expecteted cases within hospital stay, expected cases being calculated via logic regression based on patient volume and patient risk profiles | | | | * | * | * |
| | | Revisions due to complications | Percentage of cases within hospital stay, unadjusted | * | * | * | * | * | * |
| | Percutaneous Coronary Intervention (PCI) | Risk-adjusted inpatient mortality ratio (SMR) | Standardized Mortality Ratio of observed vs. expecteted cases within hospital stay, expected cases being calculated via logic regression based on patient volume and patient risk profiles | | | * | * | * | |
| | | Inpatient mortality | Percentage of cases within hospital stay, unadjusted | | | * | * | * | |
| | | Achieving the recanalization target in PCI | Percentage of successfully recanalized coronary arteries after indication acute coronary syndrome with ST elevation up to 24 h, unadjusted | * | * | * | * | * | |
| Administrative Data of the AOK sickness funds; patient-based data including events after hospital discharge; calculated centrally by the WIdO research institute and aggregated on hospital-level; not publicly available | Stroke | 30-day risk-adjusted mortality ratio (SMR) | Standardized Mortality Ratio of observed vs. expecteted cases including events up to 30 days after discharge, expected cases being calculated via logic regression based on patient volume and patient risk profiles | * | * | * | * | * | * |
| | | 30-day mortality | Percentage of cases including events up to 30 days after discharge, unadjusted | * | * | * | * | * | * |
| | | 90-day readmissions | Percentage of cases including events up to 90 days after discharge, unadjusted | * | * | * | * | * | * |
| | Acute Myocardial Infarction (AMI) | 30-day risk-adjusted mortality ratio (SMR) | Standardized Mortality Ratio of observed vs. expecteted cases including events up to 30 days after discharge, expected cases being calculated via logic regression based on patient volume and patient risk profiles | * | * | * | * | * | * |
| | | 30-day mortality | Percentage of cases including events up to 30 days after discharge, unadjusted | * | * | * | * | * | * |
| | | 90-day readmissions | Percentage of cases including events up to 90 days after discharge, unadjusted | * | * | * | * | * | * |

risk-adjusted outcome ratios – on both sides of the equation [38]. On the right hand of the regression equation, we include a dummy variable that takes on the value of 1 if the outcome (the dependent variable) is zero. On the left hand side, we add 1 to the outcome if it is zero before taking the log to adjust for zero values (optimal quality) in the hospital-level outcomes since the log of zero is undefined [38,39].

We use robust Huber-White standard errors $\varepsilon_{it}$ and respective patient volumes as analytical weights. To test the robustness of the simple parametric linear regression, we also model a non-parametric approach with dummy variables for years 2008, 2010, 2012, 2013, and 2014, depending on available data and base year.

To illustrate outcome variation between top and bottom performing hospitals, we divide the outcome distributions for the most recent data year into quintiles and calculate mean for each. This approach allows us to illustrate outcome differences between hospitals [1,4,6] and highlight the on-average higher mortality and complications rates when receiving treatment in low-quality hospitals rather than high-quality hospitals.

## 3. Results

To examine the trends in outcome quality, we summarize the box plots for risk-adjusted and unadjusted outcomes by treatment area in Fig. 1. The trend differs considerably across treatment areas. Overall quality improvement can be identified; however, outcome variation between hospitals has improved to a lesser degree and remains.

For cholecystectomy, both the risk-adjusted ratio of interventions and mean re-interventions due to complications decreased for the time period observed, with constant variation among hospitals. For pacemaker implantation, the median outcome rates improved for all three indicators examined; however, outcome variation has only decreased for unadjusted surgical complications. The median unadjusted ratio of re-interventions of hip replacement surgery decreased from 1.9% in 2006 to 1.1% in 2014. Variation also reduced notably. In contrast, both risk-adjusted and unadjusted inpatient mortality for PCI increased from 2010 to 2013; the variation for unadjusted mortality increased.

For the emergency condition stroke, risk-adjusted 30-day SMR and the unadjusted 30-day mortality decreased between 2006 and 2014. Similarly, outcome variation has narrowed. 90-day readmission on the other hand increased from 22% to 24%, and variation remained constant. For AMI, the risk-adjusted 30-day SMR and unadjusted mortality rate have decreased, with a strong decrease in unadjusted mortality from 15% to 12%. In contrast, AMI readmissions increased substantially both in median and variation.

Table 2 provides descriptive statistics (mean, standard deviation, and number of observations) for the selected risk-adjusted and unadjusted outcome indicators. Particularly noteworthy is the decline (from reporting year 2010 onwards) in hospital observations for the unadjusted indicators reported by the mandatory national quality monitoring system. The number of observations declined due to data privacy restrictions imposed in 2010 and 2012, which limits reporting to hospitals with more than three cases in the numerator or denominator. As a result, many hospitals with low case volumes exited the sample (see also Online Appendix). Table A1 in the Appendix provides additional insights on the overall composition of the dataset.

Linear parametric and non-parametric regressions allow for another perspective on the outcome trend. Table 3 presents the results for the linear regressions by treatment area, indicator, and for the parametric and non-parametric regression models. The non-parametric confirms the parametric approach and the regression coefficients confirm the visual trends identified in the box plots for four out of the six treatment areas. The quality improvement trends identified visually for cholecystectomy and pacemaker implantation could not be confirmed. Instead, the coefficients show statistically weak deteriorating quality trends. For robustness, we also ran parametric regressions only with observations that treat more than 20 patients per year in the respective treatment areas. The results remain comparable.

By examining the quintile distribution for risk-adjusted outcomes for 2013/2014 in Fig. 2, we highlight the outcome differences between high and low quality hospitals. Given that all indicators are risk-adjusted, there is a large amount of unexplained variation.

For cholecystectomy, the best performing (1st) hospital quintile has a low re-intervention ratio of 0.1. The worst performing (5th) quintile has an average risk-adjusted ratio of 2.4, which is 140% higher than expected (standardized ratio of 2.4 vs. 1.0). The risk of re-interventions is 4.6 times higher in 5th quintile hospitals compared to the 2nd quintile average, and 35 times higher relative to the average in the best performing quintile.

For hip replacement, the 1st quintile has no re-interventions. The average standardized ratio of re-interventions in the 5th quintile is 4.3. Compared to the 2nd quintile, the risk of re-interventions is 12.2 times higher in the 5th quintile. Similar patterns of high outcome variation can be observed for the procedures pacemaker implantation and PCI.

Outcome variation is only slightly smaller for the emergency conditions stroke and AMI. For stroke, the best performing quintile has an average 30-day SMR of 0.04, while the worst performing 5th quintile has a 30-day SMR ratio of 2.1. On average, 110% more patients die than expected based on hospitals' patient populations in a 5th quintile hospital. When moving from the 2nd quintile to the 5th quintile, mortality risk increases by a factor of 3.3. The jump in the mortality risk from the 1st compared to the 5th quintile is still higher. For AMI, the 1st quintile has a 30-day SMR of 0.2, while the 5th hospital quintile has a SMR ratio of 2.1. This is 110% higher than expected. The mortality risk is 2.9 times higher in the 5th quintile relative to the 2nd quintile; it is 11 times higher relative to the 1st quintile of best performing hospitals.

Similar patterns of high outcome variation can be observed for more confined hospital markets in city regions such as Berlin (Fig. 3), Hamburg, and Munich (Online Appendix, Figures A2 & A3). While variation captured through the quintile distributions decreased from 2006 to 2013/2014 in line with results in Fig. 1, substantial outcome differences between high and low quality hospitals remain. To test the robustness of the variation results, we excluded hospitals with less than 10 and 20 patient cases per annum for the respective procedures or diagnoses. Though the outcome variation decreases somewhat, it remains high.

## 4. Discussion

German hospitals exhibit substantial quality variation and mixed quality trends depending on treatment area and outcome measure. The three treatment areas cholecystectomy, pacemaker implantation, and hip replacement surgery show a positive quality trend over time, while PCI exhibits a negative quality trend. For the emergency conditions stroke and AMI, 30-day mortality has decreased, while 90-day readmissions have increased. This trend was also observed in English hospitals, and can be driven by an increasing number of severely ill and fragile patients that survive AIM or stroke, and the higher probability that these patients may be readmitted [40]. The descriptively identified positive quality trends for hip replacement surgery, stroke, and AMI 30-day mortality are supported by the results of time trend regression models. Likewise, the negative quality trend for 90-day stroke, AMI readmissions, and PCI inpatient mortality could also be verified with a

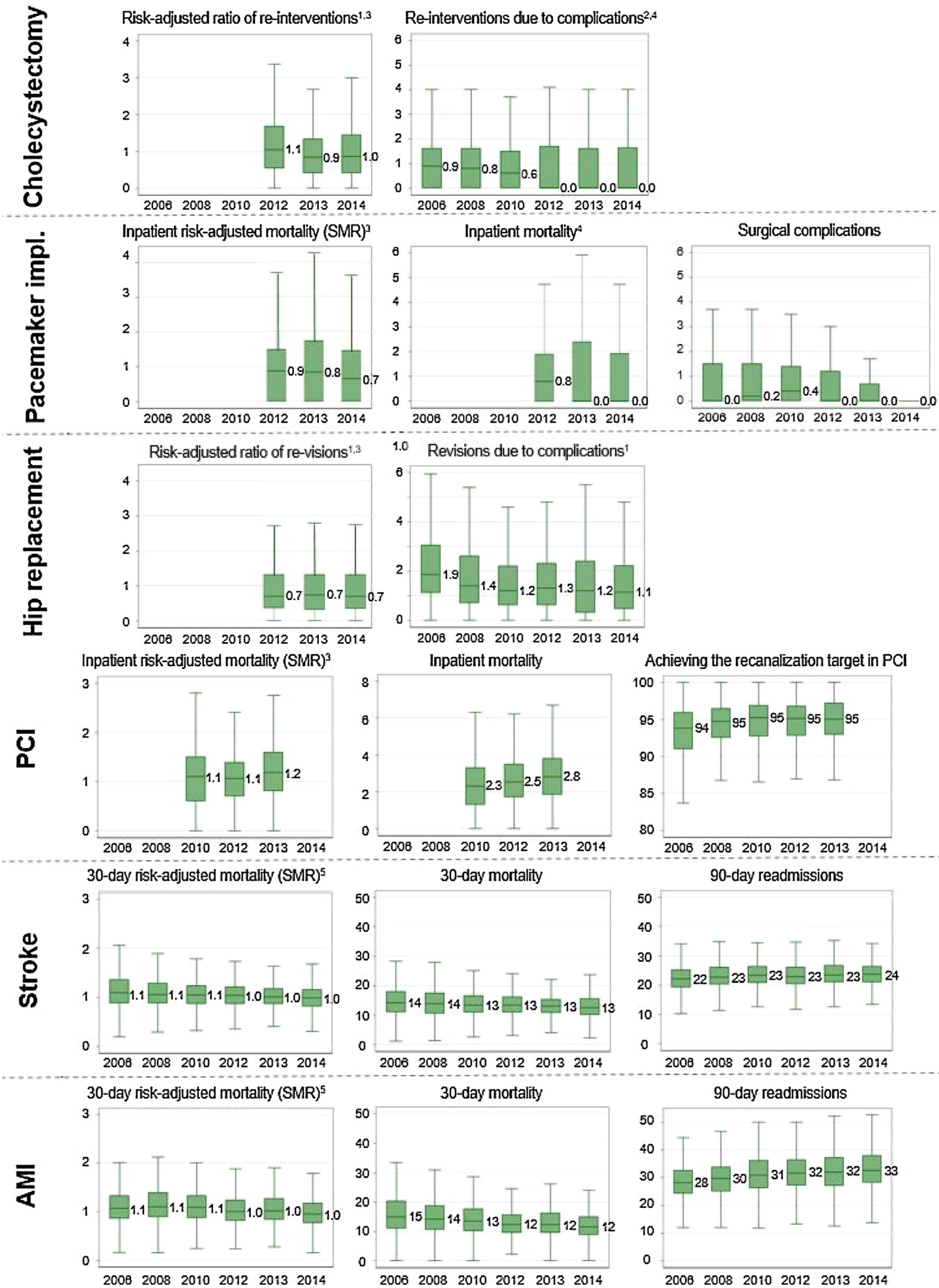



**Fig. 1.** Box plots for risk-adjusted[5] (left) and unadjusted outcome indicators[6] (middle and right).
Note: Box plots display median, IQRs and 1.5 × IQR whiskers, with box plot values weighted by patient volume in respective treatment areas as analytical weights; Outside values excluded (outside values > 1.5* interquartile range above the upper quartile/below the lower quartile); 1. Re-interventions due to complications; 2. Re-intervention due to complications after laparoscopic surgery; 3. Risk factors consistent across years, but standardization completed annually limiting year on year comparability if crude rate exhibits large improvement; 4. Crude outcome indicators loose substantial amount of observations with small patient volumes due to data privacy concerns (see Table 2 and Fig. A2); 5. Risk-adjusted ratios compare observed vs. expected evens (1.0 with same number of observed and expected events); expected mortality for all data years calculated based on the latest 2014 risk-adjustment model ensuring full comparability across years 6. Non-risk-adjusted rates are in%, indicating the mortality or complications share of all patients treated in the treatment area and hospital.

**Table 2**
Descriptive statistics for outcome indicators by treatment area and year (mean and standard deviation, number of observations)

| Treat. area | Indicator name | 2006 | | | 2008 | | | 2010 | | | 2012 | | | 2013 | | | 2014 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | SD | # hosp. | Mean | SD | # hosp. | Mean | SD | # hosp. | Mean | SD | # hosp. | Mean | SD | # hosp. | Mean | SD | # hosp. |
| CHOL | Risk-adjusted ratio of re-inter-vention due to complications (O/E) | | | | | | | | | | 1.2 | 0.9 | 1088 | 1.0 | 0.7 | 1147 | 1.1 | 0.8 | 1123 |
| | Re-interventions due to complica-tions after laparoscopic surgery (%) | 1.1 | 1.2 | 1022 | 1.1 | 1.3 | 1158 | 0.9 | 1.2 | 1129 | 1.0 | 1.6 | 754 | 0.9 | 1.5 | 571 | 0.9 | 1.5 | 555 |
| PACE | Risk-adjusted inpatient mortality ratio (SMR, O/E) | | | | | | | | | | 1.1 | 1.3 | 982 | 1.2 | 1.5 | 1023 | 1.0 | 1.3 | 995 |
| | Inpatient mortality (%) | | | | | | | | | | 1.2 | 1.6 | 707 | 1.2 | 1.6 | 576 | 1.1 | 1.7 | 574 |
| | Surgical complications (%) | 0.8 | 1.4 | 735 | 0.9 | 1.5 | 984 | 1.0 | 1.5 | 985 | 0.7 | 1.4 | 736 | 0.6 | 1.5 | 608 | 0.5 | 1.1 | 634 |
| HIP | Risk-adjusted ratio of re-inter-ventions due to complications (O/E) | | | | | | | | | | 1.0 | 0.9 | 1064 | 1.0 | 1.0 | 1117 | 1.0 | 1.2 | 1111 |
| | Revisions due to complications (%) | 2.2 | 1.7 | 904 | 1.9 | 1.8 | 1092 | 1.6 | 1.6 | 1084 | 1.7 | 1.7 | 717 | 1.6 | 1.7 | 587 | 1.5 | 1.6 | 551 |
| PCI | Risk-adjusted inpatient mortality ratio (SMR, O/E) | | | | | | | 0.9 | 1.4 | 711 | 0.9 | 1.0 | 747 | 1.1 | 2.0 | 815 | | | |
| | Inpatient mortality (%) | | | | | | | 2.5 | 1.7 | 655 | 2.7 | 1.6 | 637 | 2.9 | 1.5 | 597 | | | |
| | Achieving the recanalization target in PCI (%)[1] | 92.0 | 11.3 | 421 | 94.1 | 4.1 | 533 | 94.4 | 4.0 | 571 | 94.4 | 5.0 | 599 | 94.7 | 3.2 | 592 | | | |
| Stroke | 30-day risk-adjusted mortality ratio (SMR, O/E) | 1.1 | 0.4 | 1331 | 1.1 | 0.4 | 1292 | 1.1 | 0.4 | 1244 | 1.1 | 0.3 | 1228 | 1.0 | 0.3 | 1204 | 1.0 | 0.3 | 1161 |
| | 30-day mortality (%) | 14.7 | 6.0 | 1331 | 14.4 | 5.9 | 1292 | 14.1 | 5.9 | 1244 | 13.9 | 5.4 | 1228 | 13.5 | 5.3 | 1204 | 13.1 | 5.5 | 1161 |
| | 90-day readmissions (%) | 22.4 | 5.7 | 1331 | 23.3 | 6.0 | 1292 | 23.9 | 0.0 | 1244 | 23.7 | 5.8 | 1228 | 24.0 | 5.7 | 1204 | 24.0 | 5.8 | 1161 |
| AMI | 30-day risk-adjusted mortality ratio (SMR, O/E) | 1.1 | 0.4 | 1238 | 1.2 | 0.4 | 1213 | 1.1 | 0.5 | 1141 | 1.1 | 0.4 | 1146 | 1.1 | 0.4 | 1130 | 1.0 | 0.4 | 1093 |
| | 30-day mortality (%) | 16.1 | 7.6 | 1238 | 15.4 | 7.3 | 1213 | 14.5 | 6.8 | 1141 | 13.3 | 6.3 | 1146 | 13.4 | 6.4 | 1130 | 12.4 | 6.1 | 1093 |
| | 90-day readmissions (%) | 28.4 | 7.4 | 1238 | 29.6 | 7.5 | 1213 | 31.2 | 0.0 | 1141 | 32.0 | 7.8 | 1146 | 32.5 | 8.3 | 1130 | 33.1 | 7.8 | 1093 |

**Note**: Weighted mean and SD (case volumes in each treatment area used analytical weights); Means in this table are higher as medians in Fig. 1 due to distributions being skewed right with many hospitals lower rates and fewer hospitals with higher rates; SMR = Standardized Mortality Ratio, which compares observed events with expected events, the number of expected events being calculated via logic regression based on patient volume and patient risk profiles; mean and standard deviation weighted by patient number in respective treatment areas; [1] = Outcome indicator representing share of successfully recanalized coronary arteries after indication acute coronary syndrome with ST elevation up to 24 h; **Sources**: German National Quality Monitoring System and AOK sickness funds, own calculations.

**Table 3**
Results for parametric and non-parametric regression.

| Treatm. area | Outcome Indicator | Parametric | | Non-parametric | | | | | Quality Trend[1] |
|---|---|---|---|---|---|---|---|---|---|
| | | Beta (p) | $R^2$ | 2008 | 2010 | 2012 | 2013 | 2014 | |
| CHOL | Risk-adjusted ratio of re-intervention | -0.02 | 0.25 | no data | no data | base year | 0.02 | omitted[2] | → |
| | Re-interventions due to complications[4] | 0.02*** | 0.16 | 0.03 | -0.03 | 0.07* | 0.16*** | 0.17*** | ↘ |
| PACE | Risk adjusted inpatient mortality ratio (SMR) | -0.03 | 0.02 | no data | no data | base year | 0.03 | -0.06 | → |
| | Inpatient mortality | 0.05* | 0.37 | | | | 0.11** | 0.10** | ↘ |
| | Surgical complications | 0.00 | 0.14 | -0.04 | -0.06 | -0.01 | 0.02 | 0.02 | → |
| HIP | Risk-adjusted ratio of re-interventions | -0.02 | 0.46 | no data | no data | base year | -0.01 | -0.04 | → |
| | Revisions due to complications | -0.02** | 0.05 | -0.18*** | -0.25*** | -0.19*** | 0.15 | -0.19 | ↗ |
| PCI | Risk-adjusted inpatient mortality ratio (SMR) | 0.04** | 0.01 | no data | base year | 0.01 | 0.14*** | no data | ↘ |
| | Inpatient mortality | 0.08*** | 0.08 | | | 0.16*** | 0.26*** | | ↘ |
| | Achieving the recanalization target[3] | 0.00*** | 0.01 | 0.01** | 0.02*** | 0.02** | 0.01*** | | → |
| Stroke | Risk-adjusted 30-day mortality ratio (SMR) | -0.01*** | 0.01 | -0.02 | -0.04*** | -0.05*** | -0.07*** | -0.11*** | ↗ |
| | 30-day mortality | -0.01*** | 0.35 | -0.02 | -0.03* | -0.04** | -0.07*** | -0.10*** | ↗ |
| | 90-day readmissions | 0.01*** | 0.39 | 0.04*** | 0.07*** | 0.06*** | 0.08*** | 0.08** | ↘ |
| AMI | Risk-adjusted 30-day mortality ratio (SMR) | -0.01*** | 0.01 | 0.03 | 0.01 | -0.06*** | -0.04** | -0.12*** | ↗ |
| | 30-day mortality | -0.03*** | 0.24 | -0.04* | -0.09*** | -0.18*** | -0.18*** | -0.26*** | ↗ |
| | 90-day readmissions | 0.02*** | 0.31 | 0.04*** | 0.1*** | 0.12*** | 0.14*** | 0.16*** | ↘ |

**Note:** Patient case volume in treatment areas used as analytical weights; stars indicate the significance level based on p-values (*90%; **95%; ***99%); 1 = trend indication, if parametric time variable insignificant at 10%, then flat trend; if time coefficient significant at 10% level, then positive quality trend if outcome ratio improves and negative quality trend if outcome ratio worsens; 2due to collinearity; 3 = Outcome indicator representing share of successfully recanalized coronary arteries after indication acute coronary syndrome with ST elevation up to 24 h; 4 = after laparoscopic surgery

time trend model. For cholecystectomy and pacemaker implantation, the visual time trends could not be confirmed. These results are in line with the few previous studies for selected unadjusted outcomes in Germany [30,41], which found mixed outcome trends.

The benefit of the G-BA hospital quality monitoring program as well as the validity of the quality improvement trends for some of its indicators are weakened by the decreasing number of hospitals reporting results. Due to data privacy concerns, mandatory reporting data specifications since 2010 require that if in a specific hospital the actual number of deaths or other outcomes is below 4 or 5 cases, no outcome data needs to be reported [42]. Smaller hospitals with low case volumes are primarily affected by this requirement, raising the possibility that the quality improvement trend for non-risk-adjusted indicators (Fig. 1) might be associated with hospitals with inferior outcomes and larger outcome variation exiting the sample due to data privacy requirements. The reduction in the number of hospital observations for stroke and AMI is due to hospital closures and mergers, and is in line with overall reductions in the number of German hospitals.

For AMI, the positive mortality trend might also be caused by greater diagnoses of less severe AMIs with lower risk of death as a result of increased use and dissemination of highly sensitive cardiac biomarkers [43,44]. In our data, the share of severe ST elevation myocardial infarction (STEMI) decreased to 41% in 2014 from 49% in 2006, with the share of non-STEMIs increasing accordingly to 59%. The different quality trends for these two different outcomes, mortality and readmission, also highlight the importance of tracking a spectrum of outcomes to comprehensively measure outcome quality, and to avoid a narrow focus on one measure (e.g. 30-day mortality) while not paying attention to other important outcomes (e.g. 90-day readmission).

Outcome variation remains high today, with substantially higher mortality or re-intervention risks observed in lower performing hospitals as opposed to better performing hospitals. On average, the risk of a negative outcome is 3–12 times greater in the 5th quintile of the risk-adjusted outcome distribution as opposed to the 2nd quintile. For the elective procedures cholecystectomy, pacemaker implantation, hip replacement, and PCI, patients (with the freedom of choice, aided by their admitting physicians) would benefit substantially from selecting a 2nd quintile hospital, lowering their adjusted risk of negative outcomes by 70–90% on average relative to the 5th hospital quintile. Similarly, the category of hospital to which emergency stroke or AMI patients are admitted matters significantly for their chances of survival. If a patient is taken to a 2nd quintile hospital, rather than a 5th quintile hospital, the risk of death is reduced by roughly 70% for both conditions.

### *4.1. Implications for health policy*

Mixed outcome trends and persistent patterns of outcome variation raise policy questions about the efficiency and effectiveness of current hospital quality measurement, reporting, and improvement systems. Instruments that help to reduce outcome variation and support patients to receive the best available care need to be strengthened. However, considering the existing system of mandatory quality monitoring and peer reviews, substantial consequences for or sanctions on weaker performing hospitals, e.g. prohibition from the treatment of certain conditions or public

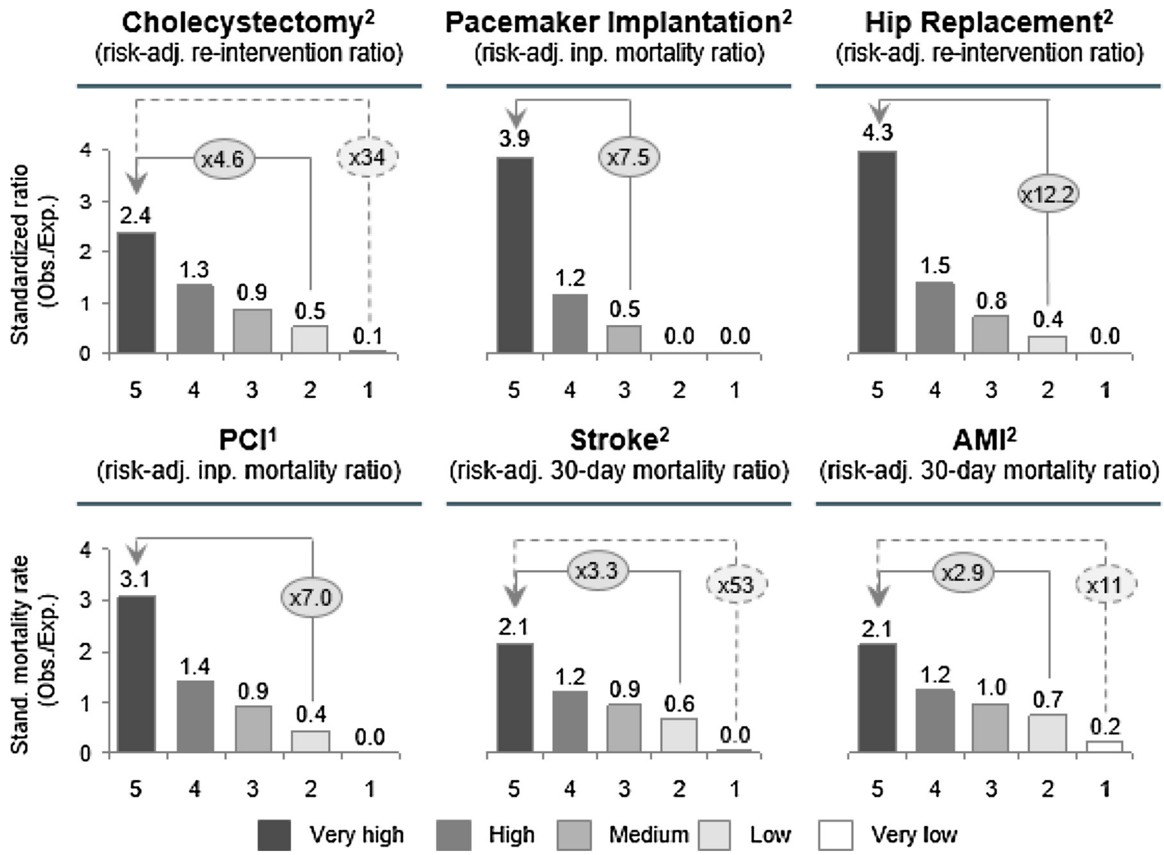



**Fig. 2.** Risk-adjusted outcome quintiles, 2013 and 2014.
Note: Distribution divided into quintiles and average for each quintile calculated and displayed, quintiles displayed from left with highest risk-adjusted mortality or re-intervention rate to right with lowest; multiple displays increased risk of negative outcome (mortality or re-intervention) when being treated in an average 5th quintile hospitals vs. an average 2nd quintile (or 1st quintile) hospital; 1. data from 2013; 2. data from 2014

disclosure of low quality, are rare to non-existent [45–47]. Furthermore, the increasing number of hospitals exempt from full reporting requirements due to data privacy considerations raises the concern that smaller hospitals with lower case volumes and frequently lower outcome quality operate outside the emerging system of quality transparency.

Focusing on quality measurement, Germany, like many other countries, is currently pursuing three major policy initiatives to improve outcomes and reduce quality of care variation in hospitals: public reporting, care centralization, and quality-related payments. However, stable or even deteriorating hospital quality trends and persistent levels of variation suggest that these initiatives need to be critically reviewed.

#### *4.1.1. Public reporting*

Several international studies have raised questions about the impact of existing public reporting systems on hospital quality and patient choice [48–51]. For the latter, these studies highlight the continued, predominantly patient relevant, factors such as distance to hospital, personal experiences, amenities, and waiting times [51,52]. Patients are often not aware of quality differences between hospitals ('no need') or the relevant information source ('no access'). Further, once patients have accessed the quality data, they might have difficulty deciding between quality of care and convenience factors, or may not understand reported results and potential choice influences ('no impact').

In Germany, as in other countries, the use of public reporting is increasing, though overall it is not meeting many enthusiasts' expectations regarding patient choice and hospital quality improvements [48,53]. Sickness funds and public reporting portals have conducted various surveys to acquire information about patients' evaluations of hospital quality. Results show that patients are increasingly aware of hospital quality variation and that patients would travel larger distances for better quality [54]. However, assessment of quality variation through available public information can be difficult for patients, who often lack the medical knowledge and cognitive abilities to understand the information provided [48,55,56].

Admitting general and specialist physicians play an important role in shaping quality of care information conveyed to patients and patient hospital choice. 90% of patients in Germany attribute a large or very large role to admitting physicians in their own hospital decisions [53]. Admitting physicians' hospital recommendations thus need to be regularly informed by independent, publicly reported hospital outcomes, for example through quality of care newsletters circulated in their regions or specialty areas (e.g. the AOK Newsletter for admitting physicians), or through their own habitual use of public reporting portals. Information about outcome variation between hospitals in local market needs to be made easily accessible to admitting physicians and patients alike.

A patient literacy and information campaign on substantial quality of care differences between hospitals, especially within local markets, can help to increase patient awareness of outcome differences, and thus their own interest in choosing high over low performing hospitals. This could also result in increased demand for information on quality and for related advice from their admitting physician. While patient involvement in decisions about their own healthcare should increasingly be the norm, many care professionals and provider organizations ignore or even reject the concept [57,58].

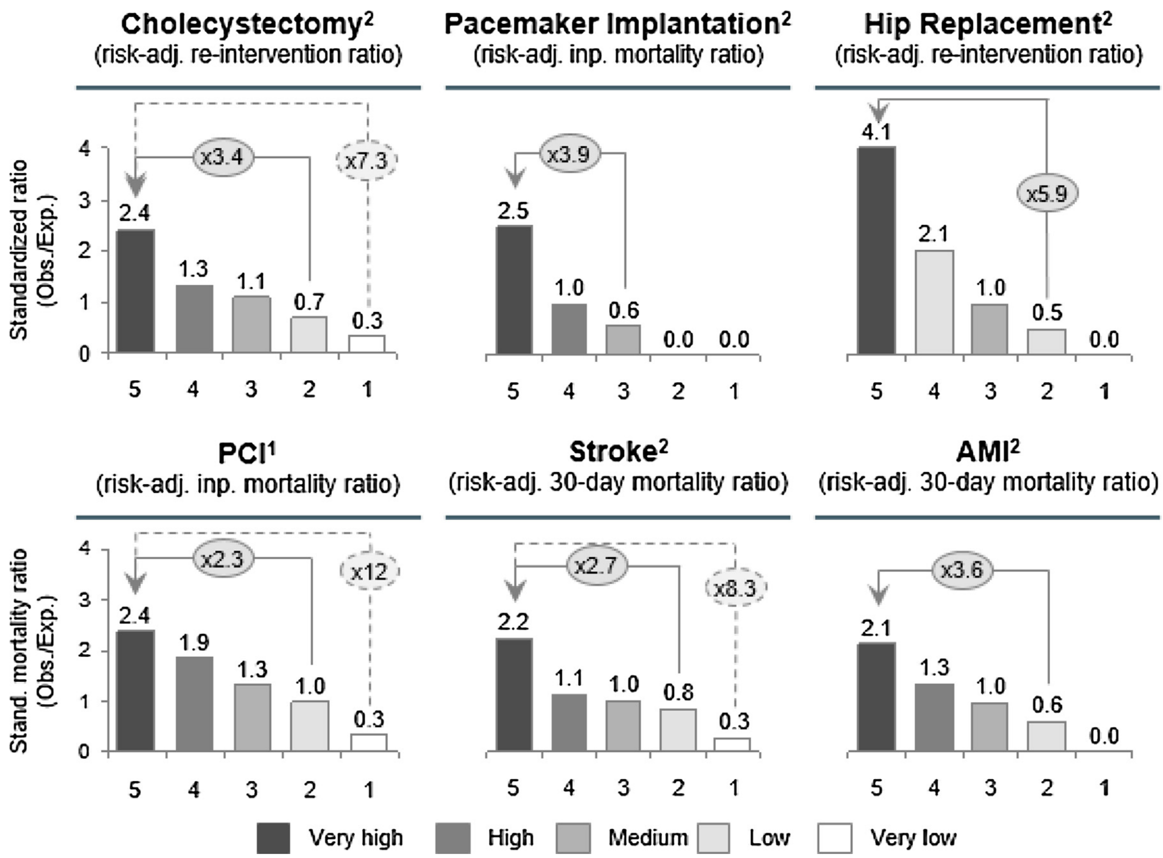



**Fig. 3.** Risk-adjusted outcome quintiles for Berlin hospitals, 2013 and 2014.
Note: Distribution divided into quintiles and average for each quintile calculated and displayed, quintiles displayed from left with highest risk-adjusted mortality or re-intervention rate to right with lowest; multiple displays increased risk of negative outcome (mortality or re-intervention) when being treated in an average 5th quintile hospitals vs. an average 2nd quintile (or 1st quintile) hospital; CHOL = 38, PACE with n = 30, HIP with n = 33, PCI with n = 22, Stroke with n = 39, and AMI with n = 41 hospital observations; 1. data from 2013 2. data from 2014

*4.1.2. Care centralization*

Quality-based centralization is especially relevant in Germany, which has significant hospital overcapacities [59,60] and a hospital payment and capacity planning system that neglects information on quality [61]. Minimum volume restrictions, quality-based hospital budgets and capacity planning, and selective contracting are emerging as the three regulatory approaches for quality-based care centralization. However, in Germany today these policy measures are ineffective or not yet implemented.

Germany introduced minimum volume standards for 7 treatment areas in 2004 (e.g. knee replacement and coronary interventions); yet, empirical analysis has shown no concentration effect. Substantial shares of the selected procedures are still being performed at hospitals with annual volumes below the minimum volume threshold [62,63]. The limited concentration is in part due to the lack of regulation or consequences for not implementing – the legislation. Quality-based hospital budgets and hospital capacity planning processes were introduced with the 2015 Hospital Structure Law [64]. However, implementation is slow, and the focus on and benefit for quality of care is unclear. Sickness funds are currently still obligated to reimburse any service provided by hospitals, though selective contracting is slowly being allowed. For example, the 2015 AOK Hessia's selective hip and knee contracts require minimum levels of outcome quality.

*4.1.3. Quality-related payments*

In contrast to countries such as the US and the UK, Germany has not implemented large-scale quality-related payment or pay-for-performance (P4P) programs. The recent Hospital Structure Law introduced a bonus-malus payment based on risk-adjusted hospital quality indicators [64]. As before, however, scope and timeline for implementation is unclear. As an alternative to the classic P4P program, top performing hospitals, for example 1st quintile hospitals in Fig. 2, could receive tied quality bonus payments to expand capacities and expertise in the treatment areas in which they achieve the best risk-adjusted outcomes. Similarly, reimbursement could be denied to hospitals for never- or sentinel events, similar to the CMS "Never Events" policy [65]. The diverging quality trend for 30-day mortality and 90-day readmission also demonstrates the need to tie payments to multiple outcome indicators measuring outcome quality comprehensively and over a longer period.

*4.2. Limitations*

The results of this study should be interpreted in light of some data and methodological limitations. The validity of self-reported hospital data might be compromised, due to reputational concerns by hospitals, competition between hospitals, and/or different coding practices. Annual, random validity checks such as basic statistical tests and cross-checks with administrative patient files, has demonstrated some validity issues affecting 15–60% of the examined reporting data [27,66] for a small share of hospital reports (5%).

To prevent false incentives and provide a more comprehensive perspective on treatment outcomes, quality indicators should also include negative outcomes that occur only after hospital discharge. The 30-day SMRs and 90-day readmission rates for stroke and AMI incorporate a certain timeframe post-discharge, but an extended period such as 180 or 365 days might provide additional insights. The stroke and AMI outcomes are also more robust as they rely on administrative data and centralized calculation whereas the outcome data from the public monitoring system relies on clinical data self-reported and collected by each hospital. Further, while the AOK QSR indicators have some advantages, they only rely on data for patients insured by the AOK sickness fund, potentially leading to biased outcome indicators. However, the high proportion of AOK insured patients in all German hospitals (35% average market share), the even higher share among hospital patients (>40%), and previous studies [67] let us assume representativeness of the AOK QSR data.

Due to the data privacy restrictions, unadjusted indicators in later data years have a smaller number of observations. In particular, smaller hospitals with low case volumes are removed from the sample. This positively skews the data in two ways and might in part result in the strong decline in the median mortality or complication ratios (e.g. zero median for cholecystecomy re-interventions). First, many hospitals with 1–3 mortality or complication cases in the numerator exit the sample. Second, hospitals with lower case volumes are also likely to have worse outcomes due to the positive outcome-volume relationship. Data privacy concerns need to be weighed against the downsides of excluding hospitals with smaller case volumes from measurement and reporting requirements.

For the risk-adjusted indicators from the public monitoring system, comparability across years is limited as the risk-adjustment is undertaken each year (based on the same risk factors and technique), and risk-factors are weighted to reach a federal average risk-adjusted ratio of 1.0, which implies that for all patients treated, the expected mortality equals the observed mortality. To display a fully comparable time trend, the underlying unadjusted outcome rates are provided as well. For stroke and AMI, the risk-adjustment is completed with the latest 2014 risk-adjustment model, which ensures full comparability across years.

## 5. Conclusion

Since the introduction of mandatory hospital quality monitoring in 2004, outcomes have improved and outcome variation has reduced for some conditions, have remained constant for others, and even deteriorated for yet other conditions and indicators. Patients face substantially different risks of death or complications depending on which hospital they go to for elective procedures or which they are admitted to for emergency care. Public reporting, its consumption and use by patients and physicians as well as care centralization at high quality centers could be two effective levers for improving quality. This could help shift care towards centers of excellence and away from hospitals with continued subpar performance. In Germany, public reporting, reimbursement, and capacity planning systems still lack relevant legislative and administrative "carrots and sticks" to substantially drive improved outcomes and centralize care. Likewise, patients and admitting physicians lack awareness and incentives to use information on quality in their hospital choice decisions.

## Funding source

Christoph Pross receives a general PhD scholarship from the Konrad Adenauer Foundation. Otherwise, no special funding for this article was received.

## Conflict of interest

No conflicts to report.

## Acknowledgements

We thank the Research Institute of the AOK (WIdO) for granting us access to the QSR outcome data used for this article. Furthermore, we thank Katherine M. Polin for a critical review and language editing of the final manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.healthpol.2017.06.009.

## References

[1] Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. The New England Journal of Medicine 2009;361:1368–75.
[2] Tsai TC, Joynt KE, Wild RC, Orav EJ, Jha AK. Medicare's bundled payment initiative: most hospitals are focused on a few high-volume conditions. Health affairs (Project Hope) 2015;34:371–80.
[3] Hawkins BM, Kennedy KF, Aronow HD, Nguyen LL, White CJ, Rosenfield K, et al. Hospital variation in carotid stenting outcomes. JACC Cardiovascular Interventions 2015;8:858–63.
[4] Chung S-C, Sundström J, Gale CP, James S, Deanfield J, Wallentin L, et al. Comparison of hospital variation in acute myocardial infarction care and outcome between Sweden and United Kingdom: population based cohort study using nationwide clinical registries. BMJ (Clinical research ed.) 2015;351:h3913.
[5] Shaw C, Bruneau C, Kutryba B, de Jongh G, Suñol R. Towards hospital standardization in Europe. International Journal for Quality in Health Care 2010;22:244–9.
[7] Fonarow GC, Smith EE, Reeves MJ, Pan W, Olson D, Hernandez AF, et al. Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. Stroke; A Journal of Cerebral Circulation 2011;42:159–66.
[8] Robertsson O, Ranstam J, Lidgren L. Variation in outcome and ranking of hospitals: an analysis from the Swedish knee arthroplasty register. Acta orthopaedica 2006;77:487–93.
[9] Kohn LT, Corrigan J, Donaldson MS. To err is human: building a safer health system. Washington D.C: National Academy Press; 2000.
[10] National Institute for Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington, D.C: National Academy Press; 2001.
[11] EEIG, Patient safety and quality in healthcare: Full report, available at http://ec.europa.eu/public_opinion/archives/ebs/ebs_327_en.pdf. (Accessed on 4 May 2016).
[12] WHO Europe, Patient Safety: Data and Statistics, available at http://www.euro.who.int/en/health-topics/Health-systems/patient-safety/data-and-statistics.
[13] Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. Medical Care 2003;41:I30–8.
[14] Kumpunen S, Trigg L, Rodrigues R. Public reporting in health and long-term care to facilitate provider choice; 2017, available at http://www.euro.who.int/en/about-us/partners/observatory/publications/policy-briefs-and-summaries/public-reporting-in-health-and-long-term-care-to-facilitate-provider-choice. (Accessed on 28 October 2015).
[15] Kroch EA, Duan M, Silow-Carroll S, Meyer JA. Hospital performance improvement: trends on quality and efficiency: a quantitative analysis of performance improvement in U.S. hospitals; 2007.
[16] Shih T, Nicholas LH, Thumma JR, Birkmeyer JD, Dimick JB. Does pay-for-performance improve surgical outcomes? An evaluation of phase 2 of the premier hospital quality incentive demonstration. Annals of Surgery 2014;259:677–81.
[17] Cohen ME, Liu Y, Ko CY, Hall BL. Improved surgical outcomes for ACS NSQIP hospitals over time: evaluation of hospital cohorts with up to 8 years of participation. Annals of Surgery 2015;263:267–73.
[18] Sugiyama T, Hasegawa K, Kobayashi Y, Takahashi O, Fukui T, Tsugawa Y. Differential time trends of outcomes and costs of care for acute myocardial infarction hospitalizations by ST elevation and type of intervention in the United States, 2001–2011. Journal of the American Heart Association 2015;4:e001445.
[19] Werner RM, Bradlow ET. Public reporting on hospital process improvements is linked to better patient outcomes. Health affairs (Project Hope) 2010;29:1319–24.
[20] Jacques RM, Fotheringham J, Campbell MJ, Nicholl J. Did hospital mortality in England change from 2005 to 2010? A retrospective cohort analysis. BMC Health Services Research 2013;13:216.
[21] Neuburger J, Currie C, Wakeman R, Tsang C, Plant F, de Stavola B, et al. The impact of a national clinician-led audit initiative on care and mortality after hip fracture in England: an external evaluation using time trends in non-audit data. Medical Care 2015;53:686–91.
[22] Duclos A, Polazzi S, Lipsitz SR, Couray-Targe S, Gawande AA, Colin C, et al. Temporal variation in surgical mortality within French hospitals. Medical Care 2013;51:1085–93.

[23] Falcoz P-E, Puyraveau M, Rivera C, Bernard A, Massard G, Mauny F, et al. The impact of hospital and surgeon volume on the 30-day mortality of lung cancer surgery: a nation-based reappraisal. The Journal of Thoracic and Cardiovascular Surgery 2014;148:841–8, discussion 848.
[24] van den Berg M, de Boer D, Gijsen R, Heijink R, Limburg L, Zwakhals S. Dutch Health Care Performance Report 2014. Bilthoven; 2015.
[25] Busse R, Nimptsch U, Mansky T. Measuring, monitoring, and managing quality in Germany's hospitals. Health affairs (Project Hope) 2009;28:w294–304.
[26] G-BA, Die gesetzlichen Qualitätsberichte 2012 der Krankenhäuser lesen und verstehen, available at https://www.g-ba.de/downloads/17-98-3049/2014-03-21_Lesehilfe-Qb.pdf=nDElXdo-sR27bLVhuoQa2g&cad=rja. (Accessed on 14 September 2015).
[27] AQUA. Qualitätsreport 2013. Göttingen: AQUA; 2014.
[28] Stausberg J, Berghof K. Qualität der stationären Versorgung in Deutschland. Eine Analyse der Entwicklung zwischen 2004 und 2008 aus Daten der externen vergleichenden Qualitätssicherung. Deutsche medizinische Wochenschrift (1946) 2014;139:181–6.
[29] Freisinger E, Fuerstenberg T, Malyar NM, Wellmann J, Keil U, Breithardt G, et al. German nationwide data on current trends and management of acute myocardial infarction: discrepancies between trials and real-life. European Heart Journal 2014;35:979–88.
[30] Kraska RA, Krummenauer F, Geraedts M. Impact of public reporting on the quality of hospital care in Germany: a controlled before-after analysis based on secondary data. Health policy (Amsterdam, Netherlands) 2016;120:770–9.
[31] Cacace M, Ettelt S, Brereton L, Pedersen J, Nolte E. How health systems make available information on service providers: experience in seven countries. Cambridge: Rand Corporation; 2011.
[32] Bundesministerium für Gesundheit. Mitglieder und Versicherte der Gesetzlichen Krankenversicherung (GKV); 2017.
[33] Copnell B, Hagger V, Wilson SG, Evans SM, Sprivulis PC, Cameron PA. Measuring the quality of hospital care: an inventory of indicators. Internal Medicine Journal 2009;39:352–60.
[34] Gandjour A, Kleinschmit F, Littmann V, Lauterbach KW. An evidence-based evaluation of quality and efficiency indicators. Quality Management in Health Care 2002;10:41–52.
[35] G-BA, Regelungen zum Qualitätsbericht der Krankenhäuser, 2013, available at https://www.g-ba.de/downloads/62-492-906/Qb-R_2014-06-19.pdf. (Accessed on 9 September 2015).
[36] Verfahren QSR. QSR-Indikatorenhandbuch 2016 für Leistungsbereiche mit Berichterstattung im AOK-Krankenhausnavigator; 2016.
[37] Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. Journal of Health Economics 2005;24:465–88.
[38] Battese GE. A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. Journal of Agricultural Economics 1997;48:250–2.
[39] Pross C, Strumann C, Geissler A, Herwartz H, Klein N. Quality and resource efficiency in hospital service provision: A geoadditive stochastic frontier analysis of stroke quality of care in Germany. Management Science 2017 (submitted).
[40] Laudicella M, Li Donni P, Smith PC. Hospital readmission rates: signal of failure or success? Journal of Health Economics 2013;32:909–21.
[41] Stausberg J, Berghof K. Quality of hospital care in Germany: analysis of the trend between 2004 and 2008 from external quality assurance data. Deutsche medizinische Wochenschrift (1946) 2014;139:181–6.
[42] G-BA. Regelungen zum Qualitätsbericht der Krankenhäuser, 2010; 2011.
[43] Maier B, Wagner K, Behrens S, Bruch L, Busse R, Schmidt D, et al. Comparing routine administrative data with registry data for assessing quality of hospital care in patients with myocardial infarction using deterministic record linkage. BMC Health Services Research 2016;16:605.
[44] Bruckel J, Liu X, Hohmann SF, Karson AS, Mort E, Shahian DM. The denominator problem: national hospital quality measures for acute myocardial infarction. BMJ Quality & Safety 2016.
[45] AQUA, Bericht zum Strukturierten Dialog 2014: Erfassungsjahr 2013, available at https://sqg.de/front_content.php?idcat=155. (Accessed on 9 September 2016).
[46] M. Dost, G., Heinze, A. Mertens, Bericht zum Strukturierten Dialog 2015: Erfassungsjahr 2014, available at https://iqtig.org/ergebnisse/strukturierter-dialog/. (Accessed on 9 October 2016).
[47] Klauber J, Geraedts M, Friedrich J, Wasem J. Krankenhaus-Report 2014: Schwerpunkt Patientensicherheit, 1st ed. Stuttgart: Schattauer G;1;mbH; 2014.
[48] Kumpunen S, Trigg L, Rodrigues R. Public reporting in health and long-term care to facilitate provider choice; 2014.
[49] Ketelaar N, Faber MJ, Flottorp S, Rygh LH, Deane KHO, Eccles MP. Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. The Cochrane Database of Systematic Reviews 2011. CD004538.
[50] James J. Health policy brief: public reporting on quality and costs. Health Affairs 2012.
[51] Dixon A, Robertson R, Appleby J, Burge P, Devlin N, Magee H. Patient choice: how patients choose and how providers respond. London: King's Fund; 2010.
[52] Boyce T, Dixon A, Fasolo B, Reutskaja E. Choosing a high-quality hospital: the role of nudges, scorecard design and information; 2016.
[53] Schwenk U, Schmidt-Kaehler S. Public reporting. Spotlight Gesundheit 2016:1.
[54] Mansky T. Was erwarten die potenziellen Patienten vom Krankenhaus? In: Böcken J, Braun B, Repschläger U, editors. Gesundheitsmonitor 2012: Bürgerorientierung im Gesundheitswesen – Kooperationsprojekt der Bertelsmann Stiftung und der BARMER GEK. Gütersloh: Bertelsmann-Stiftung; 2013. p. 136–59.
[55] Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annual Review of Public Health 2003;24:413–33.
[56] Hibbard JH, Peters E, Dixon A, Tusler M. Consumer competencies and the use of comparative quality information: it isn't just about literacy. Medical Care Research and Review MCRR 2007;64:379–94.
[57] Ocloo J, Matthews R. From tokenism to empowerment: progressing patient and public involvement in healthcare improvement. BMJ Quality & Safety 2016;25:626–32.
[58] Topol EJ. The patient will see you now: the future of medicine is in your hands. New York: Basic Books; 2015.
[59] Kumar A, Schoenstein M. Managing hospital volumes: Germany and experiences from OECD countries. OECD Health Working Papers 2013;64.
[60] Busse R, Ganten D, Huster S, Reinhardt E, Sutturp N, Wiesing U. Zum Verhältnis von Medizin und Ökonomie im deutschen Gesundheitssystem. Halle: Nationale Akademie der Wissenschaften; 2016.
[61] DKG, Bestandsaufnahme zur Krankenhausplanung und Investitionsfinanzierung in den Bundesländern, available at http://www.dkgev.de/media/file/11986.DKG_Bestandsaufnahme_KH-Planung_Investitionsfinanzierung_030812.pdf. (Accessed on 17 April 2016).
[62] de Cruppe W, Malik M, Geraedts M. Minimum volume standards in German hospitals: do they get along with procedure centralization? A retrospective longitudinal data analysis. BMC Health Services Research 2015;15:279.
[63] Nimptsch U, Peschke D, Mansky T. Mindestmengen und Krankenhaussterblichkeit – Beobachtungsstudie mit deutschlandweiten Krankenhausabrechnungsdaten von 2006 bis 2013. Germany: Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes); 2016.
[64] BT-Drs. 18/5372, Drucksache des Deutschen Bundestages 18/5372 vom 30. Juni 2015: Entwurf eines Gesetzes zur Reform der Strukturen der Krankenhausversorgung (Krankenhausstrukturgesetz – KHSG), Bundestagdrucksache (2015).
[65] AHRQ, Never Events Patient Safety Primer, available at https://psnet.ahrq.gov/primers/primer/3/never-events. (Accessed on 15 August 2016).
[66] AQUA. Bericht zur Datenvalidierung; 2013.
[67] WIdO. Qualitätssicherung der stationären Versorgung mit Routinedaten: (QSR), 1st ed. Bonn: Abschlussbericht, Wiss. Inst. der AOK; 2007.

# Kapitel 4: Nutzerverhalten auf Public Reporting Portalen am Beispiel Weisse Liste

Christoph Pross, Lars Averdunk, Josip Stjepanovic, Reinhard Busse, Alexander Geissler

Postprint, erschienen im April 2017 als: Pross C, Averdunk L, Stjepanovic J, Busse R, Geissler A. Health care public reporting utilization - user clusters, web trails, and usage barriers on Germany's public reporting portal WeisseListe.de. *BMC Medical Informatics and Decision Making*. BMC Series – open, inclusive and trusted, 2017. https://doi.org/10.1186/s12911-017-0440-6. 

**Hintergrund**: Public Reporting veröffentlicht Struktur-, Prozess- und Ergebnisindikatoren, damit Patienten und Einweiser Krankenhäuser mit guter Behandlungsqualität auswählen können. Doch nutzen Patienten die verfügbaren Daten in ihrem Entscheidungsprozess nur selten, weil das Bewusstsein über Datenverfügbarkeit und persönlichen Mehrwert fehlt oder die Darstellung nur schwer verständlich und widersprüchlich ist. Sowohl in Deutschland als auch international gibt es bisher keine Studie, die auf Basis von detaillierten Nutzerdaten eines Public Reporting Webportals analysiert, wie sich Nutzer tatsächlich verhalten, welche Informationen sie suchen und wann und warum sie die Informationssuche abbrechen.

**Ziele**: Die Arbeit analysiert die reale Nutzung von Public Reporting Portalen. Die Nutzungsentwicklung und regionalen Unterschiede werden dargestellt. Es wird geprüft, ob angebotene Qualitätsinformationen den nachgefragten entsprechen. Auf Basis des Nutzungsverhaltens werden Nutzergruppen, deren Charakteristika und Klickpfade identifiziert. Die Erkenntnisse sollen für die Verbesserung von Public Reporting genutzt werden.

**Methoden**: Die Datengrundlage bilden ca. 17 Millionen Web-Serverlog-Einträge aus dem Zeitraum Dezember 2012 bis Mai 2015 des Krankenhausvergleichsportals Weisse Liste.de. Sie enthalten Informationen über Zugriffsdatum und -zeit, Themen der aufgerufenen Seiten, Suchanfragen sowie anonymisierte IP-Adressen und Systemkonfiguration der Nutzer. Anhand dieser Daten lassen sich Besuchsverläufe erstellen und eine umfassende Datenbereinigung zum Ausschluss von fehlerhaften Anfragen durchführen. Insgesamt werden drei Analyseblöcke durchgeführt. Auf Basis von Postleitzahlen und ICD und OPS Codes analysieren wir regionales Suchverhalten und Nachfrage/Angebot der Leistungsbereiche. Mit Hilfe der Ward-Minimum-Varianz -Methode und Variablen, welche die Nutzungscharakteristiken und Interessen abbilden, erstellen wir spezifische Nutzerprofile. Mit Hilfe der Markov-Ketten-Modellierung zeigen wir die häufigsten Navigationspfade und Ausstiegspunkte auf WeisseListe.de.

**Ergebnisse**: Die Anzahl täglicher Nutzer ist von 1 445 Nutzern in 2013 auf 2 753 Nutzer in 2015 angestiegen (38% Wachstum pro Jahr). Die Anzahl der Besuche pro 1 000 Krankenhauseinweisungen haben sich im gleichen Zeitraum von 28 auf 52 fast verdoppelt. Die mobile Nutzung, die Absprungrate und der Anteil der über die Suchmaschine Google gekommenen Nutzer sind stark angestiegen, auf 25%, 38% und 52% in 2015. Der Anteil der Nutzer, die sich Qualitätsinformationen mindestens von einem Krankenhaus anschauen, lag in 2015 bei nur 48%. Die weiterführende Nutzung der Benchmarking-Funktion ist limitiert. Von den 20 meist gesuchten Diagnosen waren 15 orthopädische Diagnosen oder Krebsdiagnosen und nur für drei dieser Diagnosen gab es Information zur Ergebnisqualität. Das WeißeListe.de-Nutzungsverhalten variiert stark, mit einer stärkeren Nutzung in Märkten mit mehr Krankenhauswahlmöglichkeiten. Es konnten zehn Nutzerprofile unterschieden werden, die sich grob entlang der Dimensionen professionelle, private, unerfahrene und Expertennutzer differenzieren lassen. Nutzerorientierte Inhalte und Darstellungsformen müssen ausgebaut werden. Die Vergleichsfunktion sollte vereinfacht und gestärkt werden. Der Nutzen einer qualitätsbasierten Krankenhaussuche sollte aktiv kommuniziert werden.

Pross et al. BMC Medical Informatics and Decision Making (2017) 17:48
DOI 10.1186/s12911-017-0440-6

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

# Health care public reporting utilization – user clusters, web trails, and usage barriers on Germany's public reporting portal *Weisse-Liste.de*

Christoph Pross[1*], Lars-Henrik Averdunk[1], Josip Stjepanovic[2], Reinhard Busse[1,3] and Alexander Geissler[1]

## Abstract

**Background:** Quality of care public reporting provides structural, process and outcome information to facilitate hospital choice and strengthen quality competition. Yet, evidence indicates that patients rarely use this information in their decision-making, due to limited awareness of the data and complex and conflicting information. While there is enthusiasm among policy makers for public reporting, clinicians and researchers doubt its overall impact. Almost no study has analyzed how users behave on public reporting portals, which information they seek out and when they abort their search.

**Methods:** This study employs web-usage mining techniques on server log data of 17 million user actions from Germany's premier provider transparency portal Weisse-Liste.de (*WL.de*) between 2012 and 2015. Postal code and ICD search requests facilitate identification of geographical and treatment area usage patterns. User clustering helps to identify user types based on parameters like session length, referrer and page topic visited. First-level markov chains illustrate common click paths and premature exits.

**Results:** In 2015, the *WL.de Hospital Search* portal had 2,750 daily users, with 25% mobile traffic, a bounce rate of 38% and 48% of users examining hospital quality information. From 2013 to 2015, user traffic grew at 38% annually. On average users spent 7 min on the portal, with 7.4 clicks and 54 s between clicks. Users request information for many oncologic and orthopedic conditions, for which no process or outcome quality indicators are available. Ten distinct user types, with particular usage patterns and interests, are identified. In particular, the different types of professional and non-professional users need to be addressed differently to avoid high premature exit rates at several key steps in the information search and view process. Of all users, 37% enter hospital information correctly upon entry, while 47% require support in their hospital search.

**Conclusions:** Several onsite and offsite improvement options are identified. Public reporting needs to be directed at the interests of its users, with more outcome quality information for oncology and orthopedics. Customized reporting can cater to the different needs and skill levels of professional and non-professional users. Search engine optimization and hospital quality advocacy can increase website traffic.

**Keywords:** Public reporting, Quality transparency, Hospital quality, Provider benchmarking portal, Web usage mining, Cluster analysis, Markov chains, Clickstream analysis

* Correspondence: christoph.pross@campus.tu-berlin.de
[1]Dept. of Health Care Management, Berlin University of Technology, Administrative office H80, Str. des 17. Juni 135, 10623 Berlin, Germany
Full list of author information is available at the end of the article

## Background

Initiatives to measure and publicly report hospital quality have been implemented in many countries. They help to reduce information deficits and empower patients, their relatives, and payers to choose and contract with the most appropriate and highest quality providers. In particular, public reporting web portals are expanding rapidly in many OECD countries [1]. In the US, the CMS website *Hospital Compare* as well as several consumer reports, such as *Healthgrades.org* or *ConsumerReports.org*, provide quality of care information. In the UK, *MyNHS* and others enact the UK open data policy and the NHS quality transparency objectives. In Germany, the transparency portal *Weisse Liste.de (WL.de)* reports the results of the mandatory quality monitoring system. While *WL.de* is the leading German portal, other initiatives such as *Qualitätskliniken.de* also offer online quality of care information for participating hospitals. In total, Germany has eight portals that report hospital quality a national level and 10 portals that report quality a regional level [2].

Awareness for quality variation and treatment differences among patients is rising. In a recent representative consumer survey in Germany, half of respondents assumed that quality variation between hospitals is large [3]. In the US, a majority of people (55%) are dissatisfied with health care quality, and to compare hospital quality, they seek information on experience in certain procedures (65%), and mortality rates (57%) [4]. If information is well marketed and sought after, it appears to influence patients' hospital decisions. An analysis of the influence of the regional Hospital Guide Rhine-Ruhr in Germany found a relative increase in patient market share for hospitals that report higher than average quality [5]. In the US, higher quality hospitals have been reported to have higher market shares and to further increase their market share over time [6]. Good hospitals have an inherent self-interest in reporting quality and stimulating quality competition, as competition in regulated health systems through other dimensions (i.e. price, staffing, location) is limited [7].

Yet, public reporting expansion and optimism among policy makers are contrary to doubts among researchers and practitioners about the actual impact of public reporting. A systematic review of 150 public reporting studies found that most public reporting tools face limited usage [1]. Overall, evidence for a positive effect of public reporting on consumer behavior or quality of care is limited, and public reporting often lacks impact on the behavior of health care professionals [8]. Reported outcomes are only one aspect influencing patients' choice of hospital, with a variety of other hospital characteristics playing a substantial role as well [9]. In another US survey, only 7% of participants actually used hospital quality of care information to make health care decisions [10, 11]. In Germany, less than 20% of outpatient specialists are aware of public reporting websites and less than 10% use them actively for patient advise [12]. Causes for the often limited impact of public reporting include: complexity of quality measures, limited user-friendliness, lack of physician support and little integration into the care pathway, missing awareness of substantial quality difference between hospitals and thus motivation to search quality information and actually choose a good hospital, a mismatch between supplied and demanded information, and confusion about conflicting results on different websites for the same provider [1, 8, 13–19].

In general, studies examining health website user data are rare [1, 19], although analyzing web customer preferences is widely spread in other industries such as fashion retail and hospitality [20–22]. As the only study investigating traffic and user preferences for online public hospital quality information, Bardach et al. (2015) analyze website analytics data from a US group of hospital or physician public reporting websites and surveyed real-time visitors to these websites. Based on aggregated data (e.g. number of visitors, arrival method) and survey responses (type of respondent, purpose of visit, and website experience), they found that more than half of patients are willing to choose providers based on the information provided and health professionals generally have a better experience with public reporting than patients [23].

Past studies have been primarily based on smaller or regional patient or clinician surveys, examining changes to hospital case volumes based on reported information or only analyzed aggregated web usage data. To the best of our knowledge, there is no study that has examined in detail, based on large scale and detailed web usage data, how users actually behave on public reporting websites, which type of content they engage in, and where they abort their information search. Furthermore, most research on public reporting has been focused on a few countries, primarily the US, the UK and the Netherlands.

This paper aims to provide insights into the actual usage of online public reporting and identify public reporting improvement areas based on identified usage patterns. We first investigate whether information supplied matches patient demand and regional variations in public reporting usage. We then identify usage frequency and intensity of different portal sections and key user groups, their usage characteristics, and usage patterns. We use descriptive analyses and web mining techniques – web user clustering and first level Markov chains – on clickstream data from 17 million user actions from the *WL.de* hospital quality transparency portal from 2013 to

2015. At an overall level, we also contrast *WL.de* usage data with new and unpublished usage data from the *Hospital Compare* website.

## Methods

### *Weisse Liste* background

Annual, self-reported hospital report cards are compiled as part of the mandatory external quality monitoring system and gather structural information (such as case volumes, equipment, staff levels) across all medical specialties as well as process, outcome and risk-adjusted outcome quality indicators for 30 diseases and diagnoses, covering around 3.1 million cases or 15% of the annual case volume in Germany [24]. On behalf of several major statutory health insurance funds, the *WL.de* carries out the government mandate for the statutory health insurance (SHI) system to publically report the information in an easily accessible and patient friendly manner [25].

In 2008, the *WL.de* project was jointly initiated by the non-profit Bertelsmann Stiftung and the main patients' and consumers' associations. The portal *WL.de* has become the largest health care quality public reporting portal in Germany, consisting of a hospital, an outpatient physician and a nursing care search portal. The hospital search has gone through several development rounds, with the latest re-launch in June 2015. *WL.de* quality data is also integrated into websites of health insurance funds such as the AOK and the BARMER GEK.

Our analysis focuses on the *WL.de Hospital Search* portal (see sitemap in Fig. 1). Users search for hospital quality information by entering geographical (postcode or city) and disease, diagnosis or procedure information. Data is either entered directly or via assisted searches (via body parts or disease and procedure catalogues). Searches can be completed for 5–500 km radius or nationwide. After a valid geographical and medical information combination is entered, the website returns an initial results list of hospitals. The user can then either examine in detail a particular hospital (and generate a PDF report for the selected hospital) or initiate a benchmarking by selecting several hospitals for comparison. Other website elements present background information, current news and explanation on information relevance. The re-launch in mid-2015 has slightly changed the set-up and flow of the website compared to our observation period.

### Usage data

We received preprocessed server log files from the statistic module of the content management system Papaya CMS for 273 million server requests between 17.12.2012 and 28.05.2015 in a MySQL database dump file. We re-imported the server log files via MySQL 5.6, re-created a 100 GB MySQL database, and operated the MySQL database via MySQL Workbench. In addition, we also received cleaned web user session data from a second cookie-based tracking program (Piwik) for the period of April 2013 until April 2015, which we used to validate the cleaned Papaya CMS data.

We completed extensive data cleaning as the website is highly frequented by robots originating from search engines indexing as well as from fraudulent data siphoning. Search engine robots, with a share of 57% of all raw log file entries, are easily identified and excluded. Masked, fraudulent robots must be detected manually

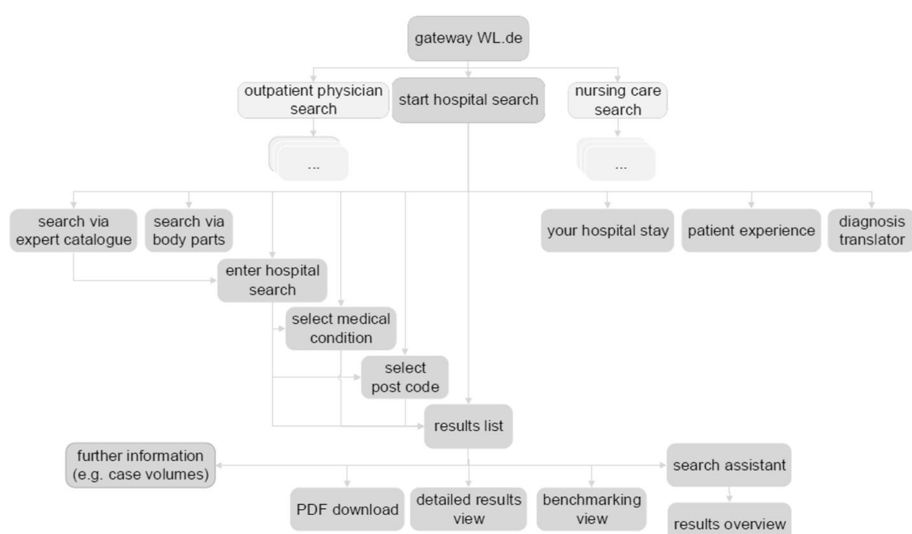



**Fig. 1** *WL.de Hospital Search* portal sitemap. Legend: Own design based on general website structure and functionality over data period from 2013–2015

by rules-based cleaning. Furthermore, we also excluded traffic generated by non-human sources, e.g. Ajax-requests and requests originating from RSS-Readers. After data cleaning, 17 million user server requests remained. In order to analyze user behavior, we reconstructed individual user sessions from the SQL server log files using established web usage mining methods [26, 27]. An example of the individual SQL server requests and the associated user sessions are displayed in Additional file 1.

The WL.de website user data is proprietary and WL.de competitive concerns as well as data usage restrictions within the public quality monitoring system do not allow data sharing beyond the limited and vetted circle of the author team. Specifically, the WL.de data privacy stipulations as well as the licensing agreement between WL.de and the SHI funds explicitly disallow data passage to external parties outside the influence of WL.de [28]. Moreover, a data usage agreement between WL.de and the author team was signed that limits the usage of this data to the scientific purposes of this study.

### Methodology

In online consumer research, clickstreams (i.e. web user trails or navigational patterns) take an increasingly important role in helping marketing professionals and researchers to uncover online consumer behavior based on large scale online shopping data. More precisely, the term “clickstream” denotes the electronic record of a user’s activity through one or more web sites and reflects a series of choices in navigating the web [20, 29, 30]. We first investigated the clickstream user session data at a more general level with descriptive methodologies for the entire data period from mid-December 2012 to end of May 2015. Afterwards, we employed e-commerce web usage mining techniques to infer detailed user patterns, usage barriers and user information gain and model user trails, or a sequence of web pages viewed by a user in a certain timeframe [31, 32]. We limited the time period for the detailed clickstream analyses to the first five months in 2015 to choose a distinct, comparable and recent time period where the site structure has not changed and to circumvent computational limitations. For the detailed analyses, we also exclude bounce visits – when users leave immediately after entering the portal [33]– and visits to the AOK and BARMER WL.de sub-portals.

The aim of the cluster analysis is to identify user groups according to their usage behavior and interests, which is captured in their clickstreams. Knowledge of the different user types, or communities, can facilitate the improvement of the website by identifying and satisfying differentiated user needs and preferences [34, 35]. For our clustering, we include two types of user information, displayed in Table 1. On the one hand, usage specific clickstream variables such as clicks per session, time of access, etc. are included. On the other hand, content related variables are taken into account. Therefore, a methodology to cluster mixed data is required to incorporate the different types of variables. To determine similarities and dissimilarities between the sessions, the Gower’s General Similarity Coefficient is used as distance measure [36]. The advantage of this measure, compared to the more frequently used Euclidean distance measure, is that Gower’s distance measure works with mixed data types and avoids an aggregation of variables and associated information loss [37].

We apply a hierarchical clustering algorithm, which is more resource intensive than the often used k-means algorithm but allows retrospective determination of cluster quantity based on stopping rules such as the Duda-Hart-Index [38] and graphical interpretation of dendrograms. Among several possible hierarchical clustering algorithms, we choose Ward’s minimum variance method as it is for the data structure fitting algorithm and capable of identifying consistent, actual user groups [37, 39–41]. Other algorithms such as the single-linkage and complete-linkage algorithms were tested and ruled out due to high outlier and data noise sensitivity [42].

To visualize navigation paths, we employ Markov chain modelling, which regards each website content area (Table 1) as a separate state and links between the topics as transitions [43]. The model contains the transition probabilities from one website topic area to another [42]. We use first-order Markov chains, where the probabilities for the next visited site depend only on the previously visited site [44]. To ensure stability of results, we ran the clustering algorithms and the Markov chain clickstream analysis multiple times for many different data samples from the first half of 2015 observation period. We also challenged the clustering and clickstream results in several workshops with different *WL.de* experts.

All analyses are conducted at an aggregated or large group level, with no individual or small user group identification. The server log files include no data privacy sensitive information. IP addresses were anonymized and used only to track returning visitors. To get access to the web portal user data, the proposed analyses and methodology were vetted by WL.de in consultation with its SHI stakeholders and found to comply with the stringent data privacy concerns [28]. Thus, our methodology and data use comply with the relevant ethical stipulation and no other approval of an additional ethics committee is required.

## Results

### Overall usage pattern

On average, *WL.de* has 10,000 daily users, of which 30% search for hospital quality information. Unique website

**Table 1** Clickstream variables and information content for clustering

| Variable | Type[1] | Mean | SD | Description |
|---|---|---|---|---|
| Number of clicks | cont | 13 | 15 | User click on website element (request) |
| Time per click | cont | 600 | 909 | Time in seconds passed between clicks |
| Successful visit | cat | 68% | | Success = view of hospital search results |
| Work time access | cat | 51% | | weekdays 9.00 am - 6.00 pm |
| Mobile device | cat | 17% | | Use of handheld device |
| Returning | cat | 22% | | Returning visitor with previous visit |
| Referrer | | | | Webpage where the user came from |
| Direct entry | cat | 21% | | *WL.de* directly entered in URL bar |
| Search engine | cat | 68% | | *WL.de* entered via search engine (e.g. google) |
| Health magazines | cat | 4% | | Patient health magazines (e.g. Apothekenumschau) |
| Health insurance | cat | 3% | | Statutory health insurance websites |
| Media | cat | 2% | | Online news sites |
| Internal link | cat | 2% | | Other *WL.de* portals (e.g. nursing care) |
| Other | cat | 1% | | |
| Website content | | | | Content visited by average user (clicks per topic) |
| Start hospital search | cont | 10% | 15% | Initiate search based on medical and geograph. info |
| Select medical condition | cont | 5% | 12% | |
| Search via body parts | cont | 3% | 11% | Select medical condition via human body part map |
| Search via catalogue | cont | 1% | 7% | Select medical condition via ICD/OPS[2] expert list |
| Select post code | cont | 1% | 3% | |
| Search results | cont | 23% | 24% | List of hospitals offering relevant care in geo area |
| Detailed results view | cont | 13% | 23% | Detailed information about one selected hospital |
| Benchmarking | cont | 1% | 4% | Direct comparison for selected criteria/hospitals |
| PDF brochure download | cont | 0% | 1% | Download info about selected hospital(s) |
| Diagnosis translator | cont | 11% | 31% | Find medical descriptions for ICD/OPS[2] codes |
| Your hospital stay | cont | 1% | 5% | |
| Patient experience | cont | 0% | 3% | Information about patient experience survey |
| Background info | cont | 0% | 3% | Background info about *WL.de* transparency project |
| Latest news | cont | 0% | 2% | |
| Sister sites | cont | 10% | 21% | Information on outpatient physicians, nursing care |

Clickstreams analyses based on 80,000 session data sample from January to May 2015 excluding bounce visits. Clustering was conducted based on 22 variables (referrer functioning as one variable); mean and standard deviation calculated for the average session in the sample
1. Variable continuous or categorical (dummy 1 = yes, 0 = no); 2. ICD = International Classification of Diseases, OPS = Operationen- und Prozedurenschlüssel (based on International Classification of Procedures in Medicine)

visits per day have increased from 1,445 in 2013 to 2,753 in 2015 (Table 2), which is an annual compound increase of 38%. In 2015, daily visits added up to 980,000 annual visits. Visits per 1,000 hospital admissions have almost doubled, from 28 in 2013 to 52 in 2015. In 2015, on average, visitors spent 6.7 min (399 s) per visit and conducted 7.4 clicks, with 54 s per click. In 2013, visitors spent substantially more time on the portal (9.4 min), conducting more clicks (10.8 clicks per visit), but taking slightly less time between clicks (52 s). Overall, time spent on the portal has decreased by 30% from 2013 to 2015. The bounce rate has increased from 22% in 2013 to 38% in 2015, which is a 73% increase and in line with increased mobile usage (11% in 2013 and 25% in 2015). During the observation period, the *WL.de* portal was not mobile-responsive, while mobile information search on *WL.de* and the general internet increased substantially [45]. The increased bounce rate can be at least partially attributed to the higher share of mobile users, which have higher bounce rates and non-successful visits (no results information).

To put the *WL.de* results into perspective, we received US CMS data on overall usage for the *Hospital Compare* portal (Table 1), where daily visits increased from 3,476 in 2013 to 3,806 visits in 2015. While absolutely still lower, usage at the *WL.de* hospital search has increased

**Table 2** Summary website traffic for WL.de and Hospital Compare hospital search 2013–2015

| | Weisse Liste.de | | | Hospital Compare | | |
|---|---|---|---|---|---|---|
| Variables | 2013 | 2014 | 2015 | 2013 | 2014 | 2015 |
| Unique visits per day | 1,445 | 2,122 | 2,753 | 3,476 | 3,072 | 3,806 |
| *Growth p.a. [%]* | - | 47 | 30 | - | −12 | 24 |
| *Visits per 1,000 hospital adm.* | 28 | 40 | 52 | 35 | 32 | 40 |
| Clicks per visit | 10.8 | 8.4 | 7.4 | 3.4 | 4.0 | 3.8 |
| Time per visit [sec] | 566 | 456 | 399 | 91 | 89 | 92 |
| Time per click [sec] | 52 | 54 | 54 | - | - | - |
| Bounce visits [%] | 22 | 32 | 38 | 37 | 32 | 34 |
| Successfull visits [%] | 66 | 53 | 48 | - | - | - |
| Mobile visits [%] | 11 | 23 | 25 | - | - | - |
| Referred via Google search [%] | 23 | 42 | 38 | - | - | - |
| Referred via Google adWords [%] | 0 | 0 | 14 | - | - | - |
| Entered directly [%] | 35 | 27 | 24 | - | - | - |

WL.de data from Q1 for each year 2013, 2014 and 2015 since data for those quarters most complete and comparable across the three years; data comparability between WL.de and Hospital Compare to the best of our knowledge; data including bounce visits

more rapidly between 2013 and 2015 than for *Hospital Compare*. Weighted by the number of hospital admissions, relative *WL.de* hospital search usage has surpassed *Hospital Compare* usage. However, in the US, many more national websites exist that report hospital quality information. In particular, *Healthgrades.com* is more commonly searched for than *Hospital Compare* [46], which implies an overall higher public reporting usage in the US than in Germany. Bounce rates for both websites are roughly equivalent and within the range of acceptable bounce rates [47]. Both clicks per visit and average time per visit are substantially longer at *WL.de*, which can be explained by the different public reporting approaches. *WL.de* reports at the medical condition, hospital and single quality indicator level (requiring more time to make the relevant selections), whereas Hospital Compare reports more generally at the aggregate hospital level, with composite information across medical conditions.

The share of hospital search users entering the website via the Google search engine has increased from 23% in 2013 to 38% in 2015 while the Google AdWords has increased from 0% in 2013 to 14% in 2015. In total, the number of daily users arriving through the Google search engine (market share of 95% in Germany) has increased from 890 in 2013 to 1,430 in 2015, which is an increase of 60% and three times the increase of use of internet search engines for consumer information search [48]. Accordingly, the share of users entering the website directly has decreased from 35% in 2013 to 24% in 2015. As a key performance indicator, the share of successful visits – users viewing hospital search results – has decreased from 66% in 2013 to 48% in 2015, which can likely be attributed to the higher share of mobile visits, as the website was not mobile responsive.

The heat map for the most recent and full year 2014 in Fig. 2 summarizes website usage based on visited website elements (topic areas). Users conduct most clicks on the initial results lists (1.1 million clicks), followed by detailed results for one hospital (0.6 million) and the hospital search entry mask (0.4 million). The initial results window is the most popular way of viewing results while the detailed and benchmarking view options are substantially less frequented. In contrast, users spent most time on the benchmarking window (170 s), followed by the detailed search results (140 s) and the results window (118 s). Once users have reached the more detailed search results, they take a longer, more detailed look.

#### Hospital quality information supplied vs. demanded

To investigate the fit between publicly reported (i.e. supplied) hospital quality information and patient demand, we determine the top 20 requested medical conditions on *WL.de* and contrast those with the outcome quality information collected and reported within the mandatory quality monitoring program (Table 3). The ranking is based on the number of user requests weighted by the 2014 disease incidence. The two primary diagnostic groups for which users want to compare hospitals are cancer and orthopedics, contributing 10 and 5 search terms to the top 20 diagnoses, respectively. While these 15 diagnoses generate substantial patient interest in public reporting of hospital quality, process or outcome quality of care indicators are only available for two orthopedic (arthrosis in knee or hip) and two cancer diagnoses

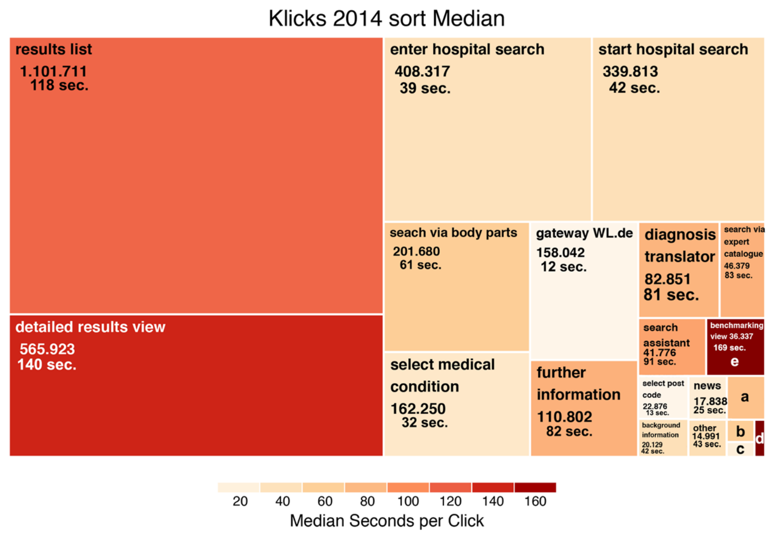



**Fig. 2** Heat map displaying number of clicks and usage intensity, fully year 2014. Legend: Own calculation. Size of rectangles captures number of clicks in topic area and color code displays average time spent in topic area. Information calculated based on 2014 data (to increase sample size). Robustness checks for full year data 2013 and 5 months' data from 2015 provide consistent results

(breast- and ovarian cancer). For highly demanded diagnoses, such as prostate, esophageal and colon cancer or spinal disc herniation, internal derangement of knee, and depression, no process or outcome quality indicators are available on *WL.de* or the mandatory quality reporting program.

## Geographical usage patterns

Figure 3 shows regional variation in public reporting usage patterns, with raw usage figures on the left. Due to population density, usage (based on search destination) is highest in the metropolitan areas Berlin, Munich and Hamburg and the state of North Rhine-

**Table 3** Top 20 diagnosis based on number of search requests weighted by incidence, 2014

Top 20 diagnosis based on number of search requests weighted by incidence, 2014

| medical condition | ICD code | numb. of searches | diagnoses, in 2014 | weighted searches[1] | search ranking | outcome indicators available |
|---|---|---|---|---|---|---|
| osteoarthritis of hip | M16 | 52,575 | 167,500 | 0.31 | 1 | ✓ |
| osteoarthritis of knee | M17 | 49,460 | 185,399 | 0.27 | 2 | ✓ |
| malignant neoplasm of prostate | C61 | 15,720 | 68,522 | 0.23 | 5 | ✗ |
| full-term uncomplicated delivery | O80 | 19,769 | 91,860 | 0.22 | 3 | ✓ |
| cervical disc disorders | M50 | 5,953 | 29,894 | 0.20 | 26 | ✗ |
| Internal derangement of knee | M23 | 17,525 | 97,990 | 0.18 | 4 | ✗ |
| benign prostatic hyperplasia | N40 | 9,349 | 57,947 | 0.16 | 12 | ✗ |
| malignant neoplasm of esophagus | C15 | 3,661 | 29,504 | 0.12 | 41 | ✗ |
| aortic aneurysm and dissection | I71 | 3,639 | 29,451 | 0.12 | 42 | ✗ |
| malignant neoplasm of breast | C50 | 15,091 | 132,926 | 0.11 | 7 | ✓ |
| malignant neoplasm of kidney | C64 | 2,416 | 23,140 | 0.10 | 59 | ✗ |
| malignant neoplasm of colon | C18 | 8,206 | 81,421 | 0.10 | 15 | ✗ |
| other intervertebral disc disorders | M51 | 15,118 | 156,893 | 0.10 | 6 | ✗ |
| acquired deformities of fingers, toes | M20 | 5,512 | 57,542 | 0.10 | 28 | ✗ |
| malignant melanoma of skin | C43 | 2,307 | 24,148 | 0.10 | 64 | ✗ |
| malignant neoplasm of rectum | C20 | 5,828 | 61,420 | 0.09 | 27 | ✗ |
| malignant neoplasm of ovary | C56 | 2,501 | 26,605 | 0.09 | 58 | ✓ |
| malignant neoplasm of pancreas | C25 | 4,566 | 48,645 | 0.09 | 36 | ✗ |
| malignant neoplasm of liver | C22 | 2,630 | 29,218 | 0.09 | 53 | ✗ |
| major depressive disorder | F32 | 11,234 | 125,623 | 0.09 | 10 | ✗ |

Note: 1. Number of searches weighted by diagnosis incidence in 2014

orthopedics cancer

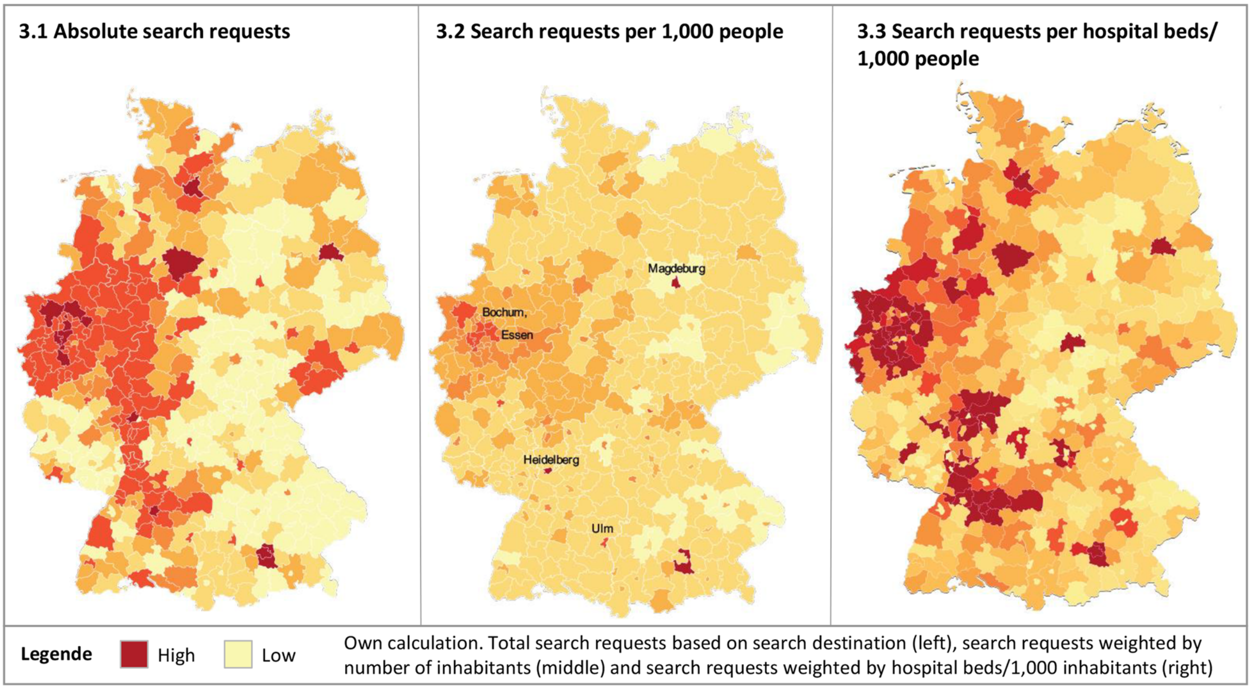



**Fig. 3** District-level maps of Germany displaying *WL.de* usage. Legend: Own calculation. Total search requests based on search destination (left), search requests weighted by number of inhabitants (middle) and search requests weighted by hospital beds/1,000 inhabitants (right)

Westphalia in West. When adjusting for number of inhabitants (middle), usage becomes more evenly distributed. However, medium-sized cities with medical universities, e.g. Magdeburg, Heidelberg und Ulm, have higher than average usage due to large catchment areas. Likewise, North Rhine-Westphalia shows higher search activities than other parts of Germany even after adjusting for population density. When adjusting for hospital capacity in each district (number of hospital beds per 1,000 inhabitants) (on the right), urban areas with lots of hospital beds stand out.

### User clustering

Results of the clustering algorithm show ten distinct user types (Table 4). These user types address similar character differentiations as previous user community studies in a health information setting [23, 49] and other e-commerce web trail studies [50, 51]. The two largest groups are the *Intensive Work Time* users (19% user share), which have 100% work time usage and a higher share of returning visitors, and the *Intensive Free Time* users (17%), which have 0% work time usage and have a small share of returning visitors. Both groups view hospital quality results in 100% of cases. On average, both groups spent 11–12 min on the website and conduct 15–16 clicks per visit. Both groups also have 100% desktop usage and all enter the *WL.de Hospital Search* via search engines.

The third largest group is the *Diagnosis Translators* (13%), which, on average, only spend 3.2 min during working hours on the website and do not view any results, but instead translate their ICD diagnosis or OPS procedure code into understandable descriptions, with a 20% share of returning users. No one of the *Diagnosis Translators* is using the hospital search function for the inquired medical condition or respective postal codes.

The fourth largest group, the *Challenged Aborts* (12%) abandon their search after only 6 clicks and 4 min on the website, without viewing any results information. All enter through search engines, one third uses a mobile device and most access the page during non-working hours. Furthermore, the *Patient Experts* (9%) access the portal directly, mostly after hours during the week. One third uses a mobile device and two thirds their desktop computers. They have the highest number of clicks (17), spent almost 16 min on the page, have a higher share of returning visitors and all of them view hospital results. Similarly, the *Professionals* (7%) spend more than 16 min on the website, conduct 16 clicks and view results in 100% of visits, but access the portal 100% during working hours and 100% through their desktop machines. Importantly, 100% of *Professionals* access the website directly and more than half of them are returning visitors (highest share of all user types).

**Table 4** User cluster and their usage characteristics

| User cluster | Share [%] | Average number clicks[2] | Average visit length [sec][2] | Time betw. clicks [sec][3] | Return visitors [%] | Viewed results [%] | Search steps/ results[1] | Visit dur. Workday [%] | Working hours [%] | Desktop usage [%] | Access via [%] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Intensive Work Timers | 19 | 15.2 | 693 | 45 | 29 | 100 | 0.45 | 100 | 100 | 100 | 100 search engine |
| Intensive Free Timers | 17 | 16.4 | 731 | 46 | 13 | 100 | 0.42 | 43 | 0 | 100 | 100 search engine |
| Diagnosis Translator | 13 | 5.7 | 182 | 32 | 19 | - | - | 100 | 100 | 100 | 100 search engine |
| Challenged Aborts | 12 | 6.1 | 255 | 48 | 8 | - | - | 53 | 12 | 69 | 100 search engine |
| Patient Experts | 9 | 16.7 | 851 | 54 | 24 | 100 | 0.33 | 63 | 12 | 66 | 100 direct |
| Curious | 7 | 14.9 | 747 | 49 | 28 | 78 | 0.60 | 83 | 53 | 83 | 35 payer, 30 media |
| Professionals | 7 | 15.9 | 884 | 53 | 56 | 100 | 0.32 | 100 | 100 | 100 | 100 direct |
| Results Mobiles | 7 | 14.5 | 696 | 49 | 8 | 100 | 0.50 | 65 | 30 | 0 | 100 search engine |
| Explorers | 4 | 13.4 | 571 | 47 | 14 | 72 | 0.67 | 77 | 48 | 80 | 100 health website |
| Other | 5 | 6.5 | 456 | 72 | 40 | | - | 79 | 42 | 74 | 100 direct |
| Average User | 100 | 12.7 | 596 | 47 | 22 | 68 | 0.42 | 76 | 51 | 83 | 67 search engine |

Clustering based on clickstream data and repeated sampling from data sample from 01/2015 – 05/2015 excluding bounce visits
1. Search steps required relative to number of results viewed; 2. all clicks or visit lengths in sec per user cluster/number of users in cluster (weighted average); 3. Calculated per session and then averaged for user cluster (simple, non-weighted average)

**Clickstream analysis**

Figure 4 displays the overall navigation trails for the *WL.de Hospital Search* portal. Most users access the hospital search portal via the *WL.de* gateway. From the hospital search entry page, 37% of users enter hospital search information correctly and go directly to the hospital results list. 48% of users require support in their hospital search, with 21% selecting a medical condition based on a drop down list, 8% via body parts display and 19% hospital search window. From the hospital search entry page, 55% of users complete the search correctly and arrive at the initial results view. 13% have to reselect a medical condition, 10% return to the body parts display and 3% re-select a post code. 10% of users exit prematurely from the hospital search entry page. When viewing the initial hospital results, 48% of users click on a specific hospital to view detailed results, 26% return to the hospital search page to change their search parameters or conduct a new search, 16% exit without viewing more detailed results, 7% of users look for more detailed explanations via the info popups and only 3% actually navigate to the benchmarking function.

A large share of users of the diagnosis translator function (76%) exit without searching for hospital quality results (i.e. the *Diagnosis Translators*). Likewise, a significant share of users exits directly from pages with additional information, such as background info (13%) and info popups (20%). Furthermore, a considerable share of users exits during the assisted search process (10% from the hospital search entry page, 9% while using the body parts display and 7% while selecting a medical condition from the drop down.

Combining the user cluster and clickstream methodology, Additional file 2: Figure S1, Additional file 3: Figure S2, Additional file 4: Figure S3, Additional file 5: Figure S4 in the Additional files separately display the clickstreams for important user clusters. The *Intensive Work Timers* as the largest cluster display a similar navigation pattern as the patterns described above for the average user. However, the *Patient Experts* as well as the *Professionals* use less frequently the assisted search

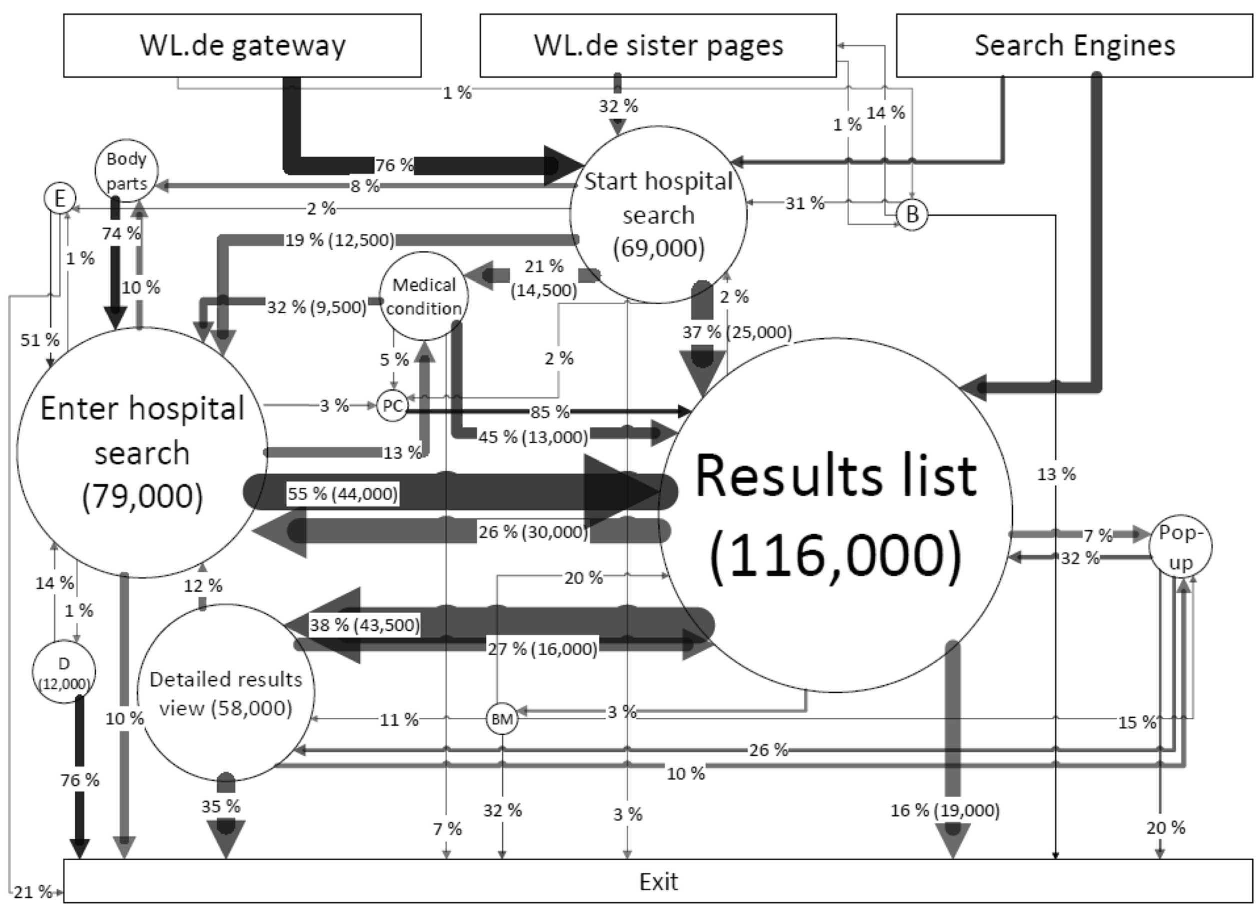



**Fig. 4** Overall navigation trails for all users. Legend: bubble size = clicks per topic area, arrow width = absolute number of transitions for entire page, arrow grayscale = share of transitions away from topic area (i.e. bubble), B = background information, BM = benchmarking view, D = diagnosis translator, PC = select post code, E = expert catalogue (ICD/OPS lists), Pop-up = detail information pop-up

functionalities. The *Challenged Aborts* display a very erratic navigational pattern and often return to a previous node, exit often from the assisted search function, the hospital search entry page, and the additional information pages, and often return to the *WL.de* gateway or sister pages without completing a hospital search.

## Discussion

*WL.de Hospital Search* usage has increased substantially, up to 2,753 daily users in 2015. Compared to 2013, users have spent less time on the webpage and more frequently not requested any hospital quality information. Relative to *Hospital Compare*, *WL.de* usage has shown a stronger growth in usage. However, since the US has several other equally or even more popular public reporting sites such as *Healthgrades.com* public reporting experiences higher usage in the US. But public reporting usage in Germany is catching up. The *WL.de* traffic growth far outpaced the overall growth of internet users in Germany [45]. As illustrated by the heat map, the more *WL.de* detailed results formats (individual hospital details view or benchmarking view) receive substantially fewer visits (only 51% and 3% relative to the 1.1 million clicks on the results page, respectively), but usage intensity is substantially higher (19% and 43% more time relative to the 118 s on the results view, respectively).

The demand vs. supply analysis has revealed a gap between hospital quality information demanded by patients on WL.de and quality indicator information provided by the quality monitoring system (Table 3). The most-searched-for diagnostic categories, for which outcome quality information is missing, are prostate, esophageal and colon cancer and the orthopedic diagnoses spinal disc herniation and internal derangement of knee, as well as depressive disorders. The lack of relevant outcome information can hinder the acceptance of public reporting as users do not find information they are looking for. Comparing usage across geographic areas, people living in Western German regions, especially the state North-Rhine-Westphalia, show a particular affinity for public reporting. One contributing factor could be the higher awareness of public reporting and the quality difference between providers, due to regular publication of the Hospital Guide Rhine-Ruhr [5, 52], which is one of the earliest public reporting products target at the general public. Another contributing factor could be higher hospital density and thus more choice relative to other states [53].

The different cluster and click chains illustrate substantial variation in user interests and behaviors, indicating both the need to provide flexibility in information access, type and detail and overall improvement potential for public reporting. On the one hand, a substantial share of users does not view any hospital results information (32%) and, on the other hand, many users do not view more detailed and possibly more informative benchmarking or detailed single hospital information. Referrer and amount of time spent on the webpage as well as interest in background and explanatory information vary among clusters.

Public reporting is supposed to encourage patients to choose high quality providers. Provider selection is also what fuels quality competition among providers and drives improvements through changes in care [54]. Since public reporting should be the basis of provider selection and the quality improvement pathway, ineffective public reporting has important consequences. Optimizing public reporting has two primary elements. *Onsite*, the right content needs to be presented in the best format and detail level for different user groups and their navigation patterns. *Offsite*, web traffic management needs to be optimized to ensure maximum traffic via search engine optimization and increased awareness of the benefits and functionality of public reporting via media communication and expert commentary.

### Onsite

The cluster analysis illustrates different usage patterns and interests for the various user groups. Different user demographics and purposes require different types and detail levels of information. For example, elderly patients or those with lower levels of education generally have more difficulty in understanding comparative health information [55, 56] and thus have distinct information needs. Certain patient groups, such as younger, highly educated, or higher income patients or patients without previous satisfactory provider interaction, have been found to search more actively for a provider [8]. While the web analytics data does not provide demographic information, a separate, 2015 *WL.de* onsite user survey sheds light on user demographics. One third of *WL.de* users are above 60 years of age and another third between 50 and 60. Next to professional and personal use, 25% of users help family members in their hospital choice. A large share of users (42%) came to the portal not having chosen a provider yet.

A site that is flexible to adapt to these differences is more likely to provide information that users want [57]. The *WL.de* portal already is an interactive website that allows personalized searches based on user background (geographical and medical information). But public reporting needs to provide more flexible and customizable search and output displays to allow different user types to navigate the page and information based on their preferences and skills levels.

An important user differentiation is the professional (outpatient physicians, health advisor at insurance funds, patient advocates) vs. patient perspective. Our clustering results show that at least 7% of users can be classified as *Professionals*. In addition, a large share of users in the *Intensive Work Time* (19%) and *Diagnosis Translators* (13%) groups also have professional backgrounds. In a *WL.de* onsite survey, 24% of users identify themselves as professional users. Professionals and patients have different requirements for technical vs. non-technical information and presentation types. Even among professional users, different technical backgrounds and the ability to take in, process, and communicate information exist. Finding the right way to address *Professionals* is critical for public reporting, as admitting physicians play a large role in patients' hospital choices, but still harbor substantial skepticism and resistance towards public reporting.

Specialists often question the credibility and usefulness of outcome data [58]. Similarly, general physicians often have a negative view of public reporting, primarily due to risks of insufficient risk-adjustments, oversimplification and patient skimming by providers [59]. Public reporting usage among specialists is limited [12]. The *WL.de* portal currently has no feature to separately address expert physician users, e.g. in tailored micro site. However, if public reporting differentiates more thoroughly between professional and non-professional users, information search, display, cognitive aids, interpretation and transfer can be more customized [23].

More customized or even personalized websites could streamline and ease the information search process for physicians, but also for patients, as returning visitors will often search for similar information (e.g. same geographic area). This information can be preselected in their personalized profiles (accessible via login). More generally, three hospital search entry buttons for new and experienced patient and professional users and customized search paths, information display and detail level can create customized public reporting.

The individual value of public reporting can be approximated by user behavior, e.g. whether the information is considered superficially or in detail. Our results show that few users navigate to the detailed result view options (detail or benchmarking view), but these website areas experience the most intensive engagement. Furthermore, 560 daily users abort the search before viewing hospital quality results, exiting prematurely from website elements such as the search function or background information.

Research consistently finds that in complex and uncertainty decision environments, consumers often make better evaluations and decisions when they are presented with less information and options about their choices. Furthermore, across display-response studies in the relevant health care literature, numerical formats that included extensive text were generally less effective than simple, more visual formats such as graphs or familiar icons [56]. Limiting consumers' choice menu to the most relevant options, via geographical filtering or additional filters such as a predetermined quality filter, can support active decision making. Likewise, ranking information can improve comprehension, particular with older patients, make options easier to assess and reduce faulty data interpretations. In general, public reporting has many applications to using nudges to guide better decision making [14].

More broadly, if consumers have a general understanding of the overall paradigm (i.e. quality difference between hospitals), they will more likely understand smaller pieces of information and integrate them into their decision process [60]. Consumers in health care lack an understanding of what a choice might actually mean, once the decision is carried out [61]. Getting patients to form awareness of the benefits of active hospital choice and choice preferences prior to their actual choice helps to simplify and improve choice processes [62]. This implies that public reporting also has a role in generating more general awareness of quality variation between hospitals and benefits of hospital choice.

### Offsite

When examining public reporting optimization potential offsite, the three primary levers are search engine optimization, expert content placement and user-orientation of quality measurement. In 2013 and 2014, *WL.de* portal was optimized for search engines (on- and off page), which increased the share of Google referrers from 23% in 2013 to 38% in 2015 (Table 2). In particular, website URLs were made more distinguishable (e.g. by including hospital names) and website metadata, which Google uses for search referencing, was individualized, by e.g. changing the reference from "Weisse Liste hospital search – detail profile" to "[hospital name] in Berlin – Weisse Liste". Tagging specific hospital names increases hit rate and relevance for users and overall traffic. Additionally, at the end of 2014 *WL.de* started to use Google Grants, the non-profit edition of Google AdWords, to advertise its hospital search. This led to a substantial share of users clicking on sponsored links – combining the terms hospital search and the requested city – at the top of the search results (14% in 2015). This also allows regional targeting to potentially increase public reporting in, e.g., areas where hospitals are possibly consistently underperforming or public reporting usage is low.

Public reporting websites can also increase their traffic via promoting expert content placement and associated media messaging. In November and December 2014, a regional German television station, the Hessian Broadcasting Corporation, ran a *WL.de*-supported program on quality in five large Hessian hospitals, such as the University Hospital Frankfurt. A central part of the program was *WL.de* quality data, which was explained by a *WL.de* expert. Furthermore, a short film promoting the *WL.de* hospital search was shown. During the first two weeks of the programming, *WL.de Hospital Search* traffic was

30% higher (3,365 visits per day on average) than during the two weeks before the first show on November 19th 2014. Similarly, the AOK-Hospital Report 2014, released on January 21st 2015, included an article about substantial medical errors in Germany. The extensive media attention also covered the AOK and *WL.de* hospital quality search portals and led to a usage spike, with 50% increased daily average traffic in two weeks after publication.

Orienting mandatory quality measurement schemes more towards the medical conditions and information users are actually searching for also increases relevance and usage of public reporting portals. Currently, patients search hospital quality information for many medical conditions for which no outcome quality indicators are available. Less than 30% of inpatient care is covered by the mandatory quality reporting [2, 63] and outcome quality information for many highly sought after oncological and orthopedic conditions are missing. Like any other industry, health care public reporting needs to identify and primarily address the needs of patients as the customers of health care provision.

### Limitations

With regards to data and methodology, we consider some shortcomings. Server log-based user tracking, as opposed to cookie-based user tracking, relies on user IP addresses, which can change due to rooter re-start or service provider maintenance. Servers can also fail to account for requests that are cached by the users' computer or proxy servers or information might be lost in communication with the client [20]. The return user tracking had to be completed manually, as the automatic user tracking via Papaya was not activated while the log files were saved. However, comparison between our server-log-based user tracking and the Piwik cookie-based user tracking showed high consistency. Analyzing web usage data often faces the challenge of changing web-site structure and content; however, for the more detailed clustering and clickstream analysis we consider a narrower timeframe with no major structural changes and we use web design predefined topic categories that remain consistent even if content within these topic changes.

As a general methodological limitation, our approach of using clickstream data (as opposed to user survey data or experiments) does not allow a clear view on what users do after they leave the webpage, like whether they actually use the information to make a decision. Furthermore, we cannot deduce what users feel or experience while using the webpage. Combining clickstream with survey response data from the same users might serve as a solution here. High dimensionality clustering (in our case 22 variables) can at times provide non-logical, impractical results; however, we verified the clustering by confirming a priori hypotheses on typical user characteristics with the revealed characteristics of our user groups and extensive discussion of user group characteristics with multiple *WL.de* experts.

## Conclusions

Presenting public reporting information in a way that is most accessible for users can help to enhance the role of quality of care in treatment and hospital decisions, leading to better outcomes for patients. Public reporting promises to affect health care markets through the individual and collective informed choice of health care consumers. However, non-professionals often find it difficult to utilize quality data as information is often complex and the decisions carry high risks. Therefore, patients seek easily accessible and understandable information to make informed choices. For public reporting to realize its promise, further efforts need to be undertaken to provide context on the need of and motivation for quality of care information usage, simplify and enhance reporting portals; provide flexible, customized or even personalized usage options; offer quality information that is demanded by users; and embed quality of care information in the treatment pathway. This is especially true, since, compared with other consumer choices, health care and hospital choice decisions are complex and involve a high degree of uncertainty.

Additional research is needed to understand large sample, actual web user response to different information displays, content and detail levels. Compartmentalizing public reporting websites and monitoring user response to design and content changes can deliver real world data on what works best to engage users and facilitate their hospital choice and professional recommendations.

## Additional files

**Additional file 1:** Data example. This supplementary material includes raw data SQL requests for two user sessions and the associated user sessions that were created based on the raw data. A short data explanation describes the data and how it was used to create user sessions. (DOCX 177 kb)

**Additional file 2: Figure S1.** Navigational trail for user group Intensive Work Timers (19% of users). The figure depicts the navigation trail for the specific user subgroup Intensive Work Timers, indicating clicks per topics area (bubble size), absolute number of transitions (arrow width) and share of transitions away from respective topic areas (arrow grayscale). (TIF 875 kb)

**Additional file 3: Figure S2.** Navigational trail for user group Patient Experts (9% of users). The figure depicts the navigation trail for the specific user subgroup Patient Experts, indicating clicks per topics area (bubble size), absolute number of transitions (arrow width) and share of transitions away from respective topic areas (arrow grayscale). (TIF 726 kb)

**Additional file 4: Figure S3.** Navigational trail for user group Professionals (7% of users). The figure depicts the navigation trail for the specific user subgroup Professionals, indicating clicks per topics area (bubble size), absolute number of transitions (arrow width) and share of transitions away from respective topic areas (arrow grayscale). (TIF 86 kb)

**Additional file 5: Figure S4.** Navigational trail for user group Challenged Aborts (12% of users). Description: The figure depicts the navigation trail for the specific user subgroup Challenged Aborts, indicating clicks per topics area (bubble size), absolute number of transitions (arrow width) and share of transitions away from respective topic areas (arrow grayscale). (TIF 784 kb)

#### Abbreviations
AOK: Allgemeine Ortskrankenkasse; CMS: Centers for medicare and medicaid services; ICD: International classification of diseases; OPS: Operationen- und Prozedurenschlüssel; Papaya CMS: Papaya content management system; RSS-Readers: Really-simply-syndication reader; UK: United Kingdom; US: United States of America; WL.de: Weisse Liste.de

#### Acknowledgements
We thank Prof. Tom Rice, professor at the Department for Health Policy and Management at the University of California, Los Angeles for helpful comments on earlier versions of this article and his support in getting Hospital Compare usage data through a Freedom of Information Act Request from the Centers for Medicare and Medicaid Services (CMS). We also thank CMS for providing the data. We also thank Hannah Wehling, Weisse Liste gGmbH, for her helpful comments on the article.

## Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Funding
CP is supported by a PhD scholarship from the Konrad-Adenauer-Foundation.

#### Availability of data and materials
The data that support the findings of this study are available from the joint project team TU Berlin, Department of Health Care Management and Weisse Liste gGmbH, but restrictions apply to the availability of these data (due to data privacy and competition concerns), which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of Weisse Liste gGmbH.

#### Authors' contributions
Lead authors were CP and LA. CP initiated and drafted the study idea, outline and implementation strategy. CP also outlined, wrote and revised the article that is being submitted. LA prepared and analyzed the *WL.de* data and contributed to the writing of the article. JS managed the data extraction, transfer and explanation for the *WL.de* and contributed to the writing of the article. RB and AG supported the study design and methods selection, methodologies and contributed to the writing of the article. Each author has read approved the final version of this article.

#### Competing interests
The authors declare that they have no competing interests.

#### Consent for publication
Not applicable.

#### Ethics approval and consent to participate
All analyses are conducted at an aggregated or large group level, with no individual or small user group identification. The server log files include no data privacy sensitive information. IP addresses were anonymized and used only to track returning visitors. To get access to the web portal user data, the proposed analyses and methodology were vetted by WL.de in consultation with its statutory health insurance stakeholders and found to comply with the stringent data privacy concerns. Thus, our methodology and data use comply with the relevant ethical stipulation and no other approval of an additional ethics committee is required.

#### Author details
[1]Dept. of Health Care Management, Berlin University of Technology, Administrative office H80, Str. des 17. Juni 135, 10623 Berlin, Germany. [2]Weisse Liste gGmbH, Leipziger Straße 124, 10117 Berlin, Germany. [3]European Observatory on Health Systems and Policies, WHO European Centre for Health Policy, Eurostation (Office 07C020), Place Victor Horta/Victor Hortaplein 40/10, 1060 Brussels, Belgium.

Received: 29 November 2016 Accepted: 4 April 2017
Published online: 21 April 2017

#### References
1. Kumpunen S, Trigg L, Rodrigues R. Public reporting in health and long-term care to facilitate provider choice. 2014. http://www.euro.who.int/en/about-us/partners/observatory/publications/policy-briefs-and-summaries/public-reporting-in-health-and-long-term-care-to-facilitate-provider-choice. Accessed 17 Aug 2016.
2. Emmert M, Hessemer S, Meszmer N, Sander U. Do German hospital report cards have the potential to improve the quality of care? Health Policy. 2014; 118:386–95. doi:10.1016/j.healthpol.2014.07.006 .
3. Schwenk U, Schmidt-Kaehler S. Public Reporting: Transparenz über Gesundheitsanbieter erhöht Qualität der Versorgung. Spotlight Gesundheit. 2016;1. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/SpotGes_PubRep_dt_final_web.pdf. Accessed 17 Aug 2016.
4. Kaiser Family Foundation. National survey on consumers' experiences with patient safety and quality information. 2004. http://kff.org/health-costs/poll-finding/national-survey-on-consumers-experiences-with-patient/. Accessed 23 Aug 2016.
5. Wübker A, Sauerland D, Wübker A. Does better information about hospital quality affect patients' choice? Empirical findings from Germany. MPRA Working Paper. 2008;10479. https://mpra.ub.uni-muenchen.de/10479/1/MPRA_paper_10479.pdf. Accessed 12 Oct 2016.
6. Chandra A, Finkelstein A, Sacarny A, Syverson C. Healthcare Exceptionalism? Performance and Allocation in the U.S. Healthcare Sector 2015. doi:10.3386/w21603 .
7. Culyer AJ, Pauly MV, editors. Handbook of health economics. Amsterdam: Elsevier; 2000.
8. Ketelaar N, Faber MJ, Flottorp S, Rygh LH, Deane KHO, Eccles MP. Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. Cochrane Database Syst Rev. 2011:CD004538. doi:10.1002/14651858.CD004538.pub2 .
9. Victoor A, Delnoij DMJ, Friele RD, Rademakers JJ, Jany JDJM. Determinants of patient choice of healthcare providers: a scoping review. BMC Health Serv Res. 2012;12:272.
10. Lindenauer P. Public reporting and pay-for-performance programs in perioperative medicine: are they meeting their goals? Cleve Clin J Med. 2009;76 Suppl 4:S3–8. doi:10.3949/ccjm.76.s4.01 .
11. Shekelle P, Lim Y-W, Mattke S, Damberg C. Does public release of performance results improve quality of care? 2008. http://www.health.org.uk/sites/health/files/DoesPublicReleaseOfPerformanceResultsImproveQualityOfCare.pdf.
12. Hermeling P, Geraedts M. Kennen und nutzen Ärzte den strukturierten Qualitatsbericht? Gesundheitswesen. 2013;75:155–9. doi:10.1055/s-0032-1321744 .
13. Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. JAMA. 2000;283:1866–74.
14. Boyce T, Dixon A, Fasolo B, Reutskaja ECHOOSINGA. High-quality hospital. 2010.
15. Austin JM, Jha AK, Romano PS, Singer SJ, Vogus TJ, Wachter RM, Pronovost PJ. National hospital ratings systems share few common scores and may generate confusion instead of clarity. Health Aff (Millwood). 2015;34:423–30. doi:10.1377/hlthaff.2014.0201 .
16. Thielscher C, Antoni B, Driedger J, Jacobi S, Krol B. Geringe Korrelation von Krankenhausführern kann zu verwirrenden Ergebnissen führen. Gesundheitsökon Qualitätsmanage. 2014;19:65–9. doi:10.1055/s-0033-1335362 .
17. Mannion R, Goddard M. Public disclosure of comparative clinical performance data: lessons from the Scottish experience. J Eval Clin Pract. 2003;9:277–86.

18. Schwartz LM, Woloshin S, Birkmeyer JD. How do elderly patients decide where to go for major surgery? Telephone interview survey. BMJ. 2005;331: 821. doi:10.1136/bmj.38614.449016.DE .
19. Moser A, Korstjens I, van der Weijden T, Tange H. Themes affecting health-care consumers' choice of a hospital for elective surgery when receiving web-based comparative consumer information. Patient Educ Couns. 2010; 78:365–71. doi:10.1016/j.pec.2009.10.027 .
20. Bucklin RE, Sismeiro C. Click here for internet insight: advances in clickstream data analysis in marketing. J Interact Mark. 2009;23:35–48. doi:10.1016/j.intmar.2008.10.004 .
21. Santos BD, Hortaçsu A, Wildenbeest MR. Testing models of consumer search using data on web browsing and purchasing behavior. Am Econ Rev. 2012;102:2955–80. doi:10.1257/aer.102.6.2955 .
22. Cezar A, Ögüt H. Analyzing conversion rates in online hotel booking. Int J Contemp Hospitality Mngt. 2016;28:286–304. doi:10.1108/IJCHM-05-2014-0249 .
23. Bardach NS, Hibbard JH, Greaves F, Dudley RA. Sources of traffic and visitors' preferences regarding online public reports of quality: web analytics and online survey results. J Med Internet Res. 2015;17:e102. doi:10.2196/jmir.3637 .
24. IQTIG. Qualitätsreport 2015. 2016. https://iqtig.org/downloads/ergebnisse/qualitaetsreport/IQTIG-Qualitaetsreport-2015.pdf.
25. Bundestag D. Gesetzliche Krankenversicherung § 136b Beschlüsse des Gemeinsamen Bundesausschusses zur Qualitätssicherung im Krankenhaus. 2016.
26. Markov Z, Larose DT. Data mining the web. Hoboken, NJ, USA: Wiley; 2007.
27. Facca FM, Lanzi PL. Mining interesting knowledge from weblogs: A survey. Data Knowl Eng. 2005;53:225–41. doi:10.1016/j.datak.2004.08.001 .
28. WeisseListe.de. Datenschutzerklärung. 2016. https://weisse-liste.de/de/informationen/datenschutz/. Accessed 12 Jan 2017.
29. Bucklin RE, Lattin JM, Ansari A, Gupta S, Bell D, Coupey E, Little JDC, et al. Choice and the internet: from clickstream to research stream. Mark Lett. 2002;13:245–58. doi:10.1023/A:1020231107662 .
30. Kalczynski PJ, Senecal S, Nantel J. Predicting on-line task completion with clickstream complexity measures: A graph-based approach. Int J Electron Commer. 2006;10:121–41. doi:10.2753/JEC1086-4415100305 .
31. Borges J, Levene M. Evaluating variable-length markov chain models for analysis of user web navigation sessions. IEEE Trans Knowl Data Eng. 2007; 19:441–52. doi:10.1109/TKDE.2007.1012 .
32. Das R, Turkoglu I. Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method. Expert Syst Appl. 2009;36:6635–44. doi:10.1016/j.eswa.2008.08.067 .
33. Bucklin RE, Sismeiro C. A model of web site browsing behavior estimated on clickstream data. J Mark Res. 2003;40:249–67. doi:10.1509/jmkr.40.3.249.19241 .
34. Paliouras G, Papatheodorou C, Karkaletsis V, Spyropoulos C. Clustering the users of large web sites into communities. In: Langley P, editor. Proceedings of the seventeenth international conference on machine learning (ICML-2000), June 29-July 2, 2000, Stanford University. San Francisco, Calif: Morgan Kaufmann Publishers; 2000. p. 719–26.
35. Hoebel N, Zicari RV. On clustering visitors of a web site by behavior and interests. In: Węgrzyn-Wolska KM, Szczepaniak PS, editors. Advances in intelligent web mastering: Proceedings of the 5th Atlantic Web Intelligence Conference–AWIC'2007, Fontainebleau, France, June 25–27, 2007. Berlin, New York: Springer; 2007. p. 160–7. doi:10.1007/978-3-540-72575-6_26 .
36. Gower JC. A general coefficient of similarity and some of its properties. Biometrics. 1971;27:857–71. doi:10.2307/2528823 .
37. Backhaus K, Erichson B, Plinke W, Weiber R. Multivariate Analysemethoden. 13th ed. Berlin: Springer-Verlag, Berlin and Heidelberg GmbH & Co. KG; 2010.
38. Duda RO, Hart PE. Pattern classification and scene analysis. New York: Wiley; 1973.
39. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc. 1963;58:236–44. doi:10.1080/01621459.1963.10500845 .
40. Punj G, Stewart DW. Cluster analysis in marketing research: review and suggestions for application. J Mark Res. 1983;20:134–48. doi:10.2307/3151680 .
41. Grabmeier J, Rudolph A. Techniques of cluster algorithms in data mining. Data Min Knowl Disc. 2002;6:303–60. doi:10.1023/A:1016308404627 .
42. Liu B. Web data mining: exploring hyperlinks, contents, and usage data. 2nd ed. Berlin: Springer; 2011.
43. Sarukkai RR. Link prediction and path analysis using Markov chains. Comput Netw. 2000;33:377–86. doi:10.1016/S1389-1286(00)00044-X .
44. Ibe OC. Markov processes for stochastic modeling. London: Elsevier; 2013.
45. Koch W, Frees B. Dynamische Entwicklung bei mobiler Internetnutzung sowie Audios und Videos. Media Perspektiven. 2016:418–37. http://www.ard-zdf-onlinestudie.de/fileadmin/Onlinestudie_2016/0916_Koch_Frees.pdf. Accessed 5 Oct 2016.
46. Huesch MD, Currid-Halkett E, Doctor JN. Public hospital quality report awareness: evidence from National and Californian Internet searches and social media mentions, 2012. BMJ Open. 2014;4:e004417. doi:10.1136/bmjopen-2013-004417 .
47. Peyton J. What's the average bounce rate for a website? 2014. http://www.gorocketfuel.com/the-rocket-blog/whats-the-average-bounce-rate-in-google-analytics/. Accessed 23 Oct 2016.
48. VuMA. Konsumenten punktgenau erreichen. 2016. https://www.vuma.de/fileadmin/user_upload/PDF/berichtsbaende/VuMA_2017_Berichtsband.pdf.
49. Huntington P, Nicholas D, Williams P. Characterising and profiling health Web user and site types: Going beyond "hits". AP. 2003;55:277–89. doi:10.1108/00012530310498851 .
50. Moe WW. Buying, searching, or browsing: differentiating between online shoppers using In-store navigational clickstream. J Consum Psychol. 2003;13: 29–39. doi:10.1207/S15327663JCP13-1&2_03 .
51. Hong T, Kim E. Segmenting customers in online stores based on factors that affect the customer's intention to purchase. Expert Syst Appl. 2012;39: 2127–31. doi:10.1016/j.eswa.2011.07.114 .
52. Ruhrgebiet I. Klinkführer Rhein-Ruhr 2005/2006. 1st ed. Essen, Ruhr: Klartext; 2005.
53. Klein-Hitpaß U, Leber W-D, Scheller-Kreinsen D. Strukturfonds: Marktaustrittshilfen für Krankenhäuser. G + G Wissenschaft. 2015;15:15–23.
54. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. Med Care. 2003;41:I30–8.
55. Hibbard JH, Slovic P, Peters E, Finucane ML, Tusler M. Is the informed-choice policy approach appropriate for medicare beneficiaries? Health Aff. 2001;20: 199–203. doi:10.1377/hlthaff.20.3.199 .
56. Kurtzman ET, Greene J. Effective presentation of health care performance information for consumer decision making: A systematic review. Patient Educ Couns. 2016;99:36–43. doi:10.1016/j.pec.2015.07.030 .
57. Vaiana ME, McGlynn EA. What cognitive science tells us about the design of reports for consumers. Med Care Res Rev. 2002;59:3–35.
58. Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists. N Engl J Med. 1996;335:251–6. doi:10.1056/NEJM199607253350406 .
59. Casalino LP, Alexander GC, Jin L, Konetzka RT. General internists' views on pay-for-performance and public reporting of quality scores: a national survey. Health Aff (Millwood). 2007;26:492–9. doi:10.1377/hlthaff.26.2.492 .
60. Greene J, Peters E, Mertz CK, Hibbard JH. Comprehension and choice of a consumer-directed health plan: an experimental study. Am J Manag Care. 2008;14:369–76.
61. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health. 2003;24:413–33. doi:10.1146/annurev.publhealth.24.100901.141005 .
62. Chernev A. When more is less and less is more: the role of ideal point availability and assortment in consumer choice. J Consum Res. 2003;30:170–83. doi:10.1086/376808 .
63. G-BA. Die gesetzlichen Qualitätsberichte 2012 der Krankenhäuser lesen und verstehen. 2014. https://www.g-ba.de/downloads/17-98-3049/2014-03-21_Lesehilfe-Qb.pdf=nDElXdo-sR27bLVhuoQa2g&cad=rja. Accessed 14 Sep 2015.

# Kapitel 5: Spezialisierung und Zertifizierung im Krankenhaus

Christoph Pross, Elke Berger, Alexander Geissler, Martin Siegel, Reinhard Busse

Preprint, zur Begutachtung eingereicht August 2017 bei Intern. Journal for Quality in Health Care (Oxford Academic): Pross C, Berger E, Geissler A, Siegel M, Busse R. Stroke Units, Certification, and Stroke Outcomes in German Hospitals: A Fixed Effects Model for Stroke Mortality from 2006 – 2014.

**Hintergrund**: Die Behandlung von Schlaganfallpatienten in spezialisierten Stroke Units (SU) hat in den letzten Jahren deutlich zugenommen, doch werden mehr als 20% aller Schlaganfallpatienten noch immer in Krankenhäusern ohne SU behandelt. International lassen sich hohe Unterschiede in der Schlaganfallmortalität feststellen; auch in Deutschland variiert die risiko-adjustierte 30-Tage Sterblichkeitsrate um ein Vielfaches zwischen den Krankenhäusern. Zwar konnten einige Studien einen positiven Zusammenhang zwischen SU-Behandlung und Ergebnissen festhalten, doch haben diese Studien oft Limitationen bezüglich Datenumfang, Zeitreihe und Aussagekraft der statistischen Modelle. Zusätzlich gibt es keine Studie, welche Behandlungsergebnisse nach Krankenhäusern mit SU, Krankenhäusern mit SU und SU-Zertifizierung und Krankenhäusern mit SUs und gesamtheitlicher Qualitätsmanagement-Zertifizierung unterscheidet.

**Ziele**: Auf Basis des Donabedianischen Qualitätsmodells untersuchen wir die Auswirkungen von Struktur- und Prozessspezialisierung in einer SU auf die Schlaganfallversorgungsqualität und evaluieren so den Nutzen von substantiellen Investitionen in SUs auf Krankenhausebene und eine zentralisierte Behandlung aller Stroke Patienten in SUs auf Versorgungssystemebene. Zusätzlich untersuchen wir, ob die Ergebnisqualität in Krankenhäusern mit einer SU-

Zertifizierung der Deutschen Schlaganfallgesellschaft (DSG) und in Krankenhäusern mit einem Qualitätsmanagementzertifikat der Kooperation für Transparenz und Qualität im Gesundheitswesen (KTQ) signifikant besser ist als in Krankenhäusern, die diese Zertifikate nicht haben. Neben dem möglichen Signaleffekt kann eine bessere Qualität durch Zertifizierung die finanziellen und organisatorischen Investitionen in Zertifikate rechtfertigen.

**Methoden:** Als Ergebnisindikator wird die 30-Tage risikoadjustierte Mortalitätsrate der Initiative Qualitätssicherung mit Routinedaten für die Jahre 2006 – 2014 genutzt. Dazu werden relevante Strukturdaten der Qualitätsberichte der deutschen Krankenhäuser und DSG und KTQ Zertifizierungsdaten verwendet. Krankenhäuser mit SU werden anhand der abgerechneten Schlaganfallkomplexbehandlungen identifiziert. Wir schätzen ein Fixed-Effects-Modell mit Dummy Variablen für SU Spezialisierung, DSG Zertifizierung und von KTQ Zertifizierung und den entsprechenden Kontrollvariablen.

**Ergebnisse:** Die Anzahl der Krankenhäuser mit SU hat sich von 276 in 2006 auf 436 in 2014 erhöht. Davon haben in 2014 219 Krankenhäuser eine DSG Zertifizierung und 74 sowohl ein DSG als auch ein KTQ Zertifikat. Deskriptive Analysen lassen eine sukzessive höhere Qualität in Krankenhäusern mit SU, in Krankenhäusern mit SU und DSG Zertifikat und in Krankenhäusern mit SU, DSG Zertifikat und KTQ Zertifikat vermuten. Doch lässt sich im Regressionsmodell nur ein statistisch signifikanter positiver Qualitätseffekt für die Behandlung in einem Krankenhaus mit SU feststellen. Die 30-Tage risikoadjustierte Sterblichkeit ist um 5.6% niedriger in Krankenhäusern mit SU als in Krankenhäusern ohne SU. Die Koeffizienten für die Zertifikate Dummy Variablen bleiben in allen Modellen insignifikant. Patienten und Gesundheitssysteme können stark von einer Zentralisierung der Schlaganfallbehandlungen profitieren. Wenn in Deutschland alle Patienten aus Krankenhäusern ohne SU in Krankenhäusern mit SU behandelt würden, könnte die Schlaganfall-Sterblichkeit um 460 Fälle reduziert werden. Die Resultate zeigen auch, dass Infrastrukturspezialisierung einen deutlich stärkeren Effekt auf die Ergebnisqualität hat als Zertifizierungen.

# Stroke Units, Certification, and Stroke Outcomes in German Hospitals: A Fixed Effects Model for Stroke Mortality from 2006 – 2014

**Running head:** Stroke Units, Certification, and Stroke Outcomes in German Hospitals

**Key Words**: stroke, stroke mortality, stroke unit, hospital specialization, certificate, certification, outcomes research

**Manuscript word count**: 4,509

**Table and figure count**: 4 tables and 2 figures

Christoph Pross MSc*; Elke Berger MPH*; Dr. Alexander Geissler*; Dr. Martin Siegel[#] °;

Prof. Dr. med. Reinhard Busse MPH FFPH*[+]°

* Department of Health Care Management; Berlin University of Technology

[#] Department of Public and Health Economics, Berlin University of Technology

[+] European Observatory on Health Systems and Policies

° Berlin Centre of Health Economics Research

**Corresponding author:**

Christoph Pross
Berlin University of Technology
Dept. of Health Care Management
Administrative office H80
Str. des 17. Juni 135
10623 Berlin, Germany
Phone: +4917680804596
Email: christoph.pross@campus.tu-berlin.de

**Statement on funding:** Lead author Christoph Pross is supported by a general PhD scholarship from the Konrad-Adenauer-Foundation. Financial support for Martin Siegel from the German Ministry of Education and Research (BMBF), funding number 01EH1604A, is gratefully acknowledged.

**Statement on conflicts of interest:** No conflicts of interest.

**ABSTRACT**

Treatment of stroke patients in stroke units (SU) has increased and studies have shown improved outcomes for SUs. However, a large share of patients is still treated in hospitals without SU. Inter- and intra-country stroke outcome variation remains substantial. The effects of service line SU and total hospital quality (THQ) certification on outcomes remain unclear. We employ 2006–2014 hospital panel data, which includes structural data and 30-day standardized mortality for 1,100–1,300 German hospitals annually. We estimate hospital- and time-fixed effects regressions. The main independent variables are dummy variables for SU infrastructure and SU and THQ certification. The number of hospitals with SU has increased considerably in our observation period. Descriptive analysis illustrates better stroke outcomes for non-certified and certified SUs and hospitals with THQ certification. In a fixed effects model, having a SU has a significant quality-enhancing effect, lowering stroke mortality by 5.6%. SU service line and THQ certification remain insignificant across models. Patients and health systems may benefit substantially from SU treatment expansion. Installing a specialized SU appears more important than getting the SU certified or a THQ certification. Hospitals should prioritize investment in SU infrastructure and health systems should centralize stroke care in SUs.

## 1. INTRODUCTION

Stroke is the second leading cause of death worldwide (GBD 2015 Mortality and Causes of Death Collaborators, 2016). Recent data shows an incidence of about 16 million first-ever strokes annually, resulting in 5.7 million deaths, substantial long term disabilities and high long term care costs (Strong et al., 2007). Stroke incidence, deaths and associated medical and economic costs are substantial in the US, European countries and developing countries, making stroke a truly global disease burden (GBD 2015 Mortality and Causes of Death Collaborators, 2016; Mozaffarian et al., 2015; Nichols et al., 2012).

Latest OECD Health Care Quality Indicator data suggest remarkable cross-country stroke outcome differences (e.g. mortality rates of 3.2 – 10.2 per 100 patients for ischemic 30-day stroke) (OECD, 2015). Next to cross-country differences, substantial outcome variation between hospitals within countries has been uncovered. US hospitals show wide outcome variation for risk-adjusted stroke mortality and readmissions, with 30-day mortality at 9.8% for the 10$^{th}$ percentile of hospitals and 17.8% for 90$^{th}$ percentile (Fonarow et al., 2011). Likewise, European hospitals show considerable differences for raw and risk-adjusted stroke survival (Häkkinen et al., 2014). In Germany, stroke outcome data from the AOK sickness fund indicates substantial hospital variation, with the risk-adjusted 30-day stroke mortality (observed/expected deaths) at 0.7 for the 2$^{nd}$ quintile and 2.1 for the 5$^{th}$ quintile (Blinded for peer review, 2017). Moreover, and despite various quality improvement initiatives, the level of stroke outcome variation has only been reduced slightly between 2006 and 2014.

To improve stroke care and reduce inter-hospital outcome variation, health policy makers, hospital stakeholders, and clinicians have developed and implemented several stroke care improvement initiatives, which are all associated with substantial financial and organizational costs. Most notably, stroke units (SUs) and their certification as well as total hospital quality (THQ) certification programs are expected to improve hospital processes and medical outcomes.

A SU provides specialized acute and rehabilitation care for stroke patients, with co-located and dedicated interdisciplinary teams of neurologists, internists, neuro- and vascular surgeons and radiologists. Round-the-clock access to radiology (e.g. CT scanners) and thrombectomy equipment as well as rehabilitation care are also generally included (Moon et al., 2003). SU care has been demonstrated to improve both short and long term stroke outcomes (Chan et al., 2013; Langhorne et al., 2010; Nimptsch and Mansky, 2014; Reistetter et al., 2014; Seenan et al., 2007; Stroke Unit Trialists' Collaboration, 2013a; Walter et al., 2009; Weimar et al., 2007) and reduce overall stroke treatment cost (Grieve et al., 2000; Saka et al., 2009). However, distributions of SUs and practices therein vary considerably in different countries, with many hospitals still providing stroke care in a conventional, non-SU care model (Leys et al., 2007; Moon et al., 2003). Based on a selection of 886 hospitals in 25 European countries in 2005, 5% of hospitals had a comprehensive SU including rehabilitation care, 4% a primary SU and 40% a minimum level of care required for a hospital admitting stroke patients (Leys et al., 2007). However, 51% of hospitals provided stroke care without even the minimum stroke care model set-up. Despite its apparent value, many hospitals and health systems still hesitate to invest in SUs and their large-scale adaptation.

The evidence for a positive relationship between service line or THQ certification and outcomes is mixed and incomplete. For stroke and acute myocardial infarction (AMI) as well as deliveries and hip fractures, a 2014 study has found a positive association between certified THQ management systems and clinical leadership, systems for patient safety, and clinical review, but not for clinical practice (Shaw et al., 2014). Similarly, a study on the Joint Commission on Accreditation of Healthcare Organization (JCAHO) certification found a positive effect on risk-adjusted mortality rates in a cross-section analysis of 965 hospitals in 1996 and 1997 (Joshi, 2003). Most studies find a weaker or non-existent effect between THQ and hospital outcomes (Hinchcliff et al., 2012; Shaw et al., 2014; Sunol et al., 2015), and more of a significant effect between service line quality systems and quality indicators (e.g. for stroke and acute myocardial infarction) (Johnson et al., 2014; Lichtman et al., 2009; Stroke Unit Trialists' Collaboration, 2013a; Sunol et al., 2015).

While the evidence for a positive relationship between SU treatment and stroke outcomes is more substantiated, studies with a robust fixed effect framework, large hospital dataset across multiple years and outcome quality data are rare. Even more, while certification schemes experience continued growth, the relationship between certification and outcomes of hospital care remains inconclusive (Brubakk et al., 2015; Makai et al., 2009), with the literature limited both in terms of evidence and quality of studies (Hinchcliff et al., 2012). Studies have often examined the link between certification and process measures of care and have not examined the association (or found only a weak association) between certification and outcome measures of care. Analyses are mostly based on cross-section data of a limited number of hospitals. To the best of our knowledge, no study exists that differentiates stroke outcomes for care in (i) a conventional, non-specialized model, (ii) a specialized, non-certified SU model, (iii) a certified SU care model and (iv) hospitals with certified SU and/or additional THQ certification, based on a large and multi-year panel dataset.

To examine the influence of SU infrastructure and process specialization and certification on quality of stroke care, we rely on Donabedian's structure, process, and outcome framework, where outcomes are influenced by both hospital structures and processes (Donabedian, 1966). Stroke care is a particularly good example to test this relationship, as SU set-up and certification place substantial structural and process requirements with the goal of achieving improved treatment outcomes. Therefore, we ask whether treatment of stroke patients in specialized facilities (i.e. SUs) improves quality and thus warrants a substantial investment at hospital and health system level. Likewise, we ask whether an additional SU certification further improve stroke outcomes, which could justify the cost of certification.

Moreover, we examine whether THQ certification and stroke case volumes influence the relationship between SU specialization, certification, and stroke outcomes. We rely on hospital-level panel data with hospital structural information and 30-day risk-adjusted stroke mortality for German acute hospitals from 2006 to 2014. We employ a fixed-effect regression model with 30-day stroke standardized mortality rate (SMR) as the dependent variable, and SU infrastructure, SU certification, and THQ certification as the main explanatory variables.

## 2. METHODS

### 2.1. Data used

We matched hospital data from different sources based on standardized institutional codes. First, we obtained structural hospital data for the years 2006, 2008, 2010, 2012, 2013 and 2014 from the German mandatory quality monitoring system, which is operated by the executive authority of the German health care system, the Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA) (G-BA, 2014). Second, we integrate stroke outcome data from the hospital report card initiative Quality Assurance with Routine Data (Qualitätssicherung mit Routinedaten, QSR) for the same data years. The QSR scheme is operated by the largest German sickness fund, the AOK, and employs routine in-hospital and outpatient data of AOK insured patients. The QSR initiative provides a risk-adjusted 30-day SMR, comparing observed vs. expected events, and stroke case volumes for AOK patients. For risk-adjustment purposes, the QSR initiative calculates 30-day expected mortality by means of logit regressions that include patient-specific risk-factors such as age, gender, and a set of co-morbidities (WIdO, 2014, 2015). To ensure comparability across years, the 2014 risk-adjustment model was applied to the AOK patient data for all data years. That ensures fully comparable and risk-adjusted time trend data. The QSR data covers three different stroke types, I. intracerebral hemorrhage (ICD Code I61), II. ischemic stroke (I63), and III. stroke not specified as hemorrhage or ischemic (I64).

Third, we integrate SU certification information from the German Stroke Society (Deutsche Schlaganfall Gesellschaft, DSG). The data includes information on which hospitals have had DSG-certified SUs and the period of certification. The DSG certificate has become the premier certification scheme for SUs in Germany (Nabavi and Ringelstein, 2015). The certificate stipulates e.g. high minimum case volume, staff level resources, training requirements as well as obligatory participation in the national stroke registry (DSG, 2015). Certification audits are conducted by neurologists and granted for a 3-year period. Beside the DSG certification, hospitals with non-certified SUs are identifiable by DRG-based administrative data and two specific procedure codes (OPS 8-891and 8-89b). Both capture complex stroke care in a SU setting, but the 8-89b procedure can also be coded by internal medicine departments without constant neurologist presence (Schilling et al., 2008). As in Nimptsch and Mansky (2014), we assume the existence of a SU when a hospital documents at least ten of these procedures per year.

Fourth, we integrate data from the THQ management certificate Cooperation for Transparency and Quality in Health Care (Kooperation für Transparenz und Qualität im Gesundheitswesen, KTQ), comparable to JCAHO accreditation. Central components of the KTQ certificate are continuous quality improvement in the dimensions patient orientation, employee orientation, patient safety, quality management, communication and transparency as well as leadership (KTQ-GmbH, 2012; Lindlbauer et al., 2016). Key components of the certification process are a survey-based self-assessment, an external audit, and an annual quality report summarizing the hospitals quality improvement initiatives. Like the DSG SU certificate, the certificate is granted for 3 years.

### 2.2. Empirical strategy

Based on Donabedian's structure, process, and outcome framework, we hypothesize better outcome quality for hospitals that organize stroke care through (ii) SU infrastructure and process specialization, (iii) SU certification, and (iv) total hospital quality (THQ) certification, relative to the (i) conventional, non-SU care model We model a fixed effects regression at the hospital level. To quantify the influence of (certified) SU care on stroke outcomes, we regress the log of stroke 30-day SMR ($SMR_{it}$) on separate dummy variables specifying the existence of a SU ($SU_{it}$), a DSG-certified SU ($acc\_SU_{it}$) and a KTQ-THQ certification ($acc\_THQ_{it}$). In contrast to several dichotomized approaches in previous studies (Joshi, 2003; Matsui et al., 2015; Wang et al., 2016), we include the numerical SMR as our dependent variable to exploit the full information content of the quality variable.

We use a log specification to reduce the impact of outliers and account for non-linear effects. We add the log of stroke case volume ($stroke\_CV_{it}$) to model stroke treatment experience and non-linear experience effects. We also add the share of stroke patients relative to all patients treated to account for relative importance of and organizational focus in stroke care. We also add hospital beds ($beds_{it}$), dummy variables for hospital teaching status and ownership type, and a category medical specialization (CMS) (Lindlbauer and Schreyögg, 2014) index to account for important time-variant characteristics.

To account for time-variant trends that affect each hospital equally such as technological advances, regulatory changes, and judicial decisions we specify time effects ($\tau_t$), excluding 2006 as the reference year. To adjust for the optimal level of stroke quality of care with a 0 SMR value (0 observed mortality), we adapt Battese's (1997) approach and include an additional dummy explanatory variable ($D_{it}^{SMR}$), which takes on the value of 1 when the SMR is 0, and add $D_{it}^{SMR}$ to $SMR_{it}$ before taking the log. We further adjust for the fact that hospitals treat different amounts of stroke patients by including the AOK patient stroke case volume as analytical weights. The main model is specified in equation 1:

$$\log(SMR_{it}) = \beta_0 + \beta_1 D_{it}^{SMR} + \beta_2 SU_{it} + \beta_3 certSU_{it} + \beta_4 certTHQ_{it} + \beta_5 \log\left(stroke_{CV_{it}}\right) + \beta_6 \frac{stroke\ cases}{all\ cases}_{it} + \beta_7 beds_{it} + \beta_8 CMS_{it} + \beta_9 teach_{it} + \beta_{10} private_{it} + \beta_{11} public_{it} + \alpha_i + \tau_t + \varepsilon_{it} \quad (1)$$

where, in addition to the variables specified above, $\beta_0$ is the intercept, $\alpha_i$ are individual time invariant hospital-fixed effects and $\varepsilon_{it}$ is the error term. To ensure robustness of our results and use the full information content of the SU dummy underlying stroke complex procedures, we also run a model where we replace the SU dummy $SU_{it}$ with the log of the SU complex procedures.

The data comprise repeated measurements at the hospital level which may involve autocorrelation in the error term $\varepsilon_{it}$. The Hausman test performed on the data indicates that a random effects specification would likely yield inconsistent estimates. We therefore use hospital fixed effects $\alpha_i$ to control for unobserved hospital characteristics and avoid inconsistencies. Testing the time-fixed effects $\tau_t$ for joint significance further indicates systematic differences in mortality between the years. All statistical inferences are based on heteroscedasticity-and-autocorrelation consistent estimates for the standard errors.

## 3. RESULTS

### 3.1. Descriptive results

For 2014, our hospital sample includes 1,162 hospital observations (Table I), which is a 13% decrease from 2006 and in line with overall closures and mergers of acute care hospitals in Germany. 726 stroke-treating hospitals did not have a SU, 436 hospitals had a SU, of which 222 SUs were DSG-certified, and 280 hospitals had a KTQ THQ certification. On average, hospitals treat 227 stroke patients per annum and have a 30-day stroke SMR of 0.99 (observed/expected mortality). 55% of hospitals have a teaching affiliation and ownership is distributed 22% private, 41% non-profit and 37% public. For 2014, our hospital sample includes 86% of all hospitals, which have recorded at least 2 stroke diagnoses. The discrepancy (Table 1) results from QSR data availability and the G-BA´s 2010 shift to reporting at site level, which increases the number of hospitals and sites in the overall, non-QSR sample.

*Insert Table I about here*

Figure 1 presents the weighted SMR averages for hospital sub-groups. From 2006 to 2010, hospitals with a DSG-certified SU and hospitals with both a DSG and KTQ THQ certification had the lowest 30-day SMR. However, from 2010 to 2012, the mean SMR for hospitals with a non-certified SU decreased from 1.06 to 1.00 while the SMR for hospitals with a certified SU and both SU and THQ certifications increased to 1.05. More than 30 larger hospitals with a relative high 30-day SMR receive a SU certification between 2010 and 2012. The 30-day SMR of these hospitals improves substantially in the following years, which causes the average to come back down. Hospitals treating stroke patients without a SU have the highest SMR for all years. Importantly, for all subgroups the SMR has decreased (positive quality trend) from 2006 to 2014.

Figure 2 presents box plots and average stroke case volumes for the different hospital subgroups for the year 2014. Hospitals without a SU treat on average only 70 stroke cases, with by far the largest outcome variation. Hospitals with SU treat on average 345 stroke patients, with substantially lower outcome variation. Hospitals with a certified SU and those with additional THQ certification treat twice as many stroke patients (636 and 675, respectively) and outcome variation is further reduced. In addition, Table II presents descriptive statistics for the relevant model variables summing across all years.

*Insert Figure 1 about here*

*Insert Figure 2 about here*

*Insert Table II about here*

### 3.2. Main regression model

Table III presents the regression results. In our main model M1, SU existence is associated with a 5.6% lower 30-day SMR. SU or THQ certification, however, show no significant additional effect on stroke outcomes. Stroke volume as well as the share of stroke cases relative to all inpatient cases do not have a significant effect on the SMR. The time fixed effects for years 2013 and 2014 have negative and significant coefficients (-0.05***, -0.08***). We consider M1 our main model as it implements our empirical strategy (equation 1) and has the lowest Bayesian Information Criterion (BIC) (Schwarz, 1978).

In M2, we replace the SU dummy with the log of complex stroke procedures. The effect of the number of complex SU procedures is significant at the 10% level and has a quality-enhancing negative effect on standardized 30-day stroke mortality (-0.002**). The BIC increases slightly for model M2. In M3, we add the interaction effect for service line and THQ certification, which is also insignificant and slightly increases the effect size of the SU certification dummy variable.

The effects are generally consistent across model specifications, both in size and significance. The models display low R-squared measures, however, since fixed effects regressions washout the explanatory effects of the intercepts, they generally show a much lower R-squared compared to ordinary least squares models.

*Insert Table III about here*

### 3.3. Robustness checks

To ensure robustness, we test the consistency of our results when using alternative variable, sample, and model specifications. Results of the robustness tests are displayed in Table IV. When limiting the sample to observations with more than 10 or 20 QSR stroke cases per year (M4 and M5), the effect magnitude is reduced. Increasing the number of QSR cases required for model inclusion eliminates many hospitals with very small case volumes, which can reduce the impact of statistical chance on the SMR (WIdO, 2014). But removing 1,800 and 2,882 of hospital observations from the sample also reduces statistical power and excludes many of the hospitals with little experience in treating stroke patients. We further alter the specification of our main SU dummy variable to ensure our results hold with different SU thresholds. Halving and doubling the complex stroke procedure threshold (M6 and M7) for the SU dummy specification does not change the results noteworthy.

We also test the effect of the two certification schemes without the SU infrastructure measure (M8). Even without the SU variable, both the SU service line certification and the THQ certification remain insignificant. Furthermore, to independently check the effect of THQ certification on AMI as another emergency condition, we estimate equation 1 without the SU infrastructure and certification dummy variables and replace the relevant stroke variables with the correspond AMI variables (M9). As before, the coefficient for the THQ certification dummy remains insignificant. Furthermore, when the 30-day SMR is substituted by the simple 30-day mortality rate as the dependent variable, the results remain comparable. Likewise, when the main model (M1) is run without AOK patient case volume as analytical weights, then the effect size doubles to -10.8% (-0.108***). When replacing the AOK stroke patient number as analytical weight with the overall stroke ICD patient number, the results also remain comparable with M1.

*Insert Table IV about here*

## 4. DISCUSSION

The stroke SMR trends for the different hospital sub-groups suggest successively better stroke outcomes in hospitals with SU infrastructure, a SU that is also DSG-certified, and a certified SU within a THQ certified hospital. However, results of the fixed effects regression models show that only having a SU significantly enhances outcome quality of care while service line and THQ certification have no significant separate effects. On average and ceteris paribus, hospitals with SU have a 5.6% lower standardized 30-day mortality compared to hospitals without a SU infrastructure, even after adjusting for stroke treatment experience (case volume) and share of stroke cases. The results are in line with previous research and confirm the benefits of treating patients in a specialized SU infrastructure within a multi-year, large sample fixed effects regression framework (Hinchcliff et al., 2012; Lichtman et al., 2009; Stroke Unit Trialists' Collaboration, 2013a; Sunol et al., 2015).

In contrast, both the service line SU certification as well as the THQ certification do not show significant results. The structural and process difference between non-certified and certified SUs might be too small to cause a significant effect since most hospitals first build up the SU infrastructure and initiate the certification afterwards for signaling purposes. The overall hospital quality management improvements associated with the THQ certification might not be specific enough to impact outcomes in emergency medical condition such as stroke or AMI. Similarly, signaling might be the primary reason for certification initiation and the process improvement delta relative to non-THQ hospitals might be too small to be systematically represented in the analyzed data.

At the hospital level, investment in SUs is associated with significantly better outcomes, which may in turn result in lower cost due to readmissions and potentially higher reimbursements in a future quality-based payment environment. Thus, next to lives saved and disabilities prevented, hospitals can benefit financially from investment in SUs. At a health system level, our results raise concerns why a large share of German stroke patients are still treated in non-specialized facilities and the change towards a centralized stroke treatment model is rather slow (Nimptsch and Mansky, 2014). Centralizing care at SU hospitals can substantially improve stroke outcomes in Germany. Treating all stroke patients at hospitals with a SU may result in a ceteris paribus decrease of the absolute 30-day stroke mortality by 5.6% from 16.2% to 15.3%. For those roughly 50,300 stroke patients currently treated at hospitals without SUs, this would correspond to 460 fewer annual stroke-related death. Considerable reductions in stroke-related disabilities as well as in medical and economic costs could be expected as an additional benefit (Stroke Unit Trialists' Collaboration, 2013b).

Experiences in other European countries such as England, Denmark or the Netherlands have demonstrated the positive outcome impact of stroke care centralization in SUs (Douw et al., 2015; Fulop et al., 2013; Lahr et al., 2012; Morris et al., 2014; Ramsay et al., 2015). Underlying the centralization argument is the positive volume-outcome relationship, which has also been shown to hold for stroke in the literature (Hannan et al., 1998; Saposnik et al., 2007; Tsugawa et al., 2013). Our results also show that even when controlling for volume (i.e. experience), specialized infrastructure and processes still improve outcomes. In the mid-term, national and regional policy makers should ensure that all stroke patients are treated in SUs, by requiring a SU infrastructure for stroke care and by centralizing stroke care with hospitals that already operate a well performing SU. Currently in Germany, as in other countries, centralization of stroke care in SUs is not implemented at sufficient scale and speed, even though hospital and emergency planning falls under responsibility of state governments. Accordingly, German health policy makers, as well as policy makers in other countries, should centralize care in existing SUs and expanded stroke unit infrastructure where the number of stroke units is still insufficient.

Although the German certification of SUs sets high procedural, personnel and infrastructural standards, in contrast to expectations, the SU service line certification shows no additional significant association with 30-day stroke SMR, when we control for SU existence. Several explanations are possible here. The DSG certification confirms the SU set-up externally, with some additional resource and process requirements. Key SU characteristics such as co-location of a dedicated and interdisciplinary team of experts as well as access to imagining equipment are already implemented in a SU and the additional infrastructure and process enhancements towards a certified SU might not have a large enough additional effect on the 30-day mortality to be detected in the data.

Although mortality is a valid and well accepted outcome parameter (Walsh et al., 2002), it is only one of the clinical outcomes that matters in stroke care (AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH, 2015). Others, such as readmissions, degree of disabilities, and quality of life are important as well (Haacke et al., 2006; Schubert et al., 2014; Stroke Unit Trialists' Collaboration, 2013a), but currently standardized and risk-adjusted data for most of these outcome parameters is not available in Germany. Certified SUs, however, might have better outcomes for these indicators because the DSG certification has a holistic approach with a focus on reducing disabilities after stroke (Nabavi and Ringelstein, 2015). Furthermore, certified SU might have better outcomes over a longer timeframe than the here examined 30 days after hospital admission.

Likewise, certified SU might provide care for more severe patients, as they have on average substantially higher case volumes (Figure 2). The standardized 30-day stroke mortality is adjusted for co-morbidities, however, stroke severity may not fully be controlled for (Wiedmann et al., 2014). Yet, the impact of severity adjustment on risk-adjusted indicators that already are adjusted for co-morbidities, age and other patient characteristics has been shown to be limited (Keyhani et al., 2012). Lastly, the suspension of the DSG SU certification process in 2008 and first months of 2009, which resulted in delays for about 100 re- or new stroke unit certifications (Busse, 2009), might have also reduced the effectiveness of the DSG certification for the time span 2008-2012 and the amount of 30-day stroke SMR improvement attributable to the DSG certification.

Like service line certification, THQ certification, captured through the KTQ certificate, showed no additional significant effect on 30-day stroke mortality. This is in line with previous studies in other countries, such as Makai et al. (2009) and Sunol et al.(2015).

The major target of the KTQ-THQ certification is the general improvement of hospital quality management. The achievement of this objective might not appropriately be reflected by 30-day mortality in one specific emergency condition. Other measures such as patient safety, patient and employee responsiveness and satisfaction, and operational efficiency at the overall hospital level might be more affected by THQ certification. For example, (Lindlbauer and Schreyögg, 2014) show improved technical efficiency for KTQ-certified hospitals. A downward bias of the THQ effect might also be caused by the fact that no consolidated and standardized data on ISO 9001 certification, which is a universal quality certificate also applied in hospitals, is available. Hospitals without a KTQ certification might alternatively have an ISO 9001 THQ certification even though they appear without THQ certification in our dataset. However, the number of ISO 9001 certifications is likely substantially smaller compared to the KTQ-certified hospitals (Lindlbauer et al., 2016).

Lastly, there are benefits from certification schemes which are not captured by outcome quality. In general, consumers of credence and experience goods often rely on third party evaluators to ascertain quality (Pauly et al., 2012). Both the SU certification and the THQ certification provide quality signals for patients, emergency response teams, and admitting physicians, which can facilitate hospital choice decisions and enhance welfare even though certification schemes might not independently reduce actual hospital mortality.

## 5. Conclusions

Our results substantiate the positive effect for SU treatment on stroke outcomes, based on a fixed effects model and large multi-year hospital sample. Hospital and health system investment in SUs may substantially improve stroke outcomes. SUs may help to save numerous life years, to reduce stroke associated disabilities, and to lower long-term stroke treatment cost considerably. Germany can learn from other country examples regarding centralization and (mandatory) emergency protocols for stroke treatment. As the first study to distinguish the potential effects of SU existence, SU certification and THQ certification, we do not find a significant effect for SU certification or THQ certification on top of the large and significant effect for SU specialization.

Our research contributes to the literatures on outcomes and operational research and how hospital quality of care can be improved through structural and process enhancements. The results have implications for the organization of stroke care in other countries as well as the academic and professional debate around the benefits of infrastructure specialization and certification in health care. Additional research can examine the effect of specialization and service line certification on other stroke outcome measures (e.g. disabilities) and outcomes in other treatment areas, such as cardiology or oncology specialized treatment units. Likewise, the effect of THQ can also be examined with other outcome indicators, with additional information on other THQ certifications and for other more elective treatment areas, where a THQ certification might possibly show a higher impact.

**Acknowledgements:** We thank the Research Institute of the AOK SHI fund (WIdO), and specifically in person Christian Günster, for granting us access to the QSR outcome data. We also thank Prof. Dr. med. Otto Busse from the German Stroke Society and Mrs. Gesine Dannenmaier and her team at the Kooperation für Qualität und Transparenz im Gesundheitswesen for their willingness to share their data and expertise with our research team.

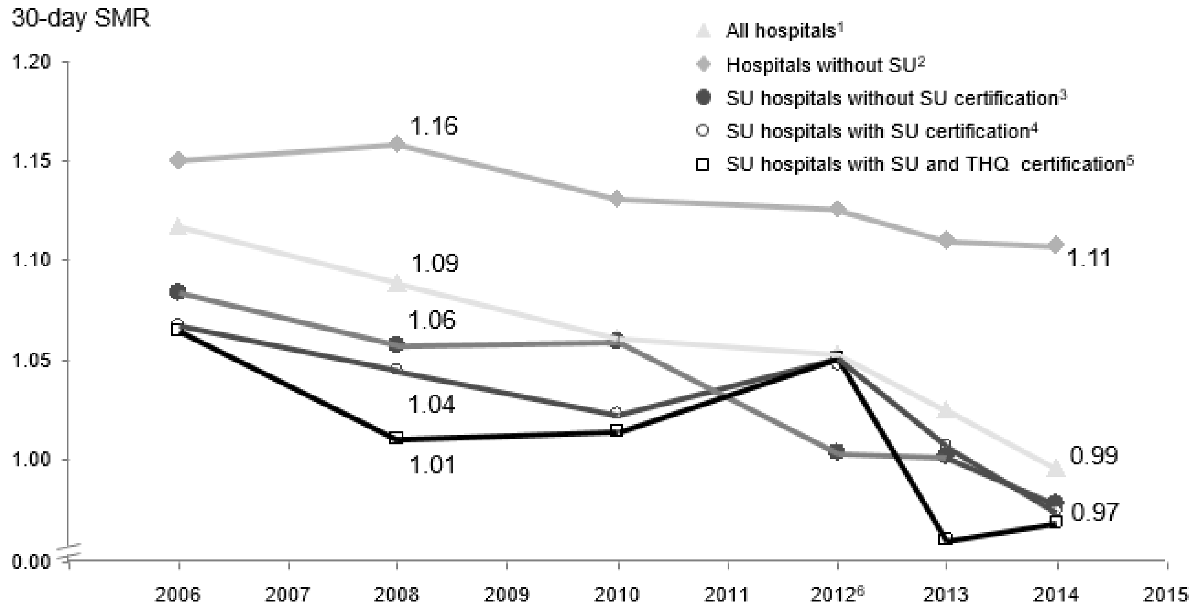
**Table I:** Overview main variable time trends

| Variables | 2006 | 2008 | 2010 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|
| Number of hospital observations[1] | 1,331 | 1,292 | 1,244 | 1,228 | 1,203 | 1,162 |
| Average 30-day stroke SMR[2] | 1.12 | 1.09 | 1.06 | 1.05 | 1.02 | 0.99 |
| Hospitals with SU[3, 4] (%) | 276 | 337 | 439 | 414 | 423 | 436 |
| Hospitals with certified SU[4, 5] | 162 | 189 | 177 | 211 | 220 | 222 |
| Hospitals with THQ certification[4] | 383 | 434 | 398 | 360 | 334 | 280 |
| Average stroke case volume[6] | 167 | 188 | 202 | 215 | 219 | 227 |
| Share stroke cases/inpatient cases, % | 3.2 | 3.3 | 3.4 | 3.5 | 3.4 | 3.6 |
| Hospitals with teaching status (%) | 511 (38) | 557 (43) | 581 (47) | 619 (50) | 625 (52) | 642 (55) |
| Average number of hospital beds | 331 | 334 | 341 | 347 | 348 | 355 |
| CMS[5] specialization index | 1.35 | 1.34 | 1.46 | 1.38 | 1.37 | 1.40 |
| Private, for-profit hospitals (%) | 225 (17) | 237 (18) | 245 (20) | 257 (21) | 255 (21) | 255 (22) |
| Private, non-for-profit hospitals (%) | 592 (44) | 574 (44) | 551 (44) | 527 (43) | 514 (43) | 477 (41) |
| Public hospitals (%) | 514 (39) | 481 (37) | 448 (36) | 444 (36) | 434 (36) | 430 (37) |
| Number of hosp. with stroke diagnoses[7] | 1,332 | 1,362 | 1,329 | 1,370 | 1,341 | 1,344 |

**Legend:** Table I illustrates the time trend for key variables in the data set

**Note**: 1. all observations that have QSR SMR stroke outcome data; 2. weighted by the AOK stroke patient volume for each hospital; 3. based on more than 10 documented complex stroke procedures (OPS codes 8_891 und 8_89b), for 2014 461 SU exist in full sample independent on whether QSR data is available; 4. and have QSR SMR stroke outcome data (especially for THQ overall more certified hospitals in Germany); 5. DSG SU certification suspended in 2008 and part of 2009, which led to a backlog of (re-) certification applications and a reduction in DSG certified hospitals in data year 2010 6. based on ICD stroke diagnoses I61 (hemorrhage), I63 (ischemic) and I64 (not further specified); 7. All hospitals which have coded 2 or more stroke ICD cases. Discrepancy in number of observations due to QSR data availability and G-BA reporting for multiple sites starting in 2010 and becoming mandatory in 2012 and 2013

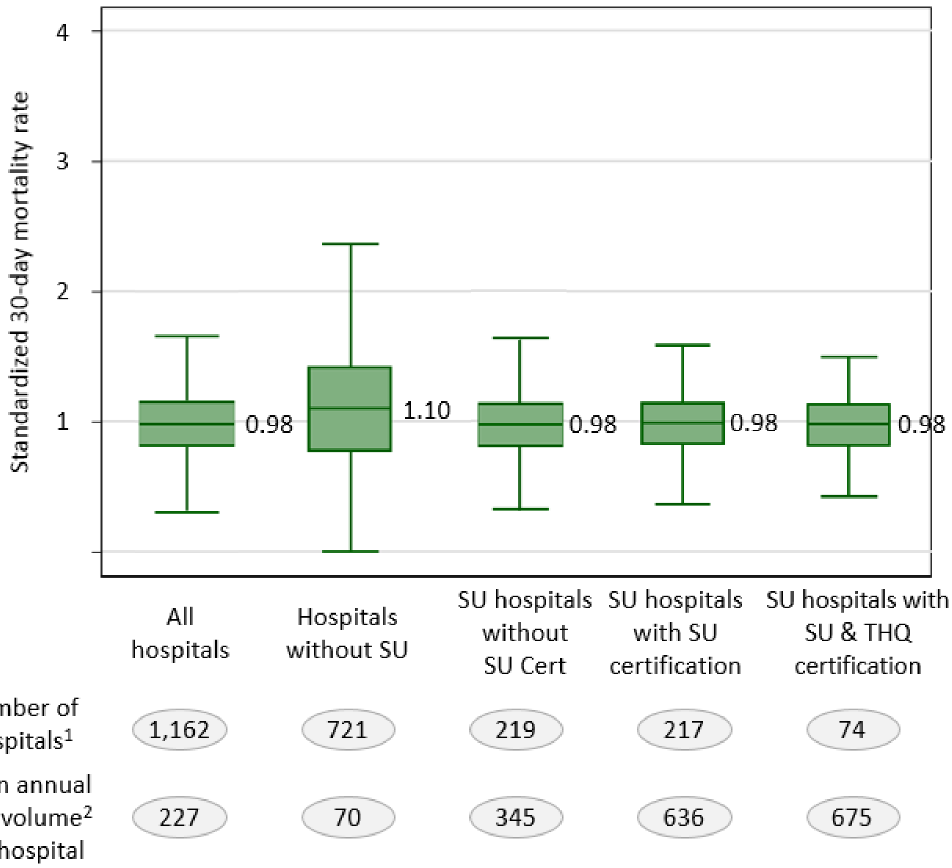
**Figure 1:** Mean 30-day SMR rates for different hospital sub-groups, 2006 – 2014




**Legend:** Figure 1 illustrates the 30-day stroke SMR time trend for the different hospital sub-groups of interest

**Note**: QSR stroke patient volume applied as analytical weights; 30-day stroke standardized mortality rate shown for different hospital subgroups; 1 = all hospitals that treat stroke patients (and have QSR stroke data); 2 = hospitals that treat stroke patients in a conventional set-up without SU (coded 10 or less stroke complex procedures); 3 = hospitals with a SU, which have not received a DSG SU certification; 4 = hospitals with SU and a DSG SU certification; 5 = hospitals with a SU, a DSG stroke unit certification and KTQ THQ certification; 6 = More than 30 larger hospitals with a relative high 30-day SMR receive a SU certification between 2010 and 2012 (16 of these hospitals also have KTQ THQ certification). These hospitals bring the average 30-day SMR up for the sub-groups "SU hospitals with SU certification" and "SU hospitals with SU and THQ certification", but improve substantially in the following years, which causes the average SMR for those subgroups to come back down.

**Figure 2:** Box plots for 30-day SMR for different hospital sub-groups, 2014




**Legend:** The box plot in Figure illustrates the median outcome and outcome variation differences between the different hospital sub-groups. Furthermore, number of hospitals and average stroke patients in the hospital sub-groups are displayed

**Note**: QSR stroke patient volume applied as analytical weights; number besides box plot indicated median value; outside values (values between lower and upper quartile and 1.5* interquartile range above or below lower/upper quartile) excluded; 1. Hospitals of sub-groups sum to all hospitals by summing all sub-groups, subtracting 74 hospitals with SU and THQ certification (already included in sub-group hospitals with certified SU) and adding 5 since 5 hospitals have no SU specialization (not coded sufficient stroke complex procedures), but have a DSG SU certification; 2. mean annual hospital stroke case volume for sub-group

**Table II:** Descriptive statistics, all years

| Variables | Mean | SD | Min | Max |
|---|---|---|---|---|
| Number of observations[1] (2006-2014) | 7462 | | | |
| Average stroke 30-day SMR[2] | 1.06 | 0.4 | 0.0 | 17.8 |
| *Log average stroke 30-day SMR[2]* | *0.02* | *0.3* | *-2.1* | *2.9* |
| Hospitals with a specialized SU | 0.31 | 0.5 | 0.0 | 1.0 |
| Average complex stroke procedures | 144.96 | 303.0 | 0.0 | 6443.0 |
| *Log average complex stroke procedures[2]* | -11.83 | 12.3 | -20.7 | 8.8 |
| Hospitals with DSG-certified SU | 0.16 | 0.4 | 0.0 | 1.0 |
| Hospitals with KTQ-TQH certification | 0.29 | 0.5 | 0.0 | 1.0 |
| Average case volume stroke cases | 202.17 | 279.0 | 0.0 | 5327.0 |
| *Log average case volume stroke cases[2]* | 3.84 | 4.1 | -20.7 | 8.6 |
| Share stroke cases / inpatient cases | 0.03 | 0.1 | 0.0 | 0.8 |
| Hospitals with teaching status | 0.47 | 0.5 | 0.0 | 1.0 |
| Average number of hospital beds | 342.23 | 302.0 | 0.0 | 3213.0 |
| CMS specialization index | 1.35 | 0.8 | 0.0 | 4.3 |
| Private, for-profit hospitals | 0.20 | 0.4 | 0.0 | 1.0 |
| Private, non-for-profit hospitals | 0.43 | 0.5 | 0.0 | 1.0 |
| Public hospitals | 0.37 | 0.5 | 0.0 | 1.0 |

**Legend:** Table II displays the descriptive statistics mean, standard deviation, minimum and maximum for the key variables used in the main regression for all years.

**Note**: 1. all hospitals that have QSR SMR stroke outcome data; 2. QSR stroke patient volume applied as analytical weights; 3. 0.000000001 or 1 added before taking the log to avoid losing observations with 0 values, results are similar for both

**Table III:** Main regression results M1-M3

| Variables | M1 | M2 | M3 |
|---|---|---|---|
| Log dummy | -0.059* | -0.059* | -0.059* |
| | (0.03) | (0.03) | (0.03) |
| Stroke unit (SU) | -0.056** | | -0.056** |
| | (0.02) | | (0.02) |
| SU certification | -0.005 | -0.005 | -0.020 |
| | (0.02) | (0.02) | (0.02) |
| THQ certification | 0.016 | 0.016 | -0.002 |
| | (0.02) | (0.02) | (0.02) |
| Log stroke case volume | -0.003 | -0.003 | -0.003 |
| | (0.00) | (0.00) | (0.00) |
| Share stroke ICD/all ICD | -0.054 | -0.063 | -0.070 |
| | (0.35) | (0.35) | (0.35) |
| Hospital beds | -0.000 | -0.000 | -0.000 |
| | (0.00) | (0.00) | (0.00) |
| Teaching hospital status | 0.007 | 0.008 | 0.008 |
| | (0.02) | (0.02) | (0.02) |
| CMS ICD specialization index | -0.003 | -0.002 | -0.003 |
| | (0.02) | (0.02) | (0.02) |
| Private, for-profit hospitals[1] | -0.022 | -0.021 | -0.022 |
| | (0.06) | (0.06) | (0.06) |
| Public hospitals[1] | -0.011 | -0.010 | -0.013 |
| | (0.07) | (0.06) | (0.06) |
| Log all OPS stroke procedures | | -0.002** | |
| | | (0.00) | |
| Interacted SU and THQ certification | | | 0.055 |
| | | | (0.03) |
| Constant | 0.182* | 0.137 | 0.188* |
| | (0.08) | (0.08) | (0.08) |
| Hospital and time fixed effects | yes | yes | yes |
| Number of observations | 7376 | 7376 | 7376 |
| R2-within | 0.016 | 0.015 | 0.017 |
| R2-between | 0.032 | 0.031 | 0.030 |
| R2-overalll | 0.018 | 0.016 | 0.017 |
| BIC | 1084 | 1088 | 1086 |
| Intraclass correlation | 0.512 | 0.512 | 0.513 |
| F-statistic | 3.9 | 3.8 | 4.0 |

**Legend:** Regression results for main model M1 and its alternatives M2 and M3
**Note**: QSR stroke patient volume applied as analytical weights; * p<0.05, ** p<0.01, *** p<0.001; time-fixed effects not displayed separately (in M1, $\beta_{2008}$ = -0.02, $\beta_{2010}$ = -0.03, $\beta_{2012}$ = -0.03, $\beta_{2013}$ = -0.05***, $\beta_{2014}$ = -0.08***), test for joint significance of time effects in M1 with F-statistic of 5.59; 1. leaving out private, non-profit hospitals as the reference category

**Table IV:** Result Robustness M4-M10

| variables | M4 | M5 | M6 | M7 | M8 | M9 |
|---|---|---|---|---|---|---|
| Log dummy | 0.181*** | 0.260** | -0.059* | -0.060* | -0.055* | |
| | (0.04) | (0.09) | (0.03) | (0.03) | (0.03) | |
| SU infrastructure (> 10 QSR patient cases) | -0.044* | | | | | |
| | (0.02) | | | | | |
| SU infrastructure (> 20 QSR patient cases) | | -0.042* | | | | |
| | | (0.02) | | | | |
| SU infrastructure (>5 complex stroke procedures) | | | -0.056** | | | |
| | | | (0.02) | | | |
| SU infrastructure (>20 complex stroke procedures) | | | | -0.056** | | |
| | | | | (0.02) | | |
| SU certification | -0.003 | -0.003 | -0.005 | -0.005 | -0.008 | |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | |
| THQ certification | 0.018 | 0.018 | 0.016 | 0.016 | 0.016 | -0.020 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Log (stroke case volume) | -0.001 | -0.001 | -0.003 | -0.003 | -0.004 | |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | |
| Share stroke ICD/all ICD | 0.108 | 0.357 | -0.056 | -0.047 | -0.104 | |
| | (0.38) | (0.36) | (0.35) | (0.35) | (0.36) | |
| Log (hospital beds) | -0.000 | -0.000 | -0.000 | -0.000 | -0.000* | -0.000 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Teaching hospital status | 0.008 | 0.002 | 0.007 | 0.007 | 0.003 | 0.019 |
| | (0.03) | (0.03) | (0.02) | (0.02) | (0.02) | (0.03) |
| CMS ICD | -0.004 | -0.006 | -0.003 | -0.003 | -0.004 | 0.014 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| Private hospital[1] | -0.029 | -0.049 | -0.022 | -0.022 | -0.027 | 0.092 |
| | (0.07) | (0.07) | (0.06) | (0.06) | (0.06) | (0.08) |
| Public hospital[1] | -0.019 | -0.036 | -0.011 | -0.011 | -0.10 | 0.100 |
| | (0.07) | (0.07) | (0.07) | (0.07) | (0.07) | (0.08) |
| Log dummy AMI | | | | | | 0.000 |
| | | | | | | (.) |
| log (AMI case volume) | | | | | | -0.016 |
| | | | | | | (0.03) |
| Share AMI ICD/all ICD | | | | | | -1.300 |
| | | | | | | (2.43) |
| constant | 0.160* | 0.176* | 0.182* | 0.181* | 0.163* | 0.094 |
| | (0.08) | (0.09) | (0.08) | (0.08) | (0.08) | (0.14) |
| Hospital and time fixed effects | yes | yes | yes | yes | yes | yes |
| number of observations | 5576 | 4494 | 7376 | 7376 | 7376 | 6188 |
| R2-within | 0.020 | 0.021 | 0.016 | 0.016 | 0.014 | 0.026 |
| R2-between | 0.000 | 0.005 | 0.032 | 0.032 | 0.015 | 0.030 |
| R2-overalll | 0.001 | 0.000 | 0.018 | 0.017 | 0.009 | 0.008 |
| BIC | 432 | -41 | 1084 | 1084 | 1090 | 2176 |
| Intraclass correlation | 0.552 | 0.615 | 0.512 | 0.512 | 0.516 | 0.621 |
| F-statistic | 5.0 | 4.0 | 3.9 | 3.9 | 3.4 | 7.0 |

**Legend:** Regression results for alternative regressions as part of the robustness check
**Note**: QSR stroke patient volume applied as analytical weights; * p<0.05, ** p<0.01, *** p<0.001; 1. Leaving out private, non-profit hospitals as the reference category

## References

AQUA - Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH (2015) *Versorgungsqualität bei Schlaganfall: Konzeptskizze für ein Qualitätssicherungsverfahren* [Online]. Available at https://www.g-ba.de/downloads/39-261-2283/2015-06-18/_AQUA/_Abnahme-Konzeptskizze-Schlaganfall.pdf.

Battese, G. E. (1997) 'A NOTE ON THE ESTIMATION OF COBB-DOUGLAS PRODUCTION FUNCTIONS WHEN SOME EXPLANATORY VARIABLES HAVE ZERO VALUES', *Journal of Agricultural Economics*, vol. 48, 1-3, pp. 250–252.

Blinded for peer review (2017) 'Hospital choice matters: a mixed method analysis of outcome quality performance and variation in German hospitals from 2006-2014', *Value In Health (under review)*.

Brubakk, K., Vist, G. E., Bukholm, G., Barach, P. and Tjomsland, O. (2015) 'A systematic review of hospital accreditation: the challenges of measuring complex intervention effects', *BMC health services research*, vol. 15, p. 280.

Busse, O. (2009) 'Mitteilungen der Deutschen Schlaganfallgesellschaft', *Nervenarzt*, vol. 6, p. 748.

Chan, D. K. Y., Cordato, D., O'Rourke, F., Chan, D. L., Pollack, M., Middleton, S. and Levi, C. (2013) 'Comprehensive stroke units: a review of comparative evidence and experience', *International journal of stroke : official journal of the International Stroke Society*, vol. 8, no. 4, pp. 260–264.

Donabedian, A. (1966) 'Evaluating the Quality of Medical Care', *The Milbank Memorial Fund Quarterly*, vol. 44, no. 3 [Online]. Available at http://www.jstor.org/stable/3348969.

Douw, K., Nielsen, C. P. and Pedersen, C. R. (2015) 'Centralising acute stroke care and moving care to the community in a Danish health region: Challenges in implementing a stroke care reform', *Health policy (Amsterdam, Netherlands)*, vol. 119, no. 8, pp. 1005–1010.

DSG (2015) *SU-Zertifizierungskriterien 2015* [Online], German Stroke Society. Available at http://www.dsg-info.de/stroke-units/zertifizierungsantraege--zertifizierungskriterien.html (Accessed 6 December 2016).

Fonarow, G. C., Smith, E. E., Reeves, M. J., Pan, W., Olson, D., Hernandez, A. F., Peterson, E. D. and Schwamm, L. H. (2011) 'Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke', *Stroke; a journal of cerebral circulation*, vol. 42, no. 1, pp. 159–166.

Fulop, N., Boaden, R., Hunter, R., McKevitt, C., Morris, S., Pursani, N., Ramsay, A. I., Rudd, A. G., Tyrrell, P. J. and DA Wolfe, C. (2013) 'Innovations in major system reconfiguration in England: a study of the effectiveness, acceptability and processes of implementation of two models of stroke care', *Implementation science : IS*, vol. 8, p. 5.

G-BA (2014) *Regelungen zum Qualitätsbericht der Krankenhäuser, 2013* [Online], Gemeinsame Bundesausschuß. Available at https://www.g-ba.de/downloads/62-492-906/Qb-R_2014-06-19.pdf (Accessed 9 September 2015).

GBD 2015 Mortality and Causes of Death Collaborators (2016) 'Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: A systematic analysis for the Global Burden of Disease Study 2015', *The Lancet*, vol. 388, no. 10053, pp. 1459–1544.

Grieve, R., Porsdal, V., Hutton, J. and Wolfe, C. (2000) ‘A comparison of the cost-effectiveness of stroke care provided in London and Copenhagen’, *International journal of technology assessment in health care*, vol. 16, no. 2, pp. 684–695.

Haacke, C., Althaus, A., Spottke, A., Siebert, U., Back, T. and Dodel, R. (2006) ‘Long-term outcome after stroke: evaluating health-related quality of life using utility measurements’, *Stroke*, vol. 37, no. 1, pp. 193–198.

Häkkinen, U., Rosenqvist, G., Peltola, M., Kapiainen, S., Rättö, H., Cots, F., Geissler, A., Or, Z., Serdén, L. and Sund, R. (2014) ‘Quality, cost, and their trade-off in treating AMI and stroke patients in European hospitals’, *Health policy (Amsterdam, Netherlands)*, vol. 117, no. 1, pp. 15–27.

Hannan, E. L., Popp, A. J., Tranmer, B., Fuestel, P., Waldman, J. and Shah, D. (1998) ‘Relationship between provider volume and mortality for carotid endarterectomies in New York state’, *Stroke; a journal of cerebral circulation*, vol. 29, no. 11, pp. 2292–2297.

Hinchcliff, R., Greenfield, D., Moldovan, M., Westbrook, J. I., Pawsey, M., Mumford, V. and Braithwaite, J. (2012) ‘Narrative synthesis of health service accreditation literature’, *BMJ quality & safety*, vol. 21, no. 12, pp. 979–991.

Johnson, A. M., Goldstein, L. B., Bennett, P., O’Brien, E. C. and Rosamond, W. D. (2014) ‘Compliance With Acute Stroke Care Quality Measures in Hospitals With and Without Primary Stroke Center Certification: The North Carolina Stroke Care Collaborative’, *Journal of the American Heart Association*, vol. 3, no. 2, e000423.

Joshi, M. (2003) ‘Hospital Quality of Care: The Link Between Accrediation and Mortality’, *Journal of Clinical Outcome Management*, vol. 10, no. 9.

Keyhani, S., Cheng, E., Arling, G., Li, X., Myers, L., Ofner, S., Williams, L. S., Phipps, M., Ordin, D. and Bravata, D. M. (2012) ‘Does the inclusion of stroke severity in a 30-day mortality model change standardized mortality rates at Veterans Affairs hospitals?’, *Circulation. Cardiovascular quality and outcomes*, vol. 5, no. 4, pp. 508–513.

KTQ-GmbH (2012) *KTQ-Manual/KTQ-Katalog: ab 2009/Version 2*, 2nd edn, Berlin, Fachverlag Matthias Grimm.

Lahr, M. M. H., Luijckx, G.-J., Vroomen, Patrick C A J, van der Zee, D.-J. and Buskens, E. (2012) ‘Proportion of patients treated with thrombolysis in a centralized versus a decentralized acute stroke care setting’, *Stroke*, vol. 43, no. 5, pp. 1336–1340.

Langhorne, P., Lewsey, J. D., Jhund, P. S., Gillies, M., Chalmers, J. W. T., Redpath, A., Briggs, A., Walters, M., Capewell, S., McMurray, J. J. V. and MacIntyre, K. (2010) ‘Estimating the impact of stroke unit care in a whole population: an epidemiological study using routine data’, *Journal of neurology, neurosurgery, and psychiatry*, vol. 81, no. 12, pp. 1301–1305.

Leys, D., Ringelstein, E. B., Kaste, M. and Hacke, W. (2007) ‘Facilities available in European hospitals treating stroke patients’, *Stroke; a journal of cerebral circulation*, vol. 38, no. 11, pp. 2985–2991.

Lichtman, J. H., Allen, N. B., Wang, Y., Watanabe, E., Jones, S. B. and Goldstein, L. B. (2009) ‘Stroke patient outcomes in US hospitals before the start of the Joint Commission Primary Stroke Center certification program’, *Stroke; a journal of cerebral circulation*, vol. 40, no. 11, pp. 3574–3579.

Lindlbauer, I. and Schreyögg, J. (2014) 'The relationship between hospital specialization and hospital efficiency: do different measures of specialization lead to different results?', *Health care management science*, vol. 17, no. 4, pp. 365–378.

Lindlbauer, I., Schreyögg, J. and Winter, V. (2016) 'Changes in technical efficiency after quality management certification: A DEA approach using difference-in-difference estimation with genetic matching in the hospital industry', *European Journal of Operational Research*, vol. 250, no. 3, pp. 1026–1036.

Makai, P., Klazinga, N., Wagner, C., Boncz, I. and Gulacsi, L. (2009) 'Quality management and patient safety: survey results from 102 Hungarian hospitals', *Health policy (Amsterdam, Netherlands)*, vol. 90, 2-3, pp. 175–180.

Matsui, H., Fushimi, K., Yasunaga, H. and Taniyama, Y. (2015) 'Variation in Risk-Standardized Mortality of Stroke among Hospitals in Japan', *PloS one*, vol. 10, no. 10, e0139216.

Moon, L., Moïse, P., Jacobzone, S. and (None) (2003) 'Stroke Care in OECD Countries: A Comparison of Treatment, Costs and Outcomes in 17 Countries', *OECD Health Working Papers*, vol. 5.

Morris, S., Hunter, R. M., Ramsay, A. I. G., Boaden, R., McKevitt, C., Perry, C., Pursani, N., Rudd, A. G., Schwamm, L. H., Turner, S. J., Tyrrell, P. J., Wolfe, C. D. A. and Fulop, N. J. (2014) 'Impact of centralising acute stroke services in English metropolitan areas on mortality and length of hospital stay: difference-in-differences analysis', *BMJ (Clinical research ed.)*, vol. 349, g4757.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., Ferranti, S. de, Despres, J.-P., Fullerton, H. J., Howard, V. J., Huffman, M. D., Judd, S. E., Kissela, B. M., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Liu, S., Mackey, R. H., Matchar, D. B., McGuire, D. K., Mohler, E. R. 3., Moy, C. S., Muntner, P., Mussolino, M. E., Nasir, K., Neumar, R. W., Nichol, G., Palaniappan, L., Pandey, D. K., Reeves, M. J., Rodriguez, C. J., Sorlie, P. D., Stein, J., Towfighi, A., Turan, T. N., Virani, S. S., Willey, J. Z., Woo, D., Yeh, R. W. and Turner, M. B. (2015) 'Heart disease and stroke statistics--2015 update: a report from the American Heart Association', *Circulation*, vol. 131, no. 4, e29-322.

Nabavi, D. G. and Ringelstein, E. B. (2015) 'Stroke Units', in Jungehülsing, G. J. and Endres, M. (eds) *Komplikationen und Folgeerkrankungen nach Schlaganfall,* Stuttgart and New York, Thieme, pp. 158–161.

Nichols, M., Townsend, N., Scarborough, P. and Rayner, M. (2012) *European cardiovascular disease statistics*, 2012nd edn, Brussels, Sophia Antipolis, European Heart Network; European Society of Cardiology.

Nimptsch, U. and Mansky, T. (2014) 'Stroke unit care and trends of in-hospital mortality for stroke in Germany 2005-2010', *International journal of stroke : official journal of the International Stroke Society*, vol. 9, no. 3, pp. 260–265.

OECD (2015) *Health at a Glance 2015*, OECD Publishing.

Pauly, M. V., McGuire, T. G. and Barros, P. P., eds. (2012) *Handbook of Health Economics*, Amsterdam, London, North Holland.

Ramsay, A. I. G., Morris, S., Hoffman, A., Hunter, R. M., Boaden, R., McKevitt, C., Perry, C., Pursani, N., Rudd, A. G., Turner, S. J., Tyrrell, P. J., Wolfe, C. D. A. and Fulop, N. J.

(2015) ‘Effects of Centralizing Acute Stroke Services on Stroke Care Provision in Two Large Metropolitan Areas in England’, *Stroke*, vol. 46, no. 8, pp. 2244–2251.

Reistetter, T. A., Karmarkar, A. M., Graham, J. E., Eschbach, K., Kuo, Y.-F., Granger, C. V., Freeman, J. and Ottenbacher, K. J. (2014) ‘Regional variation in stroke rehabilitation outcomes’, *Archives of physical medicine and rehabilitation*, vol. 95, no. 1, pp. 29–38.

Saka, O., Serra, V., Samyshkin, Y., McGuire, A. and Wolfe, Charles C D A (2009) ‘Cost-effectiveness of stroke unit care followed by early supported discharge’, *Stroke; a journal of cerebral circulation*, vol. 40, no. 1, pp. 24–29.

Saposnik, G., Baibergenova, A., O'Donnell, M., Hill, M., Kapral, M. and Hachinski, V. (2007) ‘Hospital volume and stroke outcome: Does it matter?’, *Neurology*, vol. 69, no. 11.

Schilling, M., Kiefer, R., Busse, O., Ferbert, A., Grond, M. and Ringelstein, E. B. (2008) *Kodierleitfaden Schlaganfall der DSG und DGN 2008*.

Schubert, I., Ihle, P., Köster, I., Küpper-Nybelen, J., Rentzsch, M., Stallmann, C., Swart, E. and Winkler, C. (2014) *Datengutachten für das Deutsche Institut für Medizinische Dokumentation und Information: Gutachten: Daten für die Versorgungsforschung. Zugang und Nutzungsmöglichkeiten*.

Schwarz, G. (1978) ‘Estimating the Dimension of a Model’, *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464.

Seenan, P., Long, M. and Langhorne, P. (2007) ‘Stroke units in their natural habitat: systematic review of observational studies’, *Stroke; a journal of cerebral circulation*, vol. 38, no. 6, pp. 1886–1892.

Shaw, C. D., Groene, O., Botje, D., Sunol, R., Kutryba, B., Klazinga, N., Bruneau, C., Hammer, A., Wang, A., Arah, O. A. and Wagner, C. (2014) ‘The effect of certification and accreditation on quality management in 4 clinical services in 73 European hospitals’, *International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua*, 26 Suppl 1, pp. 100–107.

Stroke Unit Trialists' Collaboration (2013a) ‘Organised inpatient (stroke unit) care for stroke’, *The Cochrane database of systematic reviews*, no. 9, CD000197.

Stroke Unit Trialists' Collaboration (2013b) ‘Organised inpatient (stroke unit) care for stroke’, *The Cochrane database of systematic reviews*, no. 9, CD000197.

Strong, K., Mathers, C. and Bonita, R. (2007) ‘Preventing stroke: Saving lives around the world’, *The Lancet Neurology*, vol. 6, no. 2, pp. 182–187.

Sunol, R., Wagner, C., Arah, O. A., Kristensen, S., Pfaff, H., Klazinga, N., Thompson, C. A., Wang, A., DerSarkissian, M., Bartels, P., Michel, P. and Groene, O. (2015) ‘Implementation of Departmental Quality Strategies Is Positively Associated with Clinical Practice: Results of a Multicenter Study in 73 Hospitals in 7 European Countries’, *PloS one*, vol. 10, no. 11, e0141157.

Tsugawa, Y., Kumamaru, H., Yasunaga, H., Hashimoto, H., Horiguchi, H. and Ayanian, J. Z. (2013) ‘The association of hospital volume with mortality and costs of care for stroke in Japan’, *Medical care*, vol. 51, no. 9, pp. 782–788.

Walsh, K., Gompertz, P. H. and Rudd, A. G. (2002) ‘Stroke care: how do we measure quality? Review’, *PostgradMedJ*, no. 78, pp. 322–326.

Walter, A., Seidel, G., Thie, A. and Raspe, H. (2009) 'Semi-intensive stroke unit versus conventional care in acute ischemic stroke or TIA--a prospective study in Germany', *Journal of the neurological sciences*, vol. 287, 1-2, pp. 131–137.

Wang, Y., Eldridge, N., Metersky, M. L., Sonnenfeld, N., Fine, J. M., Pandolfi, M. M., Eckenrode, S., Bakullari, A., Galusha, D. H., Jaser, L., Verzier, N. R., Nuti, S. V., Hunt, D., Normand, S. T. and Krumholz, H. M. (2016) 'Association Between Hospital Performance on Patient Safety and 30-Day Mortality and Unplanned Readmission for Medicare Fee-for-Service Patients With Acute Myocardial Infarction', *Journal of the American Heart Association*, vol. 5, no. 7, e003731.

Weimar, C., Ringelstein, E. B. and Diener, H.-C. (2007) 'Monitoring stroke units: management, outcome, efficiency', *Der Nervenarzt*, vol. 78, no. 8, pp. 957–966.

WIdO (2014) *Methods* [Online], Berlin, Research Institute of the AOK. Available at http://www.qualitaetssicherung-mit-routinedaten.de/methoden/ (Accessed 20 February 2015).

WIdO (2015) *Indicator handbook for treatment areas not covered in AOK Hospitalnavigator: QSR Verfahren* [Online], Berlin, Wissenschaftliches Institut der AOK.

Wiedmann, S., Heuschmann, P. U., Hillmann, S., Busse, O., Wiethölter, H., Walter, G. M., Seidel, G., Misselwitz, B., Janssen, A., Berger, K., Burmeister, C., Matthis, C., Kolominsky-Rabas, P. and Hermanek, Peter für die Arbeitsgemeinschaft Deutscher- Schlaganfall- Register (2014) 'Qualität der Behandlung des akuten Schlaganfalls: Auswertung evidenzbasierter Indikatoren von 260 000 Patientendaten', *Deutsches Ärzteblatt*, vol. 111, no. 45, pp. 759–765.

# **Kapitel 6**: Qualitätswettbewerb im deutschen Krankenhausmarkt

Christoph Pross

Preprint, zur Begutachtung geplant bei European Journal of Health Economics als: Does competition impact hospital quality of care? A fixed effect analysis of stroke, AMI and hip replacement outcomes from 2006 – 2014

**Hintergrund**: In einem hochregulativen und teilweise noch intransparenten Krankenhausmarkt soll der Krankenhauswettbewerb um Patienten neben dem effizienten Einsatz von Ressourcen auch zu einer Qualitätssteigerung führen. International haben die bisherigen empirischen Untersuchungen zum Effekt von Krankenhauswettbewerb auf Behandlungsqualität zu einem schwachen Konsensus für eine positive Wirkung in Einheitspreissystemen wie dem amerikanischen Medicare System oder dem britischen NHS System geführt. Doch sind die Resultate über Indikatoren, Leistungsbereiche und Modellansätze hinweg sehr unterschiedlich und wenig stabil. Auch wurden in Deutschland trotz der hohen Bedeutung dieser Forschungsfrage und dem zunehmenden regulativen und politischen Fokus auf das Thema Qualitätswettbewerb in diesem Bereich noch keine empirischen Studien durchgeführt.

**Ziele**: Für die Notfallindikationen Schlaganfall und Herzinfarkt und die elektive Prozedur Hüft-Total-Endoprothesenersatz soll im ersten Schritt die Entwicklung der Marktkonzentration über die Zeit und für nicht, mäßig und hochkonzentrierte Märkte betrachtet werden. Im zweiten Schritt wird mit Hilfe eines Fixed-Effekt-Modells der Einfluss von Krankenhauswettbewerb auf die Ergebnisqualität geschätzt. Damit soll ein Ausgangspunkt für die

versorgungswissenschaftliche Detailierung und weitere empirische Untersuchung etabliert werden.

**Methodik**: In Bezug auf den Produktmarkt unterscheiden wir zwischen den Leistungsbereichen Schlaganfall, Herzinfarkt und Hüft-Total-Endoprothese. Zur Abgrenzung des geographischen Marktes nutzen wir Adressdaten der Krankenhäuser und einen Radius von 30 KM. Wir inkludieren alle Krankenhäuser und deren leistungsbereichspezifische Fallzahlen in diesem Markt mit fixiertem Radius und berechnen für diese Märkt den Herfindahl-Hirschman-Index (HHI) auf Basis der Marktanteile.  Zur Messung der Ergebnisqualität nutzen wir risiko-adjustierte 30-Tage Sterblichkeit und Komplikationsraten als auch nicht-risiko-adjustierte 90-Tage Wiedereinlieferung und Re-operationen. Mit Ergebnisqualität als abhängige Variable und Wettbewerbsintensität als relevante erklärende Variable rechnen wir Fixed-Effects-Regressionen für die verschiedenen Leistungsbereiche und Indikatoren. Um die mögliche Endogenität zwischen Qualität und Wettbewerb zu adressieren verwenden wir auch ein Modell mit Instrumentenschätzung.

**Ergebnisse**: Während die Marktkonzentration für Schlaganfall zwischen 2006 und 2014 zugenommen hat, ist sie für Herzinfarkt relativ konstant geblieben und hat leicht abgenommen für Hüft-Total-Endoprothesenersatz. Für Schlaganfall und Hüft-Total-Endoprothesenersatz ist die Ergebnisqualität am schlechtesten für Krankenhäuser in hochkonzentrierten Märkten, für Herzinfarkt ist sie am schlechtesten für Krankenhäuser in mäßig konzentrierten Märkten. In den verschiedenen Regressionsmodellen konnte kein signifikanter Effekt von Wettbewerb auf Ergebnisqualität festgestellt werden. Dies mag auf der einen Seite an dem fehlenden geringen Bewusstsein für Qualitätsunterschiede zwischen Krankenhäusern und den dazu verfügbaren Public Reporting Daten liegen. Auf der anderen Seite können sich auch gegenläufig Effekte – Konzentration und Spezialisierung mit positiver Qualitätswirkung bei geringeren Fallzahlen und Wettbewerbsreduktion mit negativer Qualitätswirkung – gegenseitig aufheben.

# Does competition impact hospital quality of care?
# A fixed effect analysis of stroke, AMI and hip replacement outcomes from 2006 – 2014

Christoph Pross MSc*

* Berlin University of Technology, Department of Health Care Management;

**Corresponding author:**
Christoph Pross
Berlin University of Technology
Dept. of Health Care Management
Administrative office H80
Str. des 17. Juni 135
10623 Berlin, Germany
Phone: +4917680804596
Email: christoph.pross@campus.tu-berlin.de

**Abstract**

We examine the effects of hospital competition on quality of care in Germany for the emergency medical conditions stroke and acute myocardial infarction and the planned surgical procedure hip replacement. Based on panel data for 1,100 annual hospital observations from 2006 to 2014, we estimate a fixed effects model with risk-adjusted mortality and complications as the dependent variable and the competition level as the main dependent variable. We find generally no significant effect of competition on quality, both in normal fixed effects and in instrumental variable fixed effects regression. Patients are still not fully aware of quality of care differences and their ability to view quality of care results at a hospital level. This might isolate hospitals from quality competitive pressures. Furthermore, positive quality effects from care centralization and increased experiences might counteract negative effects of less competition in more concentrated markets. Quality transparency needs to be further increased and the effects from care centralization differentiated from competition dynamics.

**Key words** Hospital quality of care, hospital competition, HHI, stroke, AMI, hip replacement surgery

**JEL Classification**: I11, D24

# 1. Introduction

Healthcare is one of the largest sectors in the German economy, and hospitals provide the large majority of health services. Patients in Germany can freely choose their hospitals and hospitals often do not operate at capacity limited. Germany has a large installed base of hospitals, with many hospitals facing negative operative margins. This has led to a consolidation in the German hospital market, both at a local level with local hospital system emerging and at the national level with hospital chains expanding and increasingly operating in various local hospital markets (Hentschker et al., 2014). At the same time, quality of care concerns among policy makers and patients are rising, with recent legislative efforts to strengthen quality transparency and quality competition among hospitals and research uncovering substantial and continued outcome variation between hospitals for multiple treatment areas (Pross, Geissler, Busse, 2017a).

In other OECD countries, regulation and policy interventions in the hospital sector have increased in recent years, to not only constrain cost increase, but more often to enhance quality competition between hospitals and encourage patients to more actively choose between good and bad quality providers (Pross, Geissler, Busse, 2017b). The US Affordable Care Act has substantially expanded quality-related payment schemes, esp. with regards to Medicare, and encouraged the creation of Accountable Care organizations and bundled payments to integrate care across the inpatient, outpatient and rehabilitation setting (Blumenthal et al., 2015). The English NHS introduced a fixed price, free hospital choice system in 2005-2006, allowed private providers to treat NHS patients, and increased the transparency of hospital quality through patient-focused quality benchmarking websites (Cooper et al., 2011). In Germany, public reporting initiatives are expanding, payers are piloting quality-based selective contracting and legislation to introduce quality-based payments has been recently passed.

Despite the pro-competition political consensus, both theoretical and empirical investigations of the effect of hospital market competition on quality of care have shown ambiguous results (Dranove and Satterthwaite, 2012; Gaynor, 2006). Most empirical studies use a structure-conduct-performance (SCP) specification and hypothesize a causal link between market structure and hospital conduct with regards to quality of care provided. The results for the effect of provider competition appear to depend on the degree of price competition (Gowrisankaran and Town, 2003; Kessler and McClellan, 2000). In markets where prices are not regulated, the empirical evidence is mixed; however, in markets with fixed prices competition between providers appears to have a quality enhancing effect (Pauly et al., 2012); yet, several studies also report negative or no competition effects for different quality indicators in a fixed price environments (Maeda and Lo Sasso, 2012; Mutter et al., 2008; Palangkaraya and Yong, 2013). The effect of competition on quality might depend on how quality is measured and conflict with centralization of care in medical condition-specific centers of excellence with high treatment volumes. Studies examining the link between competition and quality are almost exclusively limited to the US and the UK health systems and focus on a few medical conditions, in particular acute myocardial infarction (AMI).

This paper expands the literature along four primary angles. First and foremost, we contribute to research on the effects of hospital competition on quality of care and provide the first study in the German market context as well as for a mix of emergency and planned hospital conditions. Second, we add to the literature on health reform, and in particular to the effects on possible downsides from consolidation and selective contracting associated with reduced competition. Third, we contribute to research on anti-trust proceedings in the hospitals sector, especially with regards to Germany. Some experts have cited the lack of country-specific empirical evidence in Germany and other countries as an important reason for the reluctance of competition regulators to use advanced econometric analysis (Schmid and Varkevisser, 2016). Fourth, we provide further empirical evidence on the hospital market structure in Germany, which is generally scarce (Hentschker et al., 2014).

The rest of the article is organized as follows. Section 2 examines the current state of the literature and relevant theoretical considerations. Section 3 introduces the dataset and empirical implementations. Section 4 presents the results, which are subsequently discussed in Section 5. We conclude in Section 5 with policy implications.

## 2. Literature review and theoretical considerations

In the past 15 years, an important discussion has emerged on the effects of competition on quality in health care provider markets. Basic economic theory suggests that competition will increase product or service quality if prices are regulated and above marginal cost. If price is fixed, firms compete for consumers on non-price dimensions such as quality. They will increase quality to gain market share as long as price is above marginal cost (Gaynor et al., 2012). This concepts has been tested empirically primarily in the US and UK health care markets. Results depend heavily on the degree of price competition between providers, the transparency of quality of care and the mobility of patients. With a slight consensus towards a positive effect of competition on quality of care in a fixed price, empirical results are mixed and vary across indicator, treatment area and study design.

### 2.1. Literature Review

In a seminal study in the US, Kessler and McClellan (2000) examine acute myocardial infarction (AMI) mortality rates for elderly Medicare beneficiaries between 1985 and 1994. They use patient location to identify the impact of variations in level of competition in patient´s local hospital markets. Post-1990, they find a clear positive effect of competition on AMI outcomes and partially explain this with increased enrollment in health maintenance organizations (HMO), which had a more positive contribution margin than Medicare patients. Differentiating between high and low severity patients in a fixed price Medicare setting, Kessler and Geppert (2005) find more competitive markets to have higher expenditures and higher quality of care for high risk AMI patients and lower expenditure and but statistically similar quality for low risk patients. Measuring competition via the Herfindahl-Hirschman-Index (HHI), Mukamel et al. (2002) find that higher market concentration is associated with lower spending for a range of high mortality conditions AMI, heart failure, pneumonia and stroke and lower spending is further associated with increased mortality. Shen (2003) finds lower market concentration in AMI care for Medicare patients to be associated with reduced mortality. Tay (2003) finds Medicare patient hospital choice to be positively responsive to quality of care in AMI. Similarly, Chou et al. (2014) examine Coronary Artery Bypass Graft outcomes for Medicare patients in hospitals in the US state of Pennsylvania and find hospitals in more competitive markets achieving better outcomes for more severe patients.

In contrast, Gowrisankaran and Town (2003) find an overall unclear effect from increased competition on health care provider quality. In particular, they find a negative effect of competition on quality for Medicare patients (despite its fixed price environment), which they attribute to often low or negative contribution margins. Likewise, Mukamel et al. (2001) find no significant association between market concentration and mortality from all causes for Medicare patients in 1990. Using a fixed effects model for Medicare hospital- and patient-level data from 2003 to 2006, Maeda and Lo Sasso (2012) find no significant effect of competition on mortality for Medicare patients with chronic heart failure.

Despite its mixed nature, recent reviews regard the empirical US evidence as a confirmation of the positive effect of provider competition on quality of care (Gaynor et al., 2012, 2013; Vogt and Town, 2006). An important limitation of most US studies is the endogeneity between competition and market structure, captured via market shares and the associated HHI, and changes in quality of care. If in fact the presence of a high quality hospital deters market entry or encourages exit, then the direction of the causal effect might be from quality to market structure and not from market structure on quality.

Several studies in the UK use changes in competition policy as a setting for a natural experiment to address the endogeneity. Propper et al. (2004) examined AMI mortality after two consecutive NHS reforms, which first promoted provider competition between 1991 and 1996 and after 1996 reduced competition between providers. While AMI mortality was reduced across board, they find a negative relationship between competition and quality of care, with mortality declining the strongest in most concentrated and least competitive markets before 1996. After 1996, mortality also declined in competitive markets. A later study by the same authors confirms the negative relationship between AMI survival and competition (Propper et al., 2008). In areas where competition is possible (due to number of hospitals) and heavily promoted by the regulator, hospitals appear to improve observable measures of quality (i.e. waiting times) at the cost of unobservable measures of quality (i.e. mortality).

In 2006, NHS policy changed back to more competition via free hospital choice, promotion of hospital quality transparency and a prospective payment system that ties hospital revenues to the patient volume treated. Gaynor et al. (2013) examine the pro-competition policy impact on quality of care in more and less concentrated markets via fixed effect, difference-in-difference framework. Comparing pre-policy year 2003 with post-policy year 2007 and in contrast with (Propper et al., 2004; Propper et al., 2008), they find a significant negative effect of competition on AMI 30-day mortality, with a 10% reduction in the HHI reducing the AMI mortality rate by 3%, and significant but smaller effect on 28-day all-cause mortality. Also examining the 2006 pro-competition shift, Cooper et al. (2011) employ a modified difference-in-difference estimator to test whether hospital quality improved more quickly in more competitive markets after the shift back to patient choice and hospital competition. For 30-day AMI mortality, they find a 0.3% reduction with one standard deviation increase in their patient flow and HHI-based competition measure.

Furthermore, Bloom et al. (2015) investigate the effect of competition on hospital management quality, and by extension clinical outcomes, in a cross-section study for the year 2006. Since hospital closure is centrally determined by NHS authorities, in part based on hospital quality of care, they use exogenous variation in hospital closures due to marginality of election outcomes and associated political power to estimate the causal impact of competition on management quality. Their results show a positive impact of competition on management quality, which is associated with lower AMI mortality.

As the most recent study from the UK context, Gutacker et al. (2016) find a substantial demand response to improvements in patient-reported outcomes for hip replacement, with a one standard deviation increase in health gains from before- to after-surgery increasing patient demand by 10%. They also find that patients respond less to cruder outcomes measures such as readmissions and mortality and conclude that hospital quality competition and its implications depend on type of quality measure employed.

As one of the few non-US/UK studies, Palangkaraya and Yong (2013) analyze around Australian 180 hospitals from 2000 and 2005 and find that hospitals in more competitive markets have lower unplanned re-admission rates, but higher mortality rates, with the former effect more significant than the later and effects differing greatly between public and private hospitals. In several studies in the Netherlands, Varkevisser and his co-authors analyze patient´s willing to choose hospitals. For the Netherlands, they find patients actively choosing hospitals for non-emergency care in neurosurgery and angioplasty depending on distance, waiting time, and quality (Varkevisser et al., 2010, 2012; Varkevisser and van der Geest, 2007). With regards to quality of care in angioplasty, a change of readmissions from 8.5% to 7.5% increases demand by 12% (Varkevisser et al., 2012). For other European countries including Germany, to the best of our knowledge, no study has examined the effect of hospital competition on quality of care.

As demonstrated above, empirical results on the impact of hospital market competition on quality of care provided are contested. In general, there is some consistency that hospital competition in a fixed price environment enhances competition. However, even the results in a fixed price environment are mixed. Most studies have examined the emergency condition heart attack, largely due to data availability and variability of outcomes. Empirical frameworks vary between countries due to different health system structures and regulation, policy changes and available data. The different empirical frameworks make comparability of results difficult.

## 2.2. Theoretical considerations for competition on quality in the German DRG context

Under prospective payment system with fixed DRG prices within a region and free patient hospital choice, hospitals compete on quality to attract patients (Pauly et al., 2012). In highly competitive markets, competition on quality is fiercer, leaving more competitive markets with on average higher quality of medical services. As (Brekke et al., 2011) identify, several assumptions underlie the positive effect of competition on quality of hospital care. These assumptions hold to different degrees for the German market context.

1. Marginal economic returns from an additional patient treated are positive, which is the case for surgical DRGs with a higher fixed cost share (e.g. lots of capex required) and lower marginal costs. DRG payments for stroke, AMI and hip replacement are generally considered to have positive returns for most hospitals. As DRGs in Germany are continuously adjusted and designed to include a small positive contribution margin, this assumption likely holds for the German market context.
2. Providers need spare capacity to be able to treat additional patients without suffering from increasing marginal costs. In Germany, hospital waiting times are minimal and hospital beds per capita are high (Geissler et al., 2010), so marginal cost concerns are generally less relevant.

3. Non-profit private and public hospitals have sufficient motivation to compete for patients. Both groups make up together 65% of German acute hospital market and many of the hospitals have negative profits, which they are trying to minimize by attracting more patients. All non-profit hospitals operate under the objective to operate at a small positive EBIT margin. More generally, their objectives is to provide the best care to as many patients as possible.
4. Demand for hospital services responds to quality and it responds relatively quickly. In Germany, public reporting for all acute hospitals has been initiated in 2008 and payers as well as admitting physicians have started to direct patients to higher quality hospitals. For some conditions, such as stroke, emergency personal might channel patients to better hospitals (e.g. those with a stroke unit) and substitute patient choice with emergency services hospital choice. However, overall awareness of public reporting and patient hospital choice based on outcome quality dimensions might still be limited (Pross, Averdunk et al., 2017).
5. Altruistic physician and hospital providers will value treatment benefits, including quality of care, for patients above the financial reward received from treating an additional patient (Ellis and McGuire, 1986). Thus, altruistic providers would increase quality without the incentive of attracting additional patients. While at an individual level, many physicians might demonstrate altruistic behavior, at a hospital system level, economic considerations generally prevail.

In general, when these assumptions are relaxed, the effect of competition on quality is reduced. Moreover, these assumptions might apply to different degrees to elective procedures and emergency medical conditions, which would result in a different competition effect on quality for elective and emergency conditions (Chandra et al., 2016; Propper et al., 2004).

In health care markets, location plays an important role for competition between providers. Health care service provision requires the consumer to travel, which makes provider competition geography based (Propper et al., 2008) and leads to substantial geographic variation in the degree of competition, in particular between urban and more rural areas. In general, the potential for hospital competition on quality declines quickly with increasing distance between hospitals (Gutacker et al., 2016). The willingness to travel for healthcare determines the size of hospital markets, with mobility differing across medical conditions, patient groups and countries. For more elective, clearly defined and "smaller" conditions such as hip replacement and cataract surgery willingness to choose and travel will be higher. For younger patients and patients with more financial resources, willingness to travel will be higher (Chandra et al., 2016; Friedrich and Beivers, 2009). In Germany, health care mobility is still more limited (as compared to the US), which can generally result in less hospital competition and more local or regional hospital markets.

Next to location and associated distance, patient choice of hospitals also considers several other non-outcome quality dimensions. Patients often choose the hospital that they have previously attended (Robertson and Burge, 2011; Schwartz et al., 2005). They also choose the hospital that their outpatient physician has recommended. Further, they often prefer large hospitals with academic affiliation (Victoor et al., 2012). Other dimensions than outcome quality can have large influence on hospital choice and might bear on the influence of competition on quality through their impact on patient hospital choice.

Patients can only exercise choice if they have an option between different hospitals offering the required medical service. Relative to other European counties, Germany still has a large number of hospitals with many offering large spectrum of medical services (Busse et al., 2016; Pross, Geissler, Busse, 2017a). The number of hospitals and their respective service offering is highly regulated and under jurisdiction of state and local governments, thus slow to respond to market changes and possibly unaffected by quality performance.

## 3. Data used and empirical implementation

We have assembled a comprehensive hospital-level panel dataset, derived from clinical, administrative and regulatory data sources on a variety of structural and outcome quality measures at the hospital level as well as demographics at the district level. To date in Germany, the dataset provides the most comprehensive data foundation for an empirical examination of the effects of hospital competition on quality of care.

### 3.1. Data used

We integrate patient-based, risk-adjusted outcome data for two emergency treatment areas stroke and AMI and admission-based, inpatient outcome data for the elective condition hip replacement from the two premier provider report card systems in Germany, the QSR outcome indicators from Germany´s largest sickness fund the AOK and the mandatory hospital quality monitoring system of the Federal Joint Committee (G-BA).

The QSR outcome indicators for the two emergency medical conditions stroke and AMI are centrally calculated by the WIdO research institute and based on administrative data of AOK insured patients. The stroke and AMI indicators are available from 2006 to 2014 and observations vary between 1,000 and 1,200 hospitals per year. 30-day risk-adjusted mortality includes events up to 30 days after hospital admittance, comparing number of observed with number of expected events. Expected events are similarly calculated via a logistic regression based on patient number and attributes such age, gender and comorbidities (QSR Verfahren, 2015).

The public report cards provide outcome data for the treatment area hip replacement as well as structural hospital information such as ownership type, number of beds and case volumes. The public report cards provide diagnosis, process and outcome indicators for 30 treatment areas, from which we select hip replacement and a subset of its outcome indicators, taking into account data restrictions both in terms of indicator comparability across years and size of hospital samples. We focus on outcome indicators as these are patient-relevant endpoints and the scientific evidence for the positive relationship between process and outcome indicators is limited (Copnell et al., 2009; Gandjour et al., 2002).

The hip replacement outcome data from the mandatory public report cards covers the inpatient hospital stay only and includes unadjusted rate of re-interventions due to complications from 2006 to 2014 and risk-adjusted rate of re-interventions, available from 2012 – 2014. Risk-adjusted indicators compare the number of observed events (e.g. mortalities) with the number of expected events, with the latter calculated through a logistic regression that include patient risk-factors such as age, comorbidities and gender and patient volume (G-BA, 2014). Risk-adjustment is centrally undertaken and ensures comparability of outcomes between different hospitals and their respective patient samples. We merge QSR indicators with the mandatory report card data based on standardized hospital IDs and address data.

### 3.2. Empirical implementation

We measure market structure and quality of care at the hospital and medical condition level. We use hospitals as the unit of analysis, a geographically fixed radius and medical condition specific competition measures and risk-adjusted outcome quality of care measures. In general, we estimate how changes in competition between hospital markets and over time affect the quality of care provided. We employ different fixed effects models and include an instrumental variable approach.

For emergency conditions AMI and stroke, patients are often taken to the nearest hospital without any hospital choice and thus much room for competition. However, by using risk-adjusted 30-day mortality for the two emergency conditions stroke and AMI, we achieve several benefits. AMI and stroke patients are often fairly sick and thus most influenced by general quality of care decisions made by the hospital. Overall response to competition and resulting adjustments in the provision of services will thus affect AMI and stroke patients in particular (Pauly et al., 2012). AMI and stroke represent complex emergency care situation, which can capture the overall organizational choice of quality of care to be provided and efficiency in diagnostic and therapeutic provision of care. Stroke care organization and quality has also received substantial management and political attention due to the stroke unit certification scheme of the German Stroke Society (DSG, 2015; Ringelstein et al., 2011). Furthermore, AMI and stroke are frequent medical conditions with a high outcome variation across hospitals (Pross, Geissler, Busse, 2017a). Selection bias will also be less of an issue for emergency conditions. Including post-discharge data allows us to control for the possibility that hospitals in more competitive areas discharge patients too early.

We also include risk-adjusted and unadjusted outcome measures for the elective conditions hip replacement. Here, risk-adjusted measures are only available from 2012 to 2014; however, unadjusted measures are available from 2006 to 2014, which we use for an additional model specification.

With regards to product markets, we employ a medical diagnosis or procedure specific competition measure. Based on the case volumes recorded for each hospital, we calculate the number of hospitals offering the respective procedure, their market share and HHI for each hospital market. In our empirical analysis, we focus on the HHI as it's a simple and robust competition measure, which puts more emphasis on hospitals with a larger market share, which is more in line with the consumer view (Laine, 1995). As described above, we focus on the two emergency conditions stroke and AMI and the elective condition hip replacement.

With regards to the geographical market, we employ a geographic radius of 30 km around each hospital´s geo-location (Figure 1). Patients travel on average 18 km to their hospital and the third quartile travel distance was 31 km in 2014 (Klauber et al., 2016). In flexible radius methods, however, articles often include the districts that add up to 80% of the patient population and a 30 km radius is also an often chosen distance (Bloom et al., 2015). We also vary our radius geographic market definition to ensure robustness of our results for 10, 20, 40 and 50 km radius.

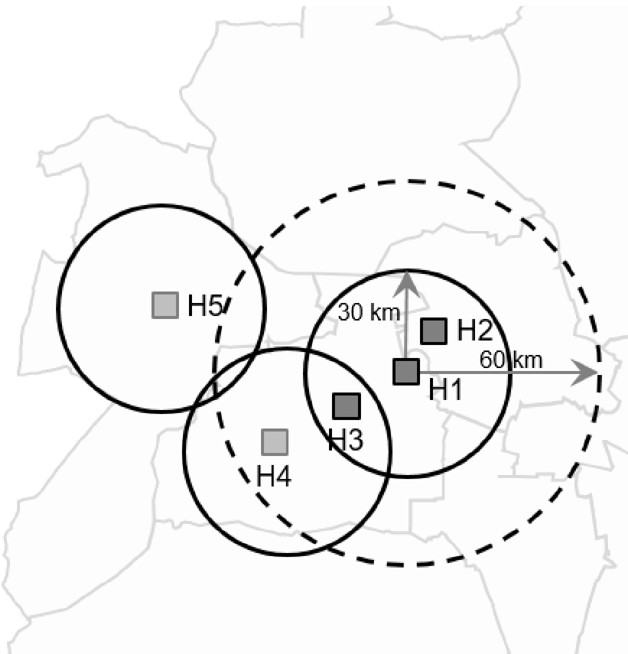



**Figure 1**: Graphical representation of geographical competition measure
Legend: Our competition measure includes all hospital that fall within a 30km radius of a certain hospital H1. H2 and H3 are included in that 30km catchment area. H3 is also included in the catchment area of H4. Any hospital (A4) with a 60km radius will have a competitive overall (to some degree) with hospital H1.

We characterize the competitive environment of the local hospital using two measures of competition. First, we consider the number of hospitals in a local market with more of 5 cases in the respective medical condition. Second, we calculate each hospital's market HHI, multiplied by 10,000 (Gaynor, 2006). Variation in competitive forces is occurring both with hospitals closing and local hospitals being acquired and expanded by a larger hospital group and by hospitals 'strategic decisions and investments to expand certain treatment areas and increase case volumes and market share. We also investigate whether a lagged competition effect might affect quality as quality investments, such as additional staff, infrastructure, medical equipment and training, might take time to realize their full impact on quality.

Much of the variation in market structure comes from urban vs. rural differences. Urban areas are more competitive with lower concentration in medical service provision, offer a higher intensity of services, with higher volumes and more experiences and thus also often better outcomes. It would be difficult to combine real world medical travel patterns, with patients travelling from rural to urban centers for medical services, with any other results.

We do not use HHI generated based on patient flows and the market from which the hospital draws large share of its patients (e.g. 80%), because reverse causality may result in endogenous patient flows (Gaynor et al., 2012). Patients may travel further for certain hospitals and certain medical conditions (e.g. high quality hospitals) and this may result in larger catchment areas which would appear less concentrated (Chou et al., 2014). At the same time, higher quality hospitals might also attract more patients within fixed, more refined market, which would lead to higher market shares for high quality hospitals and less concentration in markets with one quality leading hospital. Furthermore, patients attracted to higher quality hospitals might be especially sick, which can bias both outcomes measures (if risk-adjustment is incomplete) and market shares. This leads to potential reverse causality, with the regressor degree of competition depending on the dependent variable quality of care. To address this endogeneity problem, we expand our empirical strategy with an instrumental variable approach. We instrument the HHI with the in Germany exogenous – due to political considerations – variable number of hospitals in a respective geographic market.

We include several, in the literature common control variables to account for important time invariant cofounders. First, we include ownership type to adjust for different profit maximizing strategies for private for-profit, private non-for-profit and public hospitals. We also include teaching status to control for possibly faster dissemination of evidence based care, but also more severe patients being treated in university affiliated hospitals. Furthermore, we control for overall hospital size, captured through number of hospital beds, to adjust for large organization advantages (e.g. larger capital budgets) and disadvantages (organizational complexity). We also adjust for share of patients in the respect treatment area as part of all patients treated to account for the treatment area's relative importance. We account for the general level of hospital specialization via a Category Medical Spezialization (CMS) index developed by (Lindlbauer and Schreyögg, 2014). To ensure robustness of our regression model, we include hospital fixed effects to pick-up observed and unobserved time-invariant differences between hospitals.

Based on the above theoretical and empirical considerations, our empirical model takes the following general form:

$$\log(QI_{it})^k = \beta_0 + \beta_1 \log(HHI_{it})^l + \beta_2 \log(CV_{it})^l + \beta_3 log(\frac{CV_{it}}{all\ cases_{it}})^l$$
$$+ \beta_4 \log(hospital\ beds)_{it} + \beta_5 \log(CMS\ ICD)_{it} + \beta_5\ teach_{it} \quad (1)$$
$$+ \beta_6\ private_{it} + \beta_7\ public_{it} + \alpha_i + \tau_t + \varepsilon_{it},$$

where $QI_{it}{}^k$, $HHI_{it}{}^k$, $CV_{it}{}^k$ $(\frac{CV_{it}}{all\ cases_{it}})^k$are a set of *k* outcome quality indicators, *k* HHI competition, *k* absolute case volume, and *k* relative case volume share measures for *l* treatment areas stroke, AMI and hip replacement. $\beta_0$ is the intercept, $\alpha_i$ are individual time invariant hospital-fixed effects and $\varepsilon_{it}$ captures robust white standard errors. In addition, $teach_{it}$, $private_{it}$ and $public_{it}$ are dummy variables that capture whether a hospital has a teaching status and has private, for-profit or public ownership, with private, non-profit being the reference group. The observations are weighted by the case volume to adjust the outcome quality effect for the relative size and impact of each hospital.

In addition to the general model specified above, we also estimate regression equations where we replace the same year competition effect $HHI_{it}{}^k$ with a lagged competition effect $HHI_{it-1}{}^k$. This accounts for the possibility that competitive pressures trigger provider responses such as increased investment in quality-enhancing equipment or changes in internal quality standards, which affect the level of outcome quality provided only with a 1-2-year delay due to organizational sluggishness and time required for training and implementation of new practices.

Furthermore, we instrument for the competition effect on quality of care via the number of hospitals offering the respective medical service. As described above, the number of hospitals offering a respective medical services will likely be less influenced by quality of care provided, as hospital capacity planning is a lengthy regulatory process upon which quality of care currently has limited impact (IQTIG, 2016; Pross, Geissler, Busse, 2017a). But hospitals mergers and the acquisition of publicly owned hospitals by privately owned hospitals is still often triggered more by financial and political considerations and less so quality of care considerations.

## 4. Results

We first examine data patterns for our main competition and outcome quality measures. Table 1 displays the time trends for average number of competing hospitals, average market share and average HHI. For stroke, the average number of hospitals in each hospital market decreases slightly from 19.9 in 2006 to 19.4 in 2014. The average market share increases slightly, but the HHI increases more substantially to 2393 in 2014 (20% increase). Since the HHI emphasizes larger market shares, stroke care has further consolidated with hospitals that had already treated a large share of hospitals before. For AMI, the number of hospital increases slightly and the HHI remains constant at roughly 2,200 points. For hip replacement surgery, the number of hospitals offering this procedure in each regional market increases from 16 to 18 and the HHI declines accordingly to 2140. In 2014, stroke care is relatively more concentrated than AMI care and hip replacement surgery.

**Table 1:** Competition and quality variables time trends, 2006 – 2014

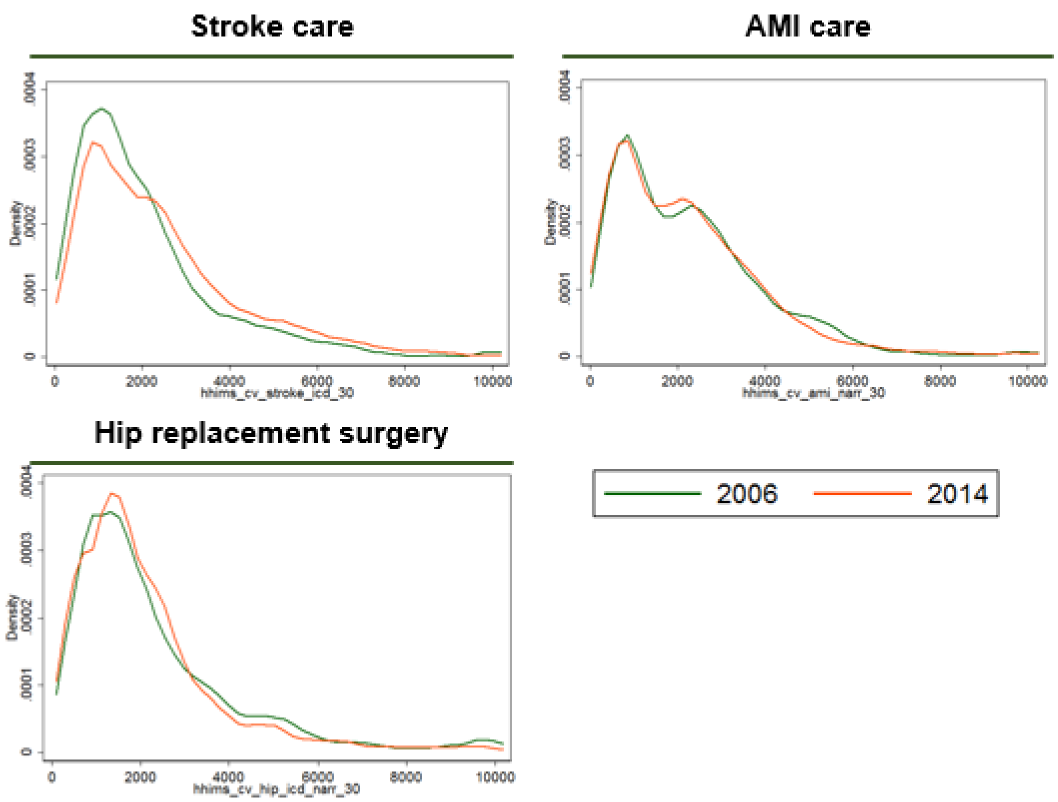
| Variables | 2006 | 2008 | 2010 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|
| Stroke number of observations | 1288 | 1245 | 1200 | 1183 | 1166 | 1125 |
| AMI number of observations | 1216 | 1193 | 1125 | 1122 | 1113 | 1074 |
| Hip number of observations | 1050 | 1133 | 1117 | 1060 | 1122 | 1122 |
| <u>*Competition measures for 30 km*</u> | | | | | | |
| *Stroke care* | | | | | | |
| Number of hospitals (#) | 19.9 | 20.3 | 19.7 | 19.9 | 19.1 | 19.4 |
| Average market shares (%) | 0.10 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
| Average HHI (index) | 1976 | 2075 | 2265 | 2292 | 2420 | 2393 |
| *AMI care* | | | | | | |
| Number of hospitals (#) | 18.7 | 19.7 | 19.3 | 19.9 | 19.1 | 19.5 |
| Average market shares (%) | 0.11 | 0.11 | 0.11 | 0.11 | 0.12 | 0.12 |
| Average HHI | 2227 | 2238 | 2247 | 2179 | 2233 | 2206 |
| *Hip replacement surgery* | | | | | | |
| Number of hospitals (#) | 16.3 | 17.0 | 17.5 | 18.3 | 18.0 | 18.3 |
| Average market shares (in %) | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |
| Average HHI | 2335 | 2191 | 2169 | 2130 | 2125 | 2140 |
| <u>*Outcome quality of care*</u> | | | | | | |
| *Stroke care* | | | | | | |
| 30-day SMR (observed/expected) | 1.12 | 1.09 | 1.06 | 1.05 | 1.03 | 0.99 |
| Unadjusted 30-day mortality (in %) | 14.7 | 14.4 | 14.1 | 13.9 | 13.5 | 13.2 |
| Unadjusted 90-day readmission (in %) | 22.4 | 23.3 | 23.8 | 23.6 | 24.0 | 24.0 |
| *AMI care* | | | | | | |
| 30-day SMR (observed/expected) | 1.12 | 1.16 | 1.13 | 1.05 | 1.07 | 1.00 |
| simple 30-day mortality (%) | 16.1 | 15.4 | 15.4 | 13.3 | 13.4 | 12.4 |
| simple 90-day readmission (%) | 28.4 | 29.6 | 31.2 | 32.0 | 32.5 | 33.1 |
| *Hip replacement surgery* | | | | | | |
| Inpatient risk-adj. rate of re-interventions (obs./exp.) | | | | 1.09 | 1.34 | 1.11 |
| Inpatient unadj. revisions due to complications (%) | 3.36 | 1.88 | 1.65 | 1.68 | 1.63 | 1.53 |

When looking at outcome quality of care trends, we noticed a substantial improvement for both risk-adjusted and unadjusted 30-day mortality in stroke and AMI care. However, the unadjusted 90-day readmission rate increases for both. For hip replacement surgery, the unadjusted rate of revisions due to complications increases substantially from 2006 to 2010, but remains fairly constant thereafter. The risk-adjusted rate of re-interventions, which is only available after 2010, increases slightly.

To further examine whether patient flows have changed, we plot kernel density estimates in 2006 and 2014 for our main measure of market concentration HHI (Figure 1). For stroke, we can identify a clear rightward shift with concentration increasing from 2006 to

2014. For AMI, the concentration levels remain roughly comparable. Hip replacement surgery becomes less concentrated. The results are in line with the trends displayed Table 1.

**Figure 1**: Kernel density estimates for the distribution of HHI, 2006 vs. 2014




If competition affects outcome quality of care positively, then we would expect the 30-day SMR to be higher in more concentrated markets. Table 2 presents the average 30-day SMR for three different categories of concentration, based on the U.S. Department of Justice horizontal merger guidelines. For stroke, the 30-day SMR is higher (quality is worse) in more concentrated markets. For AMI, the SMR increases for moderately concentrated markets, but decreases again for highly concentrated markets. For hip replacement surgery, outcome quality is highest in moderately concentrated markets and lowest for highly concentrated markets. Across treatment areas, no clear pattern emerges.

**Table 2:** Risk-adjusted outcome indicators in differently concentrated markets, all years

| | | All hospitals | Unconcentrated (HHI < 1500) | Moderately concentrated (1500 ≤ HHI ≤ 2500) | Highly concentrated (HHI<2500) |
|---|---|---|---|---|---|
| **Stroke care** | 30-day SMR | 1.06 | 1.05 | 1.05 | 1.07 |
| | # of obs. | 7207 | 3017 | 1871 | 2319 |
| **AMI care** | 30-day SMR | 1.09 | 1.08 | 1.11 | 1.08 |
| | # of obs. | 6843 | 2801 | 1503 | 2539 |
| **Hip replacement surgery** | Risk-adjusted complications | 1.2 | 1.2 | 0.9 | 1.3 |
| | # of obs. | 3236 | 1439 | 907 | 890 |

**Legend**: HHI categories based on department of justice definition, all years included in calculation of HHI groups, count of hospitals in each group displayed

Before presenting our regression results, we also display descriptive statistics for our regression variables in Table 3. The overview shows outcome quality measures and competition measures for each treatment area stroke, AMI, and hip replacement as well as control variables. Risk-adjusted 30-day mortality and unadjusted 90-day readmissions for stroke and AMI outcomes have a standard deviation that is 25-30% of the mean. The distribution for risk-adjusted rate of re-interventions and unadjusted rate of revisions due to complications for hip replacement surgery has a substantially wider spread. The hip replacement indicators also have a lower number of observations, with the risk-adjusted hip outcome indicator only with data from 2012 – 2014 and the unadjusted outcome indicator with less observations since some hospitals exit the sample due to low case volumes and new data privacy specifications.

**Table 3.** Descriptive statistics (Number of observations, mean, standard deviation, minimum, maximum)

| Variables | Observations | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| *Outcome quality measures* | | | | | |
| Risk-adjusted 30-day mortality stroke (SMR) | 7207 | 0.71 | 0.18 | 0 | 2.94 |
| 90 day readmissions stroke | 7207 | 3.44 | 0.90 | 0 | 5.08 |
| Risk-adjusted 30-day mortality Ami (SMR) | 6843 | 0.72 | 0.20 | 0 | 3.05 |
| 90 day readmissions Ami | 6843 | 3.32 | 0.88 | 0 | 5.12 |
| Risk -adjusted rate of re-interventions | 3304 | 0.40 | 0.72 | 0 | 5.66 |
| Revision due to complications | 5154 | 0.87 | 0.59 | 0 | 7.51 |
| *Competition measures* | | | | | |
| Log (number of hospitals stroke) | 9977 | 2.59 | 0.86 | 0 | 4.50 |
| Log (HHI Stroke) | 9977 | 7.45 | 0.77 | 5.42 | 9.21 |
| Log (number of hospitals AMI) | 9971 | 2.53 | 0.89 | 0 | 4.49 |
| Log (HHI_AMI) | 9971 | 7.42 | 0.81 | 5.54 | 9.21 |
| Log (number of hospitals hip replacement) | 9883 | 1.27 | 0.88 | 0 | 3.69 |
| Log (HHI Hip replacement) | 9883 | 7.41 | 0.76 | 5.58 | 9.21 |
| *Control variables* | | | | | |
| Log (stroke case volume) | 8048 | 4.31 | 1.63 | 0 | 8.58 |
| Log (AMI case volume) | 7636 | 4.28 | 1.51 | 0 | 7.72 |
| Log (Hip replacement case volume) | 7616 | 3.90 | 1.64 | 0 | 7.65 |
| Share stroke ICD/all ICD | 9920 | 0.02 | 0.05 | 0 | 1.21 |
| Share AMI ICD/all ICD | 9920 | 0.01 | 0.01 | 0 | 0.57 |
| Share hip replacement ICD/all ICD | 9920 | 0.01 | 0.03 | 0 | 0.57 |
| Log (hospital beds) | 9920 | 5.41 | 0.76 | 3.91 | 8.07 |
| Teaching hospital | 9977 | 0.07 | 0.66 | -3.14 | 1.52 |
| CMS ICD | 9977 | 0.43 | 0.49 | 0 | 1.00 |
| Private | 9977 | 0.22 | 0.42 | 0 | 1.00 |
| Public | 9977 | 0.35 | 0.48 | 0 | 1.00 |

Table 4 displays the first set of regression results. For our three treatment areas stroke, AMI and hip replacement, we regress risk-adjusted outcome quality on the 30 km HHI for the ICD case volume market shares. The competition effect on outcome quality for stroke and hip replacement is positive, for AMI the effect is negative; however, all three do not exhibit statistical significance. Next to our main predictor of interest, the case volumes as well as the share of patient cases in the analyzed treatment areas relative to the other treatment are insignificant. The F statistics for all models indicates overall model significance despite the lack of individual coefficients. The models display low R-squared measures, however, since fixed effects regressions washout the explanatory effects of the intercepts, they generally show a much lower R-squared compared to ordinary least squares models. The results are compared for regression with 15km and 45 km radius competition variables.

**Table 4.** Regression results for risk-adjusted outcome quality indicators

| **Variables** | **M1** (Stroke) | **M2** (AMI) | **M3** (Hip) |
|---|---|---|---|
| Log (HHI Stroke) | 0.025 | | |
| | (0.01) | | |
| Log (HHI AMI) | | -0.006 | |
| | | (0.02) | |
| Log (HHI hip replacement) | | | 0.096 |
| | | | (0.21) |
| Log (stroke case volume) | 0.006 | | |
| | (0.01) | | |
| Log (AMI case volume) | | 0.000 | |
| | | (0.01) | |
| Log (hip replacement case volume) | | | -0.038 |
| | | | (0.07) |
| Share stroke ICD/all ICD | 0.008 | | |
| | (0.06) | | |
| Share AMI ICD/all ICD | | -0.257 | |
| | | (0.20) | |
| Share hip replacement ICD/all ICD | | | 0.060 |
| | | | (0.66) |
| Log (hospital beds) | -0.048 | -0.031 | -0.423*** |
| | (0.02) | (0.00) | (0.12) |
| Teaching hospital | 0.004 | 0.004 | -0.045 |
| | (0.01) | (0.01) | (0.11) |
| Log (CMS ICD) | -0.004 | 0.013 | 0.370** |
| | (0.01) | (0.02) | (0.14) |
| Private | -0.015 | 0.043 | 0.006 |
| | (0.03) | (0.04) | (0.12) |
| Public | -0.014 | 0.048 | -0.037 |
| | (0.03) | (0.04) | (0.18) |
| Constant | 0.818*** | 0.920*** | 2.248 |
| | (0.20) | (0.26) | (1.69) |
| Hospital fixed effects | yes | yes | yes |
| Time fixed effects | yes | yes | yes |
| Number of observations | 7044 | 6688 | 3248 |
| R2-within | 0.017 | 0.022 | 0.008 |
| R2-between | 0.008 | 0.004 | 0.031 |
| R2-overalll | 0.002 | 0.000 | 0.010 |
| BIC | -7411.06 | -5876.85 | 5204.88 |
| Intraclass correlation | 0.74 | 0.73 | 0.47 |
| F-statistic | 5.35 | 7.44 | 2.12 |

Note: * p<0.05, ** p<0.01, *** p<0.001; the coefficients for the time effect show high significance in the stroke and AMI equation (for stroke the coefficient for 2008 is -0.020**, for 2010 -0.030***, for 2012 -0.033***, for 2013 -0.047***, for 2014 -0.064***; for AMI the coefficient for 2008 is 0.017*, for 2010 0.006, for 2012 -0.027**, for 2013 -0.018*, for 2014 -0.054***)

**Table 5.** Regression results for unadjusted outcome quality indicators

| Variables | M4 (stroke) | M5 (AMI) | M6 (hip) |
|---|---|---|---|
| Log (HHI_Stroke) | -0.074<br>(0.04) | | |
| Log (HHI_AMI) | | -0.022<br>(0.06) | |
| Log (HHI hip replacement) | | | -0.150<br>(0.08) |
| Log (stroke case volume) | 0.411***<br>(0.06) | | |
| Log (AMI case volume) | | 0.521***<br>(0.05) | |
| Log (hip replacement case volume) | | | 0.027<br>(0.03) |
| Share stroke ICD/all ICD | 0.213<br>(0.38) | | |
| Share AMI ICD/all ICD | | -1.486<br>(0.92) | |
| Share hip replacement ICD/all ICD | | | -1.179*<br>(0.59) |
| Log (hospital beds) | -0.031<br>(0.10) | 0.026<br>(0.09) | -0.041<br>(0.10) |
| Teaching hospital | -0.068*<br>(0.03) | 0.073*<br>(0.03) | -0.032<br>(0.05) |
| Log (CMS ICD) | -0.100***<br>(0.04) | -0.137***<br>(0.04) | -0.036<br>(0.04) |
| Private | -0.057<br>(0.06) | 0.024<br>(0.09) | 0.137<br>(0.09) |
| Public | 0.045<br>(0.04) | 0.078<br>(0.08) | 0.148<br>(0.12) |
| Constant | 1.849*<br>(0.82) | 0.322<br>(0.68) | 2.259**<br>(0.85) |
| Hospital fixed effects | yes | yes | yes |
| Time fixed effects | yes | yes | yes |
| Number of observations | 7044 | 6688 | 4999 |
| R2-within | 0.225 | 0.301 | 0.048 |
| R2-between | 0.695 | 0.804 | 0.003 |
| R2-overalll | 0.623 | 0.718 | 0.017 |
| BIC | 1299.89 | 1493.59 | 6038.77 |
| Intraclass correlation | 0.87 | 0.76 | 0.5 |
| F-statistic | 16.21 | 24.06 | 10.78 |

Note: * p<0.05, ** p<0.01, *** p<0.001

The results for the unadjusted outcome quality indicator regressions are displayed in Table 5. The coefficients for the competition variables are negative and, as before, insignificant. In line with other studies, the case volume becomes highly significant with large positive coefficients (Pross et al.; Pross, Berger et al., 2017). For both 90-day readmission rates for stroke and for AMI, a one unit increase in case volumes increases the readmission rate by 40-50%. This might be attributed to the fact that larger hospitals in tendency will care for a higher share of more complex cases; however, the magnitude of the effect is surprising. The fact that these outcome indicators are not risk-adjusted reduces robustness of these results. However, the independence between hospital competition levels and outcome quality is also confirmed for these non-risk-adjusted indicators.

Regression models M7 – M11 in Table 6 return to the risk-adjusted outcome measures 30-day standardized mortality rate for stroke and AMI and reintervention rate for hip replacement. Model M7 and M8 include a 2-year lagged competition effect for the data years 2008, 2010, 2012 and 2014. As before, no significant relationship between competition and quality of care can be detected.

With M9-M11, we implement an instrumental variable approach. We instrument our main competition measure 30 km HHI with the number of hospital offering the respective medical service. The signs for the competition coefficients remain as in M1-M3, the effect size increases; however, the coefficients remain insignificant.

As robustness checks, we also estimate different model specifications with HHI competition measures based on the district radius 10 and 20 km and 40 and 50 km. The results are comparable to the results displayed in Table 4-6.

**Table 6.** Regression results for unadjusted outcome quality indicators

| Variables | M7 (stroke) | M8 (AMI) | M9 (stroke) | M10 (AMI) | M11 (hip) |
|---|---|---|---|---|---|
| Lagged log (HHI Stroke) | 0.009 | | | | |
| | (0.01) | | | | |
| Lagged log (HHI AMI) | | 0.011 | | | |
| | | (0.03) | | | |
| Log (HHI_Stroke) | | | 0.066 | | |
| | | | (0.03) | | |
| Log (HHI_AMI) | | | | -0.030 | |
| | | | | (0.05) | |
| Log (HHI hip replacement) | | | | | -0.168 |
| | | | | | (4.32) |
| Log (stroke case volume) | 0.021 | | -0.003 | | |
| | (0.01) | | (0.01) | | |
| Log (AMI case volume) | | -0.004 | | -0.006 | |
| | | (0.01) | | (0.01) | |
| Log (hip replacement case volume) | | | | | -0.043 |
| | | | | | (0.01) |
| Share stroke ICD/all ICD | -0.170 | | -0.008 | | |
| | (0.21) | | (0.09) | | |
| Share AMI ICD/all ICD | | -0.061 | | -0.065 | |
| | | (0.46) | | (0.20) | |
| Share hip replacement ICD/all ICD | | | | | 0.328 |
| | | | | | (6.38) |
| Log (hospital beds) | -0.054 | -0.054 | -0.035 | -0.030 | -0.420** |
| | (0.03) | (0.04) | (0.03) | (0.03) | (0.03) |
| Teaching hospital | -0.000 | 0.010 | -0.021 | -0.014 | -0.037 |
| | (0.02) | (0.02) | (0.01) | (0.01) | (0.18) |
| log (CMS ICD) | -0.012 | -0.016 | -0.001 | 0.011 | 0.377* |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.17) |
| Private | -0.010 | 0.095* | -0.044 | 0.023 | 0.022 |
| | (0.03) | (0.05) | (0.03) | (0.04) | (0.29) |
| Public | -0.046 | 0.048 | -0.021 | 0.043 | -0.033 |
| | (0.03) | (0.04) | (0.03) | (0.03) | (0.20) |
| Constant | 0.899*** | 0.953** | | | |
| | (0.22) | (0.33) | | | |
| Hospital fixed effects | yes | yes | yes | yes | yes |
| Time fixed effects | yes | yes | yes | yes | yes |
| Number of observations | 4619 | 4398 | 6935 | 6580 | 3203 |
| R2-within | 0.016 | 0.037 | | | |
| R2-between | 0.000 | 0.018 | | | |
| R2-overalll | 0.000 | 0.001 | | | |
| BIC | -5288.84 | -4413.31 | -7130.99 | -5700.73 | 5156.97 |
| Intraclass correlation | 0.75 | 0.75 | | | |
| F-statistic | 3.82 | 8.75 | 1.97 | 5.44 | 1.75 |

Note: * p<0.05, ** p<0.01, *** p<0.001

## 5. Discussion

From the multivariate fixed-effect as well instrumental variable-enhanced fixed-effect analyses, we found no significant relationship between market competition measures and outcome quality for the treatment areas stroke, AMI and hip replacement. Because we suspected that changes to the market structure might affect quality only after some time, we also tested a model with lagged competition variables, but the coefficients were similarly insignificant. To ensure causal robustness, we also instrumented the competition effect via number of hospitals that offer the respective medical service; however, the results remain insignificant.

While several previous studies have found significant positive effects of competition on quality, many other studies have also found mixed effects for different indicators in the same treatment areas (mortality vs. readmission) (Palangkaraya and Yong, 2013), no effect of competition on quality of care (Ho and Hamilton, 2000; Maeda and Lo Sasso, 2012) or even a significant negative effect of competition on quality of care (Propper et al., 2004). Our results are thus in line with the mixed and contested literature (Goddard, 2015). Furthermore, based on somewhat altruistic providers and a general cost structure, Brekke et al. (2011) provide a theoretical underpinning for the empirically mixed effects.

### 5.1. Factors limiting hospital competition on quality in Germany

The apparently non-existent empirical relationship between competition and quality of care can also be explained by several important structural limitations on quality of care competition in Germany. Despite increased public reporting efforts in Germany over the past years, hospital competition on quality might still be constrained by a lack of a general awareness on quality differences between providers (Langner et al., 2016) and limited meaningful and actionable information on quality of care. Current quality of care transparency has a two-year lag which obscures the quality signal substantially (Pross, Geissler, Busse, 2017a). Furthermore, current public reporting via portals such as Weisse Liste.de might be ineffective in guiding patients to high quality hospitals. The number of users is still relative small, the abortion rate (without seeing hospital quality information) is almost 50%, and the usage of the benchmarking function is negligible (Pross, Averdunk et al., 2017). A large share of patients is also not aware of information availabilities (Langner et al., 2016). More generally, research has shown that many patients still do not actively choose their hospital, rather they just go to their local hospital or the one they previously attended, even for a totally different and unrelated condition (Victoor et al., 2012). Patients hospital choice is also heavily influenced by guidance from their outpatient physicians. However, general practioners as well specialists often question usefulness and credibility of public reporting data and their usage of public reporting information (Hermeling and Geraedts, 2013; Pross, Averdunk et al., 2017; Schneider and Epstein, 1996). Public reporting, which is supposed to make quality of care transparent to stimulate quality competition, has not had the impact many have hoped or still expect (Kumpunen et al., 2014).

In Germany, quality competition is further limited by the collective reimbursement, in which payers have to reimburse hospitals independent of the quality of the health care provided (GKV Spitzenverband, 2014). Selective reimbursement contracts between payers and providers have only recently emerged as a possible mechanism to increase quality competition (AOK Hessen, 2016; BT-Drs. 18/5372, 2015). When selective contracting increased in the US, many hospitals merged to increase their bargaining power. Between 1997-2006, the HHI increases by 1,000 points or more in 39 metropolitan areas (Pauly et al., 2012) Part of this competitive pressure, triggered via selective contracting, was channeled into competition on quality of care (Chou et al., 2014; Kessler and Geppert, 2005; Kessler and McClellan, 2000).

Provider competition and patient choice of hospitals might occur on other quality of care dimensions, which we cannot capture with the available hospital data. Patients choose hospitals that improve their self-reported health. Quality as measured by readmission and mortality rates is less important for patients (Gutacker et al., 2016). Patient experiences, e.g., captured via the Patient Experience Questionnaire (PEQ) are another dimension, which might actually be affected by provider competition.

Furthermore, the discrete nature of hospital market boundaries used in this paper assumes that providers are either fully inside or completely outside of the relevant geographic market. This leads to measurement error in geographic markets and the associated competition measure, which in turn can bias the estimated effect of competition toward zero (Kessler and McClellan, 2000).

Lastly, the trade-off between centralization and competition between providers needs to be taken into account. The literature shows a robust positive effect of concentration on quality of care, especially when patient volumes increase from a lower to a medium-high level (Birkmeyer et al., 2002; Hannan et al., 2005; Hernandez-Boussard et al., 2012). At larger volumes, the effect is often more ambiguous and might be counterweight by reduced quality of care competition at very high levels of competition. Care concentration in high volume, high experience centers necessarily implies even lower volumes or no service offering in hospitals with originally smaller volumes and experience. The threshold at which point higher care concentration might actually have a negative impact on quality of care competition is unclear and needs to be examined more closely. Going forward the benefits of specialization and centralization of care need to be weighted by the benefits of patient choice and competition between providers.

In light of these concerns, the German competition authority the Bundeskartellamt (BKartA – Federal Cartel Office) has recently launched an investigation into the hospital sector to examine competitive conditions and provide a fact base for the many ongoing and expected hospital merger control proceedings (Bundeskartellamt, 2016). The examination focuses in particular on quality competition between hospitals, how patients choose hospitals and hospitals´ capabilities to respond to market changes within the constrain of state hospital planning and several other regulations, such as the social law (Bundeskartellamt, 2016).

## 5.2. Limitations

The degree of market concentration might be inaccurately measured due to two concerns. On the one hand, as mentioned above, we employ a fixed radius methodology to calculate market shares due to lack of patient level data. For some hospitals, this will lead to an overestimation of the market size and thus incorrectly low market shares while other hospitals will draw from an even larger market, especially for elective medical conditions such as hip replacement. In fact, the relevant market might differ by condition and region, so using patient data and choice models for market definition and competition levels will likely lead to more accurate market sizes (Zwanziger et al., 1994). On the other hand, two hospitals in one market might owned by the same entity, which we do not control for as we measure competition at the single hospital or hospital site level and not the ownership group level. Not taking into account system membership may potential underestimate the level of market concentration (Hentschker et al., 2014). This is currently being addressed via standardization of the ownership information across years in the underlying database.

Furthermore, we employ a fixed-effect framework that allows us to indirectly control for hospital characteristics for which we do not have data, such as hospital management sophistication and hospital organization culture. However, due to the annual measures of outcome rates in the panel dataset, we also have to content with the higher noise associated with annual measures (Propper et al., 2004). This might contribute to the strongly insignificant results.

Lastly, financial hospital data, at an overall level or at a medical condition level, is not available in standardized fashion, so the effect of investment capacity or profitability cannot be accounted for. Since more than a third of  Germany operate at negative margins (Augurzky et al., 2016), the theoretical precondition that medical service for additional patients has positive contribution margin might not hold for all hospitals and cannot be controlled for due to the lack of data availability.

## 6. Conclusion

We have examined the impact of competition on the quality of care in three different medical conditions: stroke AMI as an emergency conditions and hip replacement as an elective procedure. We find no evidence that hospital competition has an effect on quality, not for the emergency treatment areas stroke and AMI nor for the elective procedure hip replacement. In line with inconclusive results in some studies and the mixed evidence in the literature, this study cannot confirm the often-intuitive positive relationship between provider competition and quality of care.

The theoretical considerations have shown that an important prerequisite for a positive effect of competition on quality is patients' awareness for the outcome differences between hospitals, their willingness to act on this information and thus 'vote with their feet' by choosing high quality hospitals. However, despite ongoing public reporting efforts and mixed evidence, this mechanism appears still rather weak in Germany.

Studies in the UK have benefited from a quasi-natural experiment due to policy changes and the streamlined regulatory structure of the NHS to examine how differences in competition over time and space affect quality before and after quality of care is made public. In Germany, the publication of hospital report cards in mid-2008 might provide a similar opportunity. Exploiting this, additional research shall employ propensity score matching and a difference-in-difference model to examine how outcome quality of care has evolved in hospitals in competitive districts and concentrated districts, ceterius paribus. In conjunction with the difference-in-difference approach, a nation-wide patient data for the relevant treatment areas and associated patient choice models can help to overcome the limitations of fixed effect radius approach.

**Acknowledgements** We thank the WIdO for granting us access to the QSR outcome data. We also thank the G-BA for making the German hospital report cards accessible.

**Contributors** Lead author is Christoph Pross.

**Funding** Christoph Pross has received a general PhD scholarship from the Konrad-Adenauer-Foundation.

**Competing interests** No competing interests exist.

**Patient consent** Not applicable.

## References


AOK Hessen (2016) *Planung von Selektivverträgen nach § 140 SGB V: Alloplastischer Gelenkersatz - Hüftendoprothetik -,* AOK Hessen [Online]. Available at http://www.aok-gesundheitspartner.de/he/krankenhaus/qs/selektivvertraege/index.html (Accessed 8 February 2017).

Augurzky, B., Krolop, S., Pilny, A., Schmidt, C. M. and Wuckel, C. (2016) *Hospital Rating Report: With tail wind into the future?*, Heidelberg, medhochzwei.

Birkmeyer, J. D., Siewers, A. E., Finlayson, E. V. A., Stukel, T. A., Lucas, F. L., Batista, I., Welch, H. G. and Wennberg, D. E. (2002) 'Hospital volume and surgical mortality in the United States', *The New England journal of medicine*, vol. 346, no. 15, pp. 1128–1137.

Bloom, N., Propper, C., Seiler, S. and van Reenen, J. (2015) 'The Impact of Competition on Management Quality: Evidence from Public Hospitals', *The Review of Economic Studies*, vol. 82, no. 2, pp. 457–489.

Blumenthal, D., Abrams, M. and Nuzum, R. (2015) 'The Affordable Care Act at 5 Years', *The New England journal of medicine*, vol. 372, no. 25, pp. 2451–2458.

Brekke, K. R., Siciliani, L. and Straume, O. R. (2011) 'Hospital Competition and Quality with Regulated Prices*', *Scandinavian Journal of Economics*, vol. 113, no. 2, pp. 444–469.

BT-Drs. 18/5372 (2015) 'Drucksache des Deutschen Bundestages 18/5372 vom 30. Juni 2015:: Entwurf eines Gesetzes zur Reform der Strukturen der Krankenhausversorgung (Krankenhausstrukturgesetz – KHSG)', *Bundestagdrucksache*.

Bundeskartellamt (2016) *Sector inquiry into hospitals* [Online], Bundeskartellamt. Available at https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2016/31_05_2016_Sektoruntersuchung_Krankenhaeuser.html;jsessionid=4F31E76D16CDB29A0B546BFDEEB4944C.1_cid387?nn=3591568 (Accessed 16 November 2016).

Busse, R., Ganten, D., Huster, S., Reinhardt, E., Sutturp, N. and Wiesing, U. (2016) *Zum Verhältnis von Medizin und Ökonomie im deutschen Gesundheitssystem*, Halle, Nationale Akademie der Wissenschaften.

Chandra, A., Finkelstein, A., Sacarny, A. and Syverson, C. (2016) 'Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector', *The American economic review*, vol. 106, no. 8, pp. 2110–2144.

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

Cooper, Z., Gibbons, S., Jones, S. and McGuire, A. (2011) 'Does Hospital Competition Save Lives? Evidence from the English NHS Patient Choice Reforms', *Economic journal (London, England)*, vol. 121, no. 554, F228-F260.

Copnell, B., Hagger, V., Wilson, S. G., Evans, S. M., Sprivulis, P. C. and Cameron, P. A. (2009) 'Measuring the quality of hospital care: an inventory of indicators', *Internal medicine journal*, vol. 39, no. 6, pp. 352–360.

Dranove, D. and Satterthwaite, M. A. (2012) 'The industrial organization of health care markets: Chapter 20', in Pauly, M. V., McGuire, T. G. and Barros, P. P. (eds) *Handbook of Health Economics,* Amsterdam, London, North Holland, pp. 1093–1139.

DSG (2015) *SU-Zertifizierungskriterien 2015* [Online], German Stroke Society. Available at http://www.dsg-info.de/stroke-units/zertifizierungsantraege--zertifizierungskriterien.html (Accessed 6 December 2016).

Ellis, R. P. and McGuire, T. G. (1986) ‘Provider behavior under prospective reimbursement. Cost sharing and supply’, *Journal of Health Economics*, vol. 5, no. 2, pp. 129–151.

Friedrich, J. and Beivers, A. (2009) ‘Patientenwege ins Krankenhaus: Räumlichliche Mobilität bei Elektiv- und Notfallleistungen am Beispiel der Hüftendoprothesen’, in Klauber, J. and Beivers, A. (eds) *Schwerpunkt: Versorgungszentren,* Stuttgart [u.a.], Schattauer, pp. 155–181.

Gandjour, A., Kleinschmit, F., Littmann, V. and Lauterbach, K. W. (2002) ‘An evidence-based evaluation of quality and efficiency indicators’, *Quality management in health care*, vol. 10, no. 4, pp. 41–52.

Gaynor, M. (2006) ‘What Do We Know About Competition and Quality in Health Care Markets?’, *NBER Working Paper*, no. 12301 [Online]. DOI: 10.3386/w12301 (Accessed 16 November 2016).

Gaynor, M., Moreno-Serra, R. and Propper, C. (2012) ‘Can competition improve outcomes in UK health care? Lessons from the past two decades’, *Journal of health services research & policy*, 17 Suppl 1, pp. 49–54.

Gaynor, M., Moreno-Serra, R. and Propper, C. (2013) ‘Death by Market Power: Reform, Competition, and Patient Outcomes in the National Health Service’, *American Economic Journal: Economic Policy*, vol. 5, no. 4, pp. 134–166.

G-BA (2014) *Regelungen zum Qualitätsbericht der Krankenhäuser, 2013* [Online], Gemeinsame Bundesausschuß. Available at https://www.g-ba.de/downloads/62-492-906/Qb-R_2014-06-19.pdf (Accessed 9 September 2015).

Geissler, A., Wörz, M. and Busse, R. (2010) ‘Deutsche Krankenhauskapazitäten im internationalen Vergleich’, in Klauber, J. and Augurzky, B. (eds) *Schwerpunkt: Krankenhausversorgung in der Krise?: Mit 100 Tabellen ; [mit Online-Zugang],* Stuttgart, Schattauer, pp. 25–40.

GKV Spitzenverband (2014) *Qualitätsorientierte Versorgungssteuerung und Vergütung: Positionen des GKV-Spitzenverbandes* [Online], Berlin, GKV Spitzenverband. Available at https://www.gkv-spitzenverband.de/krankenversicherung/qualitaetssicherung_2/qualitaetssicherung_2.jsp (Accessed 23 December 2016).

Goddard, M. (2015) ‘Competition in Healthcare: Good, Bad or Ugly?’, *International journal of health policy and management*, vol. 4, no. 9, pp. 567–569.

Gowrisankaran, G. and Town, R. J. (2003) ‘Competition, Payers, and Hospital Quality1’, *Health Services Research*, vol. 38, 6p1, pp. 1403–1422.

Gutacker, N., Siciliani, L., Moscelli, G. and Gravelle, H. (2016) ‘Choice of hospital: Which type of quality matters?’, *Journal of Health Economics*, vol. 50, pp. 230–246.

Hannan, E. L., Wu, C., Walford, G., King, S. B. 3., Holmes, D. R., JR, Ambrose, J. A., Sharma, S., Katz, S., Clark, L. T. and Jones, R. H. (2005) ‘Volume-outcome relationships for percutaneous coronary interventions in the stent era’, *Circulation*, vol. 112, no. 8, pp. 1171–1179.

Hentschker, C., Mennicken, R. and Schmid, A. (2014) ‘Defining hospital markets - an application to the German hospital sector’, *Health economics review*, vol. 4, no. 1, p. 28.

Hermeling, P. and Geraedts, M. (2013) ‘Kennen und nutzen arzte den strukturierten qualitatsbericht?’, *Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany))*, vol. 75, no. 3, pp. 155–159.

Hernandez-Boussard, T., Downey, J. R., McDonald, K. and Morton, J. M. (2012) ‘Relationship between patient safety and hospital surgical volume’, *Health Services Research*, vol. 47, no. 2, pp. 756–769.

Ho, V. and Hamilton, B. H. (2000) ‘Hospital mergers and acquisitions: Does market consolidation harm patients?’, *Journal of Health Economics*, vol. 19, no. 5, pp. 767–791.

IQTIG (2016) *Qualitätsreport 2015* [Online], Berlin, Institut für Qualitätssicherung und Transparenz im Gesundheitswesen. Available at https://www.iqtig.org/ergebnisse/qualitaetsreport/ (Accessed 17 October 2016).

Kessler, D. P. and Geppert, J. J. (2005) 'The Effects of Competition on Variation in the Quality and Cost of Medical Care', *Journal of Economics <html_ent glyph="@amp;" ascii="&"/> Management Strategy*, vol. 14, no. 3, pp. 575–589.

Kessler, D. P. and McClellan, M. B. (2000) 'Is Hospital Competition Socially Wasteful?', *The Quarterly Journal of Economics*, vol. 115, no. 2, pp. 577–615.

Klauber, J., Geraedts, M., Friedrich, J. and Wasem, J., eds. (2016) *Schwerpunkt: Ambulant im Krankenhaus*, Stuttgart, Schattauer.

Kumpunen, S., Trigg, L. and Rodrigues, R. (2014) *Public reporting in health and long-term care to facilitate provider choice,* World Health Organization, Policy Summary 13 [Online]. Available at http://www.euro.who.int/en/about-us/partners/observatory/publications/policy-briefs-and-summaries/public-reporting-in-health-and-long-term-care-to-facilitate-provider-choice (Accessed 28 October 2015).

Laine, C. R. (1995) 'The Herfindahl-Hirschman Index: A Concentration Measure Taking the Consumer's Point of View', *The Antitrust Bulletin*, vol. 40, no. 2, pp. 423–432.

Langner, D., Garling, M., Krzyzanowski, M., Bestman, B., Verheyen, F. and Meusch, A. (2016) *Qualität: Die Patientenperspektive,* WINEG, WINEG WISSEN.

Lindlbauer, I. and Schreyögg, J. (2014) 'The relationship between hospital specialization and hospital efficiency: do different measures of specialization lead to different results?', *Health care management science*, vol. 17, no. 4, pp. 365–378.

Maeda, J. L. K. and Lo Sasso, A. T. (2012) 'The relationship between hospital market competition, evidence-based performance measures, and mortality for chronic heart failure', *Inquiry : a journal of medical care organization, provision and financing*, vol. 49, no. 2, pp. 164–175.

Mukamel, D. B., Zwanziger, J. and Bamezai, A. (2002) 'Hospital competition, resource allocation and quality of care', *BMC Health Services Research*, vol. 2, no. 1, p. 3241.

Mukamel, D. B., Zwanziger, J. and Tomaszewski, K. J. (2001) 'HMO penetration, competition, and risk-adjusted hospital mortality', *Health Services Research*, vol. 36, 6 Pt 1, pp. 1019–1035.

Mutter, R. L., Wong, H. S. and Goldfarb, M. G. (2008) 'The effects of hospital competition on inpatient quality of care', *Inquiry : a journal of medical care organization, provision and financing*, vol. 45, no. 3, pp. 263–279.

Palangkaraya, A. and Yong, J. (2013) 'Effects of competition on hospital quality: an examination using hospital administrative data', *The European journal of health economics : HEPAC : health economics in prevention and care*, vol. 14, no. 3, pp. 415–429.

Pauly, M. V., McGuire, T. G. and Barros, P. P., eds. (2012) *Handbook of Health Economics*, Amsterdam, London, North Holland.

Propper, C., Burgess, S. and Gossage, D. (2008) 'Competition and Quality: Evidence from the NHS Internal Market 1991-9*', *The Economic Journal*, vol. 118, no. 525, pp. 138–170.

Propper, C., Burgess, S. and Green, K. (2004) 'Does competition between hospitals improve the quality of care?', *Journal of Public Economics*, vol. 88, 7-8, pp. 1247–1272.

Pross, C., Averdunk, L.-H., Stjepanovic, J., Busse, R. and Geissler, A. (2017) 'Health care public reporting utilization: User clusters, web trails, and usage barriers on Germany's public reporting portal Weisse Liste.de', *BMC Medical Informatics and Decision Making (in press))*.

Pross, C., Berger, E., Geissler, A., Martin, S. and Reinhard, B. (2017) ‘Stroke Units, Certification, and Stroke Outcomes in German Hospitals: A Fixed Effect Model for Stroke Mortality from 2006 - 2014’, *Health Economics (submitted)*.

Pross, C., Geissler, A. and Busse, R. (2017a) ‘Hospital Choice Matters: a mixed method analysis of outcome quality performance and variation in German hospitals from 2006-2014’, *Under review at Health Policy*.

Pross, C., Geissler, A. and Busse, R. (2017b) ‘Measuring, Reporting, and Rewarding Quality of Care in 5 Nations: 5 Policy Levers to Enhance Hospital Quality Accountability’, *The Milbank quarterly*, vol. 95, no. 1, pp. 136–183.

Pross, C., Strumann, C., Geissler, A., Herwartz, H. and Klein, N. ‘Quality and Efficiency in Hospital Service Provision: A Geoadditive Stochastic Frontier Analysis of Stroke Quality of Care in Germany’, *Journal of Health Economics (submitted)*, vol. 2017.

QSR Verfahren (2015) *Indicator handbook for treatment areas not covered in AOK Hospitalnavigator* [Online], Berlin, Wissenschaftliches Institut der AOK.

Ringelstein, E. B., Muller-Jensen, A., Nabavi, D. G., Grotemeyer, K.-H. and Busse, O. (2011) ‘Erweiterte Stroke-Unit’, *Der Nervenarzt*, vol. 82, no. 6, pp. 778–784.

Robertson, R. and Burge, P. (2011) ‘The impact of patient choice of provider on equity: Analysis of a patient survey’, *Journal of health services research & policy*, 16 Suppl 1, pp. 22–28.

Schmid, A. and Varkevisser, M. (2016) ‘Hospital merger control in Germany, the Netherlands and England: Experiences and challenges’, *Health policy (Amsterdam, Netherlands)*, vol. 120, no. 1, pp. 16–25.

Schneider, E. C. and Epstein, A. M. (1996) ‘Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists’, *The New England journal of medicine*, vol. 335, no. 4, pp. 251–256.

Schwartz, L. M., Woloshin, S. and Birkmeyer, J. D. (2005) ‘How do elderly patients decide where to go for major surgery? Telephone interview survey’, *BMJ (Clinical research ed.)*, vol. 331, no. 7520, p. 821.

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

Tay, A. (2003) ‘Assessing competition in hospital care markets: the importance of accounting for quality differentiation’, *The Rand journal of economics*, vol. 34, no. 4, pp. 786–814.

Varkevisser, M. and van der Geest, S. A. (2007) ‘Why do patients bypass the nearest hospital? An empirical analysis for orthopaedic care and neurosurgery in the Netherlands’, *The European journal of health economics : HEPAC : health economics in prevention and care*, vol. 8, no. 3, pp. 287–295.

Varkevisser, M., van der Geest, S. A. and Schut, F. T. (2010) ‘Assessing hospital competition when prices don't matter to patients: the use of time-elasticities’, *International journal of health care finance and economics*, vol. 10, no. 1, pp. 43–60.

Varkevisser, M., van der Geest, S. A. and Schut, F. T. (2012) ‘Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands’, *Journal of Health Economics*, vol. 31, no. 2, pp. 371–378.

Victoor, A., Delnoij, D. M. J., Friele, R. D. and Rademakers, J. J. D. J. M. (2012) ‘Determinants of patient choice of healthcare providers: a scoping review’, *BMC health services research*, vol. 12, p. 272.

Vogt, W. B. and Town, R. (2006) *How Has Hospital Competition Affected the Price and Quality of Hospital Care?,* Robert Wood Johnson FOundation, Research Synthesis Report 9 [Online]. Available at http://www.rwjf.org/en/library/research/2006/02/how-has-hospital-consolidation-affected-the-price-and-quality-of.html (Accessed 3 February 2017).

Zwanziger, J., Melnick, G. and Eyre, K. (1994) 'Hospitals and antitrust: defining markets, setting standards', *Journal of health politics, policy and law*, vol. 19, no. 2, pp. 423–477.

# Kapitel 7: Qualität und Ressourceneffizienz im Krankenhaus

Christoph Pross, Christoph Strumann, Alexander Geissler, Helmut Herwartz, Nadja Klein

Preprint, zur Begutachtung bei Management Science als: Pross C, Strumann C, Geissler A, Herwartz H, Klein N. Quality and Resource Efficiency in Hospital Service Provision: A Geoadditive Stochastic Frontier Analysis of Stroke Quality of Care in Germany

**Hintergrund**: Zunehmender Kostendruck im Krankenhaus führt zu einem Zielkonflikt zwischen möglichst optimaler Behandlungsqualität und Kostenreduktion. Daher sorgen sich Experten und Patienten um die Behandlungsqualität im Krankenhaus. Dennoch haben die meisten Studien zu technischer Krankenhauseffizienz die Qualität bisher ausgeklammert, weil umfangreiche und vergleichbare Qualitätsdaten nicht oder nur schwer verfügbar waren. Effizienzstudien mit Behandlungsqualität als Parameter haben widersprüchliche Ergebnisse geliefert. Zusätzlich werden räumliche Interaktionen in den Krankenhauseffizienzstudien bisher unzureichend betrachtet, obwohl die Krankenhauseffizienz abhängig von Interaktionen mit räumlichen Einflüssen wie Wettbewerb mit anderen Krankenhäusern, der ambulanten Versorgung und demografischen Faktoren ist.

**Ziele**: Auf Grundlage des stochastischen Effizienzgrenzansatzes und standardisierter Ergebnisindikatoren wird eine Methodik entwickelt, mit der die qualitätsadjustierte technische Effizienz und relevante Einflussfaktoren geschätzt werden können. Mit Hilfe dieser neuen Methodik untersucht die Arbeit Effizienz in der Schlaganfallversorgung in Bezug auf Behandlungsqualität und Ressourceneinsatz und evaluiert Struktur- und Raumfaktoren, die diese Effizienz beeinflussen. Auf Basis der empirischen Ergebnisse werden Empfehlungen zur Besserung der Effizienz und Qualität in der Schlaganfallversorgung ausgesprochen.

**Methodik**: Die Arbeit nutzt einen Paneldatensatz mit Strukturdaten auf Krankenhaus- und Fachabteilungsebene (G-BA), standardisierte Ergebnisindikatoren zu Mortalität und Wiedereinlieferung bei Schlaganfall (AOK), DSG Stroke Unit Zertifizierungsdaten und sozioökonomische Daten aus der INKAR Datenbank. Der verfügbare Datenzeitraum reicht von 2006 bis 2013. Die Datenelemente sind verknüpft über standardisierte Institutskennzeichen und Adressdaten der Krankenhäuser. Es wird eine Produktionsfunktion mit risikoadjustiertem Patientenvolumen als Output, Qualitätsfaktoren (realisierte Sterbefälle und Wiedereinlieferungen) und Ressourcen (Personaleinsatz von Arzt- und Pflegepersonal) als zu minimierende Inputs und Struktur- und Raumfaktoren als Determinanten der Ineffizienz geschätzt. Effizienzverbesserungsmaßnahmen werden durch Größe und Richtung der marginalen Effekte der Strukturfaktoren auf die Ineffizienz identifiziert.

**Ergebnisse**: Die Ergebnisse zeigen ein klassisches Trade-off zwischen Behandlungsqualität und Ressourceneinsatz, da Krankenhäuser mit hoher Behandlungsqualität gleichzeitig einen hohen Ressourceneinsatz vorweisen. Räumliche Faktoren haben einen signifikanten Einfluss auf die Schätzer. Mit einem durchschnittlichen Effizienzlevel von 0.73 gibt es ein hohes Effizienzsteigerungspotential, welches sowohl durch die Reduktion der Schlaganfallsterblichkeit als auch durch Ressourceneinsparungen erreicht werden kann. Insgesamt könnte bei Erreichung der Effizienzgrenze die Schlaganfallsterblichkeit um 2 630 Sterbefälle und Wiedereinlieferung um 2 951 Fälle reduziert werden. Wichtige effizienzsteigernde Faktoren sind der Aufbau einer SU, eine SU Zertifizierung, die regionale Zentralisierung der Behandlung und organisatorische Zentralisierung der Behandlung innerhalb eines Krankenhauses. Die Methodik kann zur Identifikation von Best Practice Krankenhäusern und für die qualitätsadjustierte Effizienzschätzung in anderen Leistungsbereichen eingesetzt werden.

# Quality and Resource Efficiency in Hospital Service Provision: A Geoadditive Stochastic Frontier Analysis of Stroke Quality of Care in Germany

Christoph Pross* Christoph Strumann† Alexander Geissler‡
Helmut Herwartz§ Nadja Klein¶

**Abstract**

We specify a Bayesian geoadditive SFA model to assess hospital performance along the dimensions of resources and quality of stroke care in German hospitals. With 1,100 annual observations and data from 2006 to 2013 and risk-adjusted patient volume as output, we introduce a production function that captures quality, resource inputs, hospital inefficiency determinants and spatial patterns of inefficiencies. We identify performance improvement mechanisms by considering marginal effects for the average hospital. Specialization and certification can substantially reduce mortality. Regional and hospital-level concentration can improve quality and resource efficiency. Finally, our results demonstrate a trade-off between quality improvement and resource or cost reduction and substantial regional variation in efficiency.

Keywords: Hospital quality; hospital efficiency; regional and spatial modelling; geoadditive stochastic frontier model; spatial correlation

*JEL: I11, I18, D24, C11, C14*

*Corresponding author. Department of Healthcare Management, Berlin University of Technology, Straße des 17. Juni 135, 10623 Berlin, christoph.pross@campus.tu-berlin.de, Work +49 03 314 28420, Fax +49 30 314 28433

†Institute for Entrepreneurship and Business Development, University of Lübeck, christoph.strumann@uni-luebeck.de

‡Department of Healthcare Management, Berlin University of Technology, a.geissler@tu-berlin.de

§Chair of Econometrics, Georg-August-University Göttingen, hherwartz@uni-goettingen.de

¶University of Melbourne, Melbourne Business School, 200 Leicester Street, Carlton VIC 3053, Australia

# 1 Introduction

Most OECD countries are characterized by a steady increase of hospital expenditures, which account for a substantial and increasing share of overall health care costs (Street et al., 2014; Martin et al., 2014; OECD, 2015; Levit et al., 2002). Reforms have been implemented to induce cost containment in the hospital sector, e.g., the introduction of activity-based hospital budgets (Piacenza et al., 2010). Increasing cost pressures have triggered concerns that hospitals face a trade-off between quantity of patients treated and service quality (Grieco and McDevitt, 2016; Wu and Shen, 2014; Pauly, 2011; Picone et al., 2003). As a consequence, the importance of hospital quality in policy-making and research is increasing. Hospitals have to optimize the relationship between cost and quality in order to provide high quality services using as few resources as possible.

Both cost and quality of health care have been shown to vary widely among countries, regions, and providers (Chung et al., 2015; Fonarow et al., 2011; Ghaferi et al., 2009; Fisher et al., 2003; Häkkinen et al., 2015; Peltola et al., 2015), suggesting an enormous potential for quality improvements and cost reductions. However, the importance of service quality has not been fully reflected in the hospital efficiency literature (McGlynn et al., 2008). Due to the lack of appropriate measures to incorporate quality in efficiency analysis (Hussey et al., 2009), and difficulties in hospital quality measurement itself (Lazar et al., 2013), models of hospital efficiency have rather focused on ad-hoc output measures such as patient and procedure volumes. The most important hospital production component, quality of care, has often been neglected (O'Neill et al., 2008; Hollingsworth, 2008).

In a seminal paper, Zuckerman et al. (1994) apply Stochastic Frontier Analysis (SFA) to estimate a multi-product cost function. To account for outcome quality differences and distinct patient populations, they include quality metrics as output variables, including a case-mix index (CMI) and 30-day risk-adjusted mortality. Their results emphasize the sensitivity of efficiency estimates with regard to the inclusion of quality measures. Mutter et al. (2008) highlight the non-trivial impact of including outcome quality measures or patient burdens of illness on efficiency estimates and hospital rankings. Based on cross-sectional regressions, Deily and McKay (2006) find higher cost inefficiency associated with a higher mortality rate in Florida hospitals, but in a separate study they fail to detect an association between cost inefficiency and care outcomes (McKay and Deily, 2008). The authors explain these contradictory results with regional differences in the relationship between hospital quality and cost inefficiency. In summary, most approaches to include quality into hospital efficiency estimations are restricted to stylized Data Envelopment Analysis (DEA) models and obtain inconclusive results (Nayar et al., 2013; Clement et al., 2008; Hollingsworth, 2008; Jha et al., 2009). As a consequence, little of the hospital efficiency research has been translated into policy or service delivery (Hollingsworth, 2012; Hollingsworth and Street, 2006).

The existing studies are often limited to cross-sectional data, hospital samples are restricted regionally or to a single ownership group. Quality indicators are not always risk-adjusted, and rarely include extended time-frames after hospital discharge. Likewise, outcome quality is often examined for the entire hospital or jointly for several medical conditions, which is problematic given condition-specific risk and resource profiles (Bradford et al., 2001). Furthermore, the existing literature has rarely accounted for spatial patterns of hospital performance, which have been detected, e.g., for Germany (Felder and Tauchmann, 2013; Herwartz and Strumann, 2012) or for England (Bloom et al., 2015; Gravelle et al., 2014). Neglecting common performance patterns of hospitals in the same or neighboring regions might induce systematic biases to inefficiency scores and estimated effects of their determinants, if quality varies not only between providers but across regions (Reistetter et al., 2015; Bechtold et al., 2015; Bech and Lauridsen, 2008).

In this study, we investigate the determinants of hospital performance and incorporate both performance dimensions: the efficient use of resources and the quality of care for a specific medical indication. Our aim is to identify the effects of care specialization, service line certification and treatment concentration on the efficiency of care provision for a hospital's risk-adjusted patient pool. As production output, we propose a risk-adjusted and medical condition-specific measure of hospital patient volume. This modification reduces the risk that unmeasured differences in hospital outputs affect estimated efficiencies. In contrast to Zuckerman et al. (1994), we do not further adjust the output term by quality, but rather include realized outcome quality as inputs. Holding the output term constant, hospitals are encouraged to minimize poor quality (mortality and readmissions) and staff levels (physicians and nurses) in order to treat patients efficiently. Thus, we propose a more comprehensive notion of (risk-adjusted) hospital performance that includes both quality of care and resource efficiency. Moreover, technical efficiency is captured at the medical condition and medical department as the relevant levels of analysis (Porter and Teisberg, 2006). At the same time, we take correlated spatial patterns among hospitals into account.

We focus our analysis on the acute care for stroke patients, which is characterized by high variation of quality of care (Wiedmann et al., 2015; Reistetter et al., 2014). Quality of care in stroke treatment is especially relevant for survival, disability, and long term cost of care. Stroke is a leading cause of disability, morbidity, and mortality in both developed and developing countries, with 6.7 million deaths globally each year due to stroke (WHO, 2014). Due to demographic changes and high costs of treatment, follow-up, and rehabilitation, stroke care poses a major challenge to all health care systems. At the same time, advances in stroke care promise improvements in care outcome quality (Stroke Unit Trialists' Collaboration, 2013; Donnan et al., 2008; Saposnik et al., 2008), and associated reductions in long term cost of care (Kalanithi et al., 2014). Hospitals adopt these care improvements to different degrees, which creates performance differentials between care providers.

For the empirical implementation, we join an information rich unbalanced panel data set for the time period 2006 to 2013 with a newly proposed, Bayesian geoadditive SFA model. The data set includes structural and quality metrics at the hospital and medical department level, with a focus on standardized stroke mortality (SMR) and readmissions, as well as the structural determinants of stroke outcomes, such as stroke unit capacities and certification. In comparison with previous DEA analyses, flexible SFA models have an important advantage. SFA addresses the concerns that DEA models quantify all departures from the best practice frontier as inefficiencies, including those caused by random events and measurement errors (Newhouse, 1994). This might lead to an overestimation of inefficiency (Rosko and Mutter, 2011). For the modelling of hospital efficiency including quality of care metrics, random events and measurement errors are of particular concern. The Bayesian SFA approach allows a distinguished analysis of individual heterogeneity and observable characteristics of inefficiency, while accounting for space dependent patterns of local performance by including geoadditive predictors of technical inefficiencies. Based on estimated efficiency scores we determine slack resources and estimate the saving potential in terms of quality improvements (saved deaths and readmissions) and staff reallocations if stroke care is provided efficiently.

# 2 Data and methods

## 2.1 Data

The empirical approach is based on an unbalanced panel data set, which combines four distinct data sources and comprises annual observations from 2006 until 2013 for more than 1,100 stroke treating German acute care hospitals. The panel data set is unique in time coverage, number of hospitals included, and scope of variables at the medical department, hospital and regional level.

As a first data source, we process structural hospital and medical department data (e.g., ownership, number of beds, disease and procedural case volumes) from publicly available hospital report cards.[1] Resource inputs (physician and nurse staff levels) are aggregated from the relevant medical departments and weighted by the share of stroke patients treated in each department. This provides a more exact measurement of the resource inputs at the relevant unit of analysis.

As a second data source, we integrate stroke outcome quality, such as the risk-adjusted 30-day mortality ratio and 30-day readmissions, as well as stroke case volumes, computed by the hospital report card initiative Quality Assurance with Routine Data (QSR). The QSR scheme is operated by the largest German sickness fund the AOK (Allgemeine Ortskrankenkasse), and employs administrative in- and outpatient data of AOK insured patients. The outcome quality measures cover three different stroke types: *Intracerebral hemorrhage* (ICD Code I61, 12% of all cases in Germany in 2014), *Ischemic stroke* (I63, 86%) and *Stroke not specified as hemorrhage or ischemic stroke* (I64, 2%). The AOK uses its administrative patient data to calculate risk-adjusted outcome rates for each hospital. The data enables detailed annual risk-adjustments by means of logistic regressions that include patient-specific risk-factors such as age, gender, and co-morbidities (diabetes, hypertension, etc.) (WIdO, 2015). The indicators cover a 30-day time period after hospital discharge, which is important for a more comprehensive quality of care assessment.

[1] The report cards are part of the mandatory external quality monitoring system operated by the executive authority in the German health care system, the Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA).

As a third data source, we integrate stroke unit information from the German Stroke Society (Deutsche Schlaganfall Gesellschaft, DSG). The data covers, inter alia, information on which hospitals operate DSG-certified stroke units and the time period of certification. The DSG stroke unit certificate stipulates an integrated, co-located, and high resource stroke care model that takes into account evidence-based care guidelines (DSG, 2015), which in many dimensions go beyond the average (non-certified) stroke unit infrastructure. As documented in Table 1, for 2013 the data set includes 219 hospitals with DSG-certified stroke units (with on average 619 stroke patients), 235 hospitals with a non DSG-certified stroke unit (340 stroke patients), and 937 hospitals without a stroke unit (64 stroke patients).

Lastly, we integrate regional district-level data on number of general practice physicians and economic indicators from the INKAR database hosted by the Federal Institute for Research on Building, Urban Affairs, and Spatial Development.

*insert Table 1 about here*

## 2.2 Stochastic frontier modelling of quality and efficiency

We operationalize the theoretical framework specified in Appendix A via an input-oriented stroke quality production function. For the output, we utilize the volume of stroke patients adjusted for their risk-profiles, which is computed by $RisAdjPatVol_{it} = ExpMor_{it}/\overline{ObsMorRate_t}$, where $ExpMor_{it}$ is the expected 30-day stroke mortality for hospital $i$ at time $t$ (calculated by the QSR initiative based on patient risk profiles and patient volumes (WIdO, 2015)) and $\overline{ObsMorRate_t}$ is the hospital average annual 30-day observed mortality rate at time $t$.

We specify two distinct types of input variables. First, the 30-day observed mortality ($ObsMor_{it}$) and the number of readmitted patients after 30 days ($ObsReadm_{it}$) serve as measures of quality of care. Second, to account for resource use, we include the number of (full-time equivalent) physicians ($Phys_{it}$) and nurses ($Nurs_{it}$) as input variables.

Relying on a Cobb-Douglas production function, we follow Battese (1997) to correct for zero input values, i.e., the optimal level of realized quality - no deaths or readmissions - by including two dummy variables ($D_{it}^{ObsMor}$) and ($D_{it}^{ObsReadm}$) in the production function. The variables take on the value of 1 when observed mortality or readmission are zero, respectively.[2] The empirical production function reads as

$$\begin{aligned}\log(RisAdjPatVol_{it}) &= \beta_0 + \beta_1 D_{it}^{ObsMor} + \beta_2 D_{it}^{ObsReadm} \\ &\quad + \beta_3 \log(ObsMor_{it}^*) + \beta_4 \log(ObsReadm_{it}^*) \\ &\quad + \beta_5 \log(Phys_{it}) + \beta_6 \log(Nurs_{it}) \\ &\quad + \lambda_t + v_{it} - u_{it}^* \exp\left\{\eta_{it}^{(u)}\right\},\end{aligned} \tag{1}$$

where $D_{it}^k = 1$ if $k_{it} = 0$, $D_{it}^k = 0$ if $k_{it} > 0$ and $k_{it}^* = \max(k_{it}, D_{it}^k)$ with $k \in \{ObsMor, ObsReadm\}$. Year effects are included as $\lambda_t$ for $t = \{2008, 2010, 2012, 2013\}$, with 2006 providing the reference year. We specify the inefficiency term ($\eta_{it}^{(u)}$) in (1) as

$$\begin{aligned}\eta_{it}^{(u)} &= f_{spat}(dist_i) + \tau_t + \delta_1 Spec_{it} \\ &\quad + \delta_2 MSStrDis_{it} + \delta_3 NumStrHospDis_{it} + \delta_4 \log(PatShaStroke_{it}) \\ &\quad + \delta_5 MedDepCon_{it} + \delta_6 SUCert_{it} + \delta_7 SUnonCert_{it} + \delta_8 GPsPerDis_{it} \\ &\quad + \delta_9 \log(HosBed_{it}) + \delta_{10} PrivHos_{it} + \delta_{11} NonProfHos_{it} + \delta_{12} Teach_{it} \\ &\quad + \delta_{13} UniHos_{it} + \delta_{14} ShaI61_{it} + \delta_{15} ShaI64_{it} + \delta_{16} DiagCon_{it},\end{aligned} \tag{2}$$

where the variables governing hospital inefficiency ($\eta_{it}^{(u)}$) allow a classification into six groups: specialization, certification, centralization, outpatient care, spatial structure and other control variables. Similar to the production function we also model year effects for the inefficiencies with $\tau_t$.

**Stroke specialization** We include a hospital-level stroke specialization measure in form of the stroke patient share of all inpatient cases treated annually in a hospital ($PatShaStroke_{it}$). With a relatively high stroke patient share, stroke treatment process and investment requirements have increased priority for the hospital and its staff. Likewise, staff is more experienced in and focused on stroke treatment.

More importantly, we consider the existence of a stroke unit as a specific structural measure of specialization. In stroke units, care is provided by an interdisciplinary team of experts, including neurologists, cardiologists, radiologists and neuro- and vas-

[2] This procedure allows us to preserve a substantial proportion of sample observations with an optimal level of quality, i.e., zero values for one or both of the variables $ObsMor_{it}$ and $ObsReadm_{it}$. Neglecting these observations might result in seriously biased estimators of the parameters of the production function (Battese, 1997).

cular surgeons. These experts are co-located in a specialized site with 24/7 availability and dedicated stroke diagnostic equipment such as CT scanners. These infrastructure and procedural conditions allow accurate and early diagnosis of type and extend of stroke as well as rapid and appropriate treatment of the different stroke types hemmorrhage (I61) and ischemic stroke (I63). The Stroke Unit Trialists' Collaboration (2013) found consistent evidence that more organized stroke care is associated with improved outcome quality, especially in seperate stroke units with dedicated staff and facilities. We capture stroke unit specialization through the variable $SUnonCert_{it}$, which marks hospitals that perform 10 or more complex stroke procedures (OPS codes 8-891 and 8-89b[3])) in a stroke unit annually Nimptsch and Mansky (2014), but have not received a special stroke service line certification.

**Certification** Hospital certificates confirm the compliance with general structural, process, and quality management standards or specific treatment guidelines set by medical specialty associations. To get approved, hospitals have to reach those higher-than-normal standards and fix systematic problems, which is supposed to result in better and more efficient inpatient care. However, empirical evaluations of the benefits of certification have shown inconclusive results with regard to improvements of patient safety and quality of care (Brubakk et al., 2015; Mumford et al., 2013; Nimptsch and Mansky, 2012; Lichtman et al., 2011). In our empirical model, the stroke unit variable $CertSU_{it}$ differentiates hospitals that have received a stroke unit certification from the German Stroke Society (DSG) from other hospitals.

**Centralization** Centralization of service provision has been demonstrated to improve outcome quality and reduce costs (Prabhakaran et al., 2013; Morris et al., 2014; Eastaugh, 2001; CIHI, 2005). In particular, centralization can lead to a more optimal regional care model based on economic considerations, patient needs, and quality of care. However, political resistance often hinder centralization of care (Turner et al., 2016). In this study, we distinguish inter- and intra-hospital centralization. The former is quantified at the district level via hospital market shares for stroke patients ($MSStrDis_{it}$), as well as the number of hospitals in one district treating stroke patients ($NumStrHospDis_{it}$). The latter is included in form of a Herfindahl Hirschman Index (HHI) of the medical department stroke patient shares within one

[3]OPS is the official German classification system for operations, procedures, and general medical activities, based on the International Classification of Procedures in Medicine from the World Health Organization.

hospital ($MedDepCon_{it}$).

**Outpatient care** General practitioners (GPs) and specialists in outpatient care monitor stroke risk factors, such as obesity, high blood pressure, and diabetes. GPs also provide the continuum of care after hospital discharge, in particular post-hospital observation and recovery from in-hospital conditions. GP density and continutiy of care can thus positively affect inpatient outcome quality if outpatient care is readily available in the hospital district (Barker et al., 2017; Bayliss et al., 2015; Cheng et al., 2010; Gill and Mainous, 1998).

**Underlying spatial structures** Several studies find that quality of care varies across regions (Reistetter et al., 2015; Bechtold et al., 2015). Moreover, as shown by Herwartz and Strumann (2012), spatial clusters characterize German hospital performance, and Gravelle et al. (2014) find similar effects for England. Ignoring this form of dependence cannot only affect estimation accuracy, but might also induce systematic biases to inefficiency scores and estimated effects of their determinants (Anselin, 1988). Moreover, if spatial clusters exist for both the dependent and explanatory variables, estimated relationships might appear stronger than they actually are (Bech and Lauridsen, 2008). Due to the difficult definition of a hospital market, we consider two rival spatial structures to ensure robustness of the empirical findings. On the one hand, we rely on the districts to define a hospitals' region and, on the other hand, on a broader classification based on the NUTS2[4] level. However, these artificial market boundaries cannot be assumed to reflect the true hospital markets, which might have an overlapping structure, e.g., due to patient flows. To capture both, correlated (overlapping) structures between the regions and region-specific effects, we separate the geoadditive effects in a structured part assigning a Markov random field prior, and an unstructured effect with identically distributed Gaussian prior (see also Section 2 for details). Moreover, through both effects, we are able to account for regional differences in health status and behaviour.

**Control variables** The empirical model also includes control variables common in the empirical hospital literature, such as hospital size measured through the number of hospital beds ($HosBed_{it}$), ownership type ($PrivHos_{it}$ and $NonProfHos_{it}$), uni-

[4]The Nomenclature of Units for Territorial Statistics (NUTS) is a geocode standard for referencing the subdivisions of countries for statistical purposes developed and maintained by the European Union.

versity hospital ($UniHos_{it}$), teaching status ($Teach_{it}$) and general, hospital-level specialization ($Spec_{it}$). While other studies include hospital beds as production inputs, we treat beds as an exogenous efficiency influence for two reasons. First, in Germany the number of hospital beds for each medical speciality is fixed through state-level hospital plans in the short- and mid-term (Geissler et al., 2011). Second, the total number of hospital beds, which aggregates number of beds in the different medical departments, does not characterize information pertinent to the actual medical departments that treat stroke patients. We control for overall hospital-level specialization, as specialized hospitals and medical departments can respond more rapidly and comprehensively to unanticipated and rare treatment complications. Specifically, we employ the Lindlbauer and Schreyögg (2014) specialization measure, which captures specialization by treatment area depending on volume thresholds, i.e., the average number of patients treated nationally (or 80% of the hospital's patients) in a specific diagnostic category ($Spec_{it}$). Furthermore, the particualar stroke types are associated with different resource requirements and distinctive survival and complication risks (Andersen et al., 2009; Lim and Cheon, 2015). To account for this, we include the composition of stroke patients for each hospital, which is approximated by the share of hemorrhage stroke relative to ischemic stroke ($sh - I61_{it}$ and $sh - I64_{it}$), as well as the distribution of the 3 types ($DiagCon_{it}$). Lastly, we control for the number of thrombectomies conducted in each hospital ($ShaTRHOMB_{it}$) to take into account high resource requirements, better treatment outcomes and more severe stroke cases associated with thrombectomy treatments (Bing et al., 2013; Minnerup et al., 2016).

*insert Table 2 about here*

# 3 Results

In this section, we discuss our results in five steps. First, in order to base the empirical analysis on the appropriate model, we evaluate various model specifications, which differ with regard to the inclusion of unobservable heterogeneity. Second, we display parameters describing the variation in the quality-expanded notion of hospital performance. The findings are checked for robustness against alternative performance definitions by excluding some input variables. Third, we analyze spatial patterns of inefficiencies in care provision when controlling for unobservable local conditions of hospital efficiency. Fourth, based on estimated efficiency scores and respective marginal effects of the underlying explanatory variables, we quantify slack resources and the associated potential for quality improvements or resource reductions as efficiency improvements. Lastly, we subject our results to a series of robustness checks.

In all models, we employ a total of 120,000 MCMC iterations. To reduce autocorrelations, we delete the first 20,000 iterations (burn-in) and store each 100th iterate (thinning). We check convergence of the chains graphically in terms of the sampling paths and autocorrelation plots. To ensure numerical stability, we center the output and input variables by subtracting their means. For the discussion of empirical results, we regard particular effects to be significant if a certain posterior credibility interval does not contain zero effects. The models were estimated in the open source software BayesX (Belitz et al., 2015), where an implementation of the geoadditive SFA models is provided.

## 3.1 Model selection

The diagnostic results of alternative models are displayed in Table 3. Starting with hospital individual effects, we extend the model by random effects varying on a regional level to take unobservable local conditions into account. Germany comprises 438 districts or 38 NUTS2 regions for which we allow spatial dependence patterns of technical inefficiencies. While each NUTS2 region hosts at least one hospital, the sample comprises hospitals which are located in 415 of the 438 districts. In consequence, for the specified spatial effects, the random Markov field (structured spatial pattern) obtains effects for all districts, while unstructured (random) spatial contributions are only determined for districts that host at least one hospital.

To assess model fit, we use the deviance information criterion (DIC, Spiegelhalter et al., 2002). The DIC of the model neglecting any spatial dependence (Model 1) exceeds the DIC of rival models. The specification of region-specific random effects to explain variations in inefficiency improves model accuracy substantially. This result indicates the presence of regional patterns in hospital performance. In general, the district level (Models 2 - 4) is more appropriate to account for these unobservable local conditions than the broader NUTS2 level (Models 5 - 7). The specification of both, a structural and a non-structural spatial random effect at the district level, achieves the best model fit (i.e., the lowest DIC). This is also confirmed in terms of the Watanabe-Akaike information criterion (WAIC, Watanabe, 2010). As a consequence, the following discussion of the results is mainly based on Model 2.

*insert Table 3 about here*

## 3.2 Estimated parameters

Table 4 provides estimated parameters of the best fitting quality-and-resource Model 2 along with the results for two models with alternative input specifications concentrating each only on one performance dimension, the quality-only Model 8 or the resource-only Model 9. On the left hand side of Table 4, we show the estimated coefficients for the production function and the model fit results for the three different models. On the right hand side, we show the estimated coefficient for the inefficiency terms. In order to analyze potential trade-offs between outcome quality and staff resources, we include the input variables as explanatory variables describing variations of inefficiencies when they are not incorporated in the production function. We also display effects (standardized coefficients) resulting from an estimation based on standardized output and respective input variables (zero mean, unit variance) to facilitate a comparison of their relative importance.

**Production function** Across all model specifications, the estimated output elasticities of the (non-zero) input variables are positive and significant. The coefficients of the dummy variables identifying zero input observations are significantly negative. These estimates can be interpreted as the mean number of risk-adjusted stroke patients, i.e., 0.7 ($= \exp(-0.329)$) or 0.8 ($= \exp(-0.186)$), if the observed mortality or readmission is zero, respectively. The negative year effects relative to the base year 2006 decrease over time until 2012 and, thus, represent a trend towards quality improvements, explained through medical technology progress and improved evidence-based guidelines for care provision. Standardized output elasticities of both resource variables (physician and nurse staff levels) are smaller than the elasticities of the quality indicators. The higher proportionality between the output term (risk-adjusted patient volume, based on the expected 30-day mortality) and the quality input terms (observed 30-day mortality and readmissions) might serve as an explanation here. The standardized effect of the physician staff level is stronger than that of nurse staff levels, since physicians are critical for care provision and outcomes by making the crucial diagnostic and treatment decisions. The output elasticity of observed mortality is about 44% higher than that of 30-day readmissions. Relative to readmissions, observed mortality has a higher proportionality with risk-adjusted patient volume, which is calculated based on each hospitals expected mortality.

**Effects on hospital performance** The estimated inefficiency effects for the different models offer some evidence for a trade-off between quality improvement and resource reduction. For instance, in Model 9, where quality is neglected in the production function, the negative effects of both quality indicators on the inefficiency term suggest that hospitals offering worse quality are able to provide stroke care relatively more resource efficient. Similarly, the effects of both dummy variables, $D^{ObsMor}$ and $D^{ObsReadm}$, identify high quality hospitals (no deaths or readmissions) as relatively less efficient. This result underlines the importance to expand the traditional approach of considering only costs or resources by integrating quality measures in hospital performance assessment. The estimated effect of physicians ($\log(Phys)$) differs substantially from that of nurses ($\log(Nurse)$) in Model 8, in which physicians (nurses) have a positive (negative) effect on quality performance. However, this interpretation might be misleading due to the high correlation of both input terms ($> 0.94$, unconditionally), which might explain the negative correlation of the coefficients.[5]

The two variables measuring centralization have both negative effects on the inefficiency. A comparison of the results across the models offers a decomposition of the efficiency improvements, since the strength of the effects varies across the model specifications. The impact of market ($MSStrDis_{it}$) and of the number of stroke care hospitals in one district ($NumStrHospDis$) is weaker when quality is not taken into account (Model 9). The concentration of stroke patients in medical departments ($MedDepCon_{it}$) has a stronger effect in comparison with models including quality (Model 2 and 8). Thus, the internal concentration (of stroke patients within the hospital across medical departments) affects mostly the resource use, while the external concentration (market share and number of stroke care hospitals) has stronger effects on the quality component of efficiency.

Specialization in stroke treatment, relative to other conditions ($PatShaStroke_{it}$),

[5]To clarify this issue, we have applied a Principal Component analysis to separate potentially different effects that are masked within both variables. We have found a negative effect on inefficiency of the component, which explains about 97% of the variation and highly correlates with both input variables ($> 0.99$). The second component, which correlates negatively (positively) with nurses (physicians), has also a negative effect. Although explaining less than 1% of the variation, the second component might govern the positive effect of the variable $\log(Nurse)$ on inefficiency.

and (non-certified) stroke-units ($SUnonCert_{it}$) have no effect on resource efficiency. However, if outcome quality of care is taken into account, both specialization measures have significantly negative effects on inefficiency. Similarly, the certification of stroke units ($CertSU_{it}$) enhances quality performance (Model 2, Model 8), but not resource efficiency (Model 9). Hospitals with a DSG-certified stroke unit exhibit less inefficiencies than hospitals with non-certified stroke units ($-1.167$ vs. $-0.592$).

Similar to other literature on hospital performance, private, for-profit hospitals face higher inefficiencies than public hospitals (Farsi and Filippini, 2008; Herwartz and Strumann, 2012). In our estimations, the latter result holds for private, non-profit hospitals only if quality is taken into account. The effect is negative if the performance is assessed only in terms of resources and contradicts with previous findings (Herwartz and Strumann, 2012; Herr, 2008). The financial straits for private, non-profit hospitals (Augurzky et al., 2016) might have provoked cost containment efforts, by increasing resource efficiency at the cost of quality performance.

The respective performance differential for university hospitals varies over the input specifications. It is only significant if quality is excluded. On average, university hospitals treat more severe cases with a higher CMI[6] compared with non-university acute care hospitals.[7] Including quality measures corrects for the more complex and resource intensive tertiary care provided by university hospitals.

The shares of the different stroke types, $sh - I61_{it}$ and $sh - I64_{it}$, obtain effects which are in line with expectations. The treatment of intracerebral hemorrhage (I61) is more complex and faces higher risks and mortality for the patients in comparison with the treatment of the reference group, i.e., ischemic stroke (I63). Furthermore, with a higher share of I61, the critical diagnostic and therapeutic distinction between I61 and I63 needs to be undertaken more frequently, which increases complexities. However, these complexities do not affect resource efficiency. Patients with an early stage or preliminary stroke, not classified as hemorrhage or ischemic stroke, (I64) are treated most efficiently, irrespective of whether quality is accounted for or not.

In contrast to our expectation, a higher concentration of stroke type diagnoses ($DiagCon_{it}$) enhances inefficiencies. Higher concentration on one particular diagno-

[6]The CMI reflects the average severity and associated resource requirement of cases treated at a certain hospital, with 1.0 representing the overall German average.

[7]Based on our dataset, German university hospitals had an average CMI of 1.51 in 2013. In contrast, non-university hospitals had an average CMI of 1.03.

sis might result in diseconomies of scale. Hospitals that face a high concentration of relatively low resource consuming stroke patients (e.g., with diagnosis I64) might also hold ready fixed resources for eventually more complex stroke admissions. With regard to the latest treatment innovations, a higher number of thrombectomy procedures increases resource efficiency, possibly due to a shorter length of stay (Ganesalingam et al., 2015). However, $ShaTRHOMB_{it}$ has no significant effect in the combined quality performance and resource efficiency model, which might be explained by thrombectomy being used in more severe cases, with severity not directly captured in the QSR risk-adjustment.

Similar to the centralization of stroke treatment at the district level, hospital beds have a negative effect on inefficiency only if quality is accounted for. Larger hospitals with more beds might benefit from improved economies of scale and more experience resulting in higher quality. Resource efficiency is not affected by hospital size.

*insert Table 4 about here*

**Spatial patterns of hospital performance** To visualize the spatial pattern of the regional effects entering the predictor of the inefficiency term ($\eta_{it}^{(u)}$) in equation (2), we display its centered posterior mean in Figure 1 for each German district. Non-colored, white areas indicate districts without hospitals. In contrast to the structural spatial effect (left hand side), the unstructured spatial effect (center) contributes to the composite spatial effect (right hand side) with markedly less variation. In total, favourable local conditions for the treatment of stroke patients are detected, in particular, for hospitals in Eastern and Southern Germany, while hospitals in North-Eastern and Western Germany are characterised by higher inefficiencies in the treatment of stroke patients.

A spatial decomposition of the inefficiencies into quality (deceased patients, Figure 2) and resources (staff levels, Figure 3) underlines the previous finding that an efficient use of resources is not always in line with high quality treatment of patients. While some regions in Western Germany show an efficient use of resources (e.g., within North Rhine-Westphalia or Rhineland-Palatinate), they lack behind other regions in Eastern Germany (e.g., within Brandenburg or Saxony) if quality performance is considered. However, some regions (e.g., within Baden-Wuerttemberg) perform well in terms of quality performance as well as resource efficiency.

*insert figures 1, 2 and 3 about here*

## 3.3 Estimated (in)efficiencies

**Slack resources** The overall posterior mean of technical efficiency is 0.730, indicating a sizeable potential for quality improvement and/or resource reallocation. To quantify such potentials for an average hospital,[8] we compute slack resources, i.e., resources in excess of those needed under full efficiency (for methodological details see (12) in Appendix B). We show the respective posterior distributions in Figure 4 for each input variable (in levels). If the average hospital boosts its performance to an efficient treatment of stroke patients, the mortality can be reduced by up to 6.26 (63.70%) deaths per year,[9] which is close to the maximum quality improvement for readmissions. Similarly, we can also compute the slack resources for each observation evaluated at the (MCMC) posterior mean, obtaining for 2013 total numbers of nationwide saved deaths and readmissions of 2630 (24.62%) and 2951 (30.78%).

*insert Figure 4 about here*

**Marginal effects** To highlight the economic relevance of those variables that describe the inefficiency, we provide their marginal effects on both technical efficiency and slack resources. If the latter are positive or negative, the effect quantifies how much of an input can be reduced or increased to achieve the same level of output (for methodological details see (9) and (13) in Appendix B). All effects shown in Table 5 are evaluated at the means of the underlying variables. Of particular interest are the marginal effects of treating patients in certified and non-certified stroke units, concentrating stroke treatment in fewer medical departments at the hospital level, and concentrating stroke treatment in fewer hospitals at the district level.

For example, hospitals with a DSG-certified stroke unit are characterized by a 0.026 (3.3%) higher technical efficiency in comparison with hospitals that treat stroke patients but do not operate a stroke unit. By closing their lacks of efficiency, these hospitals could increase their treatment quality by reducing stroke mortality by 0.927 or readmissions by 1.079 per year. When moving stroke treatment from hospitals

[8]The average hospital is defined by the sample means of all input and explanatory variables.

[9]In 2013 the means of observed mortality and readmission were 9.83 and 8.82, respectively.

without a stroke unit to hospitals with a non-certified stroke unit, mortality (readmissions) can be reduced by 0.401 (0.472).

*insert Table 5 about here*

## 3.4 Robustness Checks

To validate our results, we subject our analysis to a series of robustness checks, based on our best fit quality-and-resource Model 2. Results are presented in Table 6 for key variables.

As a more general formulation of our Cobb-Douglas production function, we estimate a translog model (10) and obtain very similar results for all relevant coefficients and model fit diagnostics. Likewise, rescaling the quality measures by adding 0.5 to each observation prior to log-transformation (Model 11) to handle optimal realized quality (zero mortality and/or readmissions) inputs achieves identical results as our first choice dummy variable approach.

To test alternative specifications of the non-certified, stroke unit variable $SUnonCert_{it}$, we half and double the required number of complex stroke procedures for stroke unit identification, based on our baseline of 10 procedures. As before, the results of these models (12-13) are qualitatively equivalent. Additionally, we add in Model 14 variables that differentiate between younger (1-2 years) and older certificates (3-4 years). The results are comparable, but older certificates show a weaker effect ($-1.070$) than younger certificates ($-1.252$). This is possibly due to the fact that organizational processes and structural provisions optimized before and during the certification process might loose some rigour over time.

To test for the robustness of the specified spatial structure, we include in Model 15 several district-level control variables such as annual mortality per 1,000 inhabitants, the average age of the population and unemployment rate in addition to the spatial structures. These variables do not improve model accuracy and their estimated effects are insignificant. This reflects the inclusion of regional health status due to the risk-adjusted patient population of each hospital and the structured and unstructured spatial structures. In addition, we examine two mixed-level models with NUTS2-level structured effects and district-level unstructured effects and vice versa (Models 16-17). Results are identical to our base-line Model 2.

Lastly, we safeguard the results of our certification variable for causal inference. Due to a potential self-selection of better performing hospitals pursuing certification, the existence of reverse causality cannot be excluded (see e.g. the findings of Dick et al., 2008). We adapt the procedure of Lindlbauer et al. (2016) to the SFA context and apply in Model 18 a combination of a matching approach and a Difference-In-Difference (DID) estimation to investigate the impact of certification on hospital performance. In a first step, we use propensity-score matching to ensure that any observed differences between certified and non-certified hospitals can be attributed to certification. To achieve a balance between both groups of hospitals in their baseline characteristics, we match each hospital becoming certified during the sample period with a non-certified hospital. In a second step, we estimate a DID specification of our best fitting Model 2. We include all observations in the estimation to achieve a comparable production frontier. However, we model specific effects on the inefficiency term for the pairs of matched hospitals. We also estimate specific effects for non-matched hospitals. The results confirm the previous finding that hospital inefficiency is reduced by certification; however, the effect in the DID Model 18 is smaller.

*insert Table 6 about here*

# 4 Discussion

In the following, we highlight four main implications. First, we discuss the improved efficiency for certified stroke units and the potential quality benefits of treating stroke patients in stroke units only, as well as the possible gains from within-hospital concentration of stroke care. Second, we examine the potential for resource reallocations due to efficiency improvements associated with specialization, certification and concentration. Third, we stress the potential benefits from regional concentration of stroke care. Fourth, we comment on the regional differences in hospital technical inefficiencies. In addition, we address some shortcomings of dataset and methodology.

**Stroke unit specialization and certification** In general, health policy makers, providers and payers could initiate substantial further consolidation and concentration of stroke care in certified centers of excellence. In 2013, 136,000 stroke patients (49%) were treated in certified stroke units while 80,000 patients (29%) were treated in non-certified stroke units and 60,000 (22%) stroke patients were treated in hospitals without a stroke unit (see Table 1). Certified and non-certified stroke units, on average, have a risk-adjusted stroke mortality of 0.9, while hospitals without a stroke unit have a risk-adjusted mortality of 1.0. A large share of stroke patients has not been treated in the best manner. Our results highlight the benefit of (further) care specialization and certification.

The twice-as-large efficiency gain for hospitals with certified stroke units can be explained by the increased expertise, better infrastructure and higher service level requirements for DSG-certified stroke units.[10] In preparation for the DSG audits, the stroke care team reviews and updates process plans, which can lead to improved outcome quality and efficiency after the certificate is granted.[11]

Exploiting the quality differences between the different stroke treating hospitals at the national level can result in substantial quality of care improvements and reductions of annual stroke deaths. If all stroke patients that are currently treated at hospitals without a stroke unit were treated at hospitals with a non-certified stroke unit, the average 30-day mortality for the 937 hospitals without a stroke unit could be reduced by 0.401 deaths, ceteris paribus. At the country level, this could result in annually 376 fewer stroke deaths after 30 days (i.e., ≈1% decrease of German national stroke mortality in 2013). Even more, if all patients from hospitals without a stroke unit were treated at hospitals with a certified stroke unit, national stroke mortality could be reduced by 868 deaths (≈2% reduction), ceteris paribus. Similar benefits could be achieved for readmissions. With regard to optimal resource allocation, not all currently stroke-treating hospitals shall set-up a (certified) stroke unit, but instead, centralizing stroke care in those hospitals that run a high-quality (i.e., certified) stroke unit could result in a more efficient and higher quality stroke care provision. To implement this concentration, a two stage policy approach appears feasible. In the first stage, regulators might require that stroke patients can only be treated in stroke units and, in a second stage, in DSG-certified stroke units only.

**Intra-hospital concentration** Next to specialization and certification, hospitals can also concentrate their acute and rehabilitative care within a specialized med-

[10]For example, the DSG requires in its stroke unit manual a minimum of 250 annual stroke patients treated and a minimum level of 1.5 FTE specialized nursing staff per stroke unit bed for a local stroke unit (Busse, 2008).

[11]As indicated by the certification timing effect identified in the robustness section confirms, the effect is particular strong in 1-2 year after certification.

ical department or stroke unit. Depending on size and specialization, hospitals often undertake care for one treatment area in several medical departments. Stroke care is a particularly good example as stroke diagnostic and therapeutic interventions can be performed by, e.g., the internal medicine, cardiology, and neurology departments. In 2013, only 381 hospitals provided stroke care in one medical department, 261 hospitals treated stroke patients in two medical departments, 128 hospitals treated stroke patients in 3 medical departments, and 339 hospitals treated stroke patients in at least 4 medical departments.

The negative and significant coefficient for hospital-level stroke care centralization ($MedDepCon_{it}$) in Model (2) (see Table 4) indicates that, in fact, within-hospital centralization can improve efficiency and care outcome quality. Following acute treatment in a stroke unit, patients can continue to be treated on this same specialized stroke unit or within a less intensive care setting rather than being relocated to different wards based on free hospital bed capacity. Considering the marginal effect for observed 30-day mortality of -0.142 and the average level of within-hospital, stroke care centralization of 0.83, mortality can be reduced by 0.023 deaths for the average hospital, if within-hospital stroke care centralization is maximally concentrated (at a HHI of 1). Extrapolating within-hospital effects to the national level, this could have a mortality reduction effect of 33 deaths per year.

**Regional concentration of stroke treatment** Regional concentration provides further potential for efficiency and outcome quality improvements. Currently, stroke care is undertaken by 1,391 hospitals (in 2013) in 415 districts in Germany. Case volumes range from 10 or less in the 110 hospitals with lowest volumes to 1,000 or more in the largest 27 stroke hospitals. There are almost 800 hospitals that provide care for less than 250 patients annually (DSG requirement for regional stroke unit), and 587 hospitals that provide care for less than 100 stroke patients. Likewise, in 2013 a single district hosts, on average, 7.5 hospitals that provide stroke care. There are 154 districts with 1 or 2 stroke hospitals, 173 districts with 3-5 hospitals, 60 districts with 6–10 hospitals, and 16 districts with more than 10 stroke hospitals.

To demonstrate the benefits from regional concentration, we highlight the inefficiency reducing effects of district hospital market share for stroke care. With a marginal effect of 0.013 on technical efficiency and an efficiency increasing elasticity of -0.453 for the average hospital, increases of the stroke hospital market share can substantially enhance efficiency and quality of care. When increasing the market share for half of the hospitals (696 hospitals) from an average of 30% to a share of 63% (mean plus one SD), annual national stroke mortality and readmissions could be reduced by 347 deaths and 408 cases per year, respectively, ceterius paribus.

Especially for emergency conditions, elapsed time until treatment is critical. In many countries, the recommended time window for stroke treatment after the onset of symptoms is 3.0 to 4.5 hours (Fassbender et al., 2013). In their review of stroke treatment guidelines and studies, Fassbender et al. (2013) emphasize the benefits of bypassing hospitals without a stroke unit in favor of treatment at more distant hospitals with a stroke unit. Similarly, several studies have shown the benefits of treating stroke patients in a more centralized model with specialized hospitals as opposed to a decentralized model with several smaller, non-specialized community hospitals (Lahr et al., 2012; Gladstone et al., 2009).

**Regional variation in technical efficiency** Focusing on resource efficiency only, some studies have found higher efficiencies for regions in Western and Northern Germany (Augursky and Schmitz, 2010). As a methodological distinction, however, we also include observed quality as an input and risk-adjust the output patient volume for specific patient risk-factors rather than considering CMIs. Furthermore, we evaluate efficiency for a specific treatment area and not for the overall hospital, which avoids the grouping of treatment areas. These methodological advances can possibly explain the following differences in the results. Including quality in the efficiency estimations, we find sizable inefficiencies for hospitals located in North Rhine-Westphalia (NRW), and, in particular, in the Rhein-Ruhr region. When excluding quality from the production function, hospitals located in NRW become relatively more efficient, which is in line with results in Augursky and Schmitz (2010). Interestingly, several regions (e.g., in Eastern Germany) with resource inefficiencies show a higher ability to produce better outcome quality.

**Trade-off between quality and resource inputs** Our results illustrate the potential to reduce or reallocate resource inputs while keeping stroke care quality measured in terms of observed mortality and readmissions constant. As the marginal effects on the staff slack resources (see Table 5) demonstrate, the efficiency enhancing effects of specialization, certification, regional concentration, and higher stroke patient share are substantial with regard to staff resources. When specialization and concentration increase, the productivity of both nurses and physicians increases, which can free up resources for reallocation, or as described above, might be invested in quality improvements. While a difficult choice, medical staff and hospital managers have to regularly decide on how to allocate limited resources more efficiently. Modelling such trade-offs can support resource allocation. Ideally, inefficiencies can be reduced through a combination of quality of care improvements and resource reallocations. This will achieve the highest reduction of the inefficiency gap since the marginal effect from reallocated units decreases the closer one gets to the efficiency frontier. Policy and management interventions should address both dimensions simultanously.

**Shortcomings** With regards to data and methodology employed, we consider some shortcomings. The QSR outcome quality indicators are based on AOK patient-level data and address the quality of care for AOK stroke patients treated in each hospital. Hence, outcome quality data is limited to AOK insured patients and quality data for patients insured with other public sickness funds and private health insurers is not included. However, the AOK is by far the largest health insurer in Germany, with an overall market share of 35% among publicly insured patients and a range of state-level market shares between 21% and 51%. This indicates that AOK patient data covers a large share of inpatient treatments and lets us assume representativeness of the AOK outcome data.[12] Furthermore, the morbidity-oriented risk structure compensation scheme within the public sickness fund system indicates that AOK patients, on average, are older and have a weaker health status compared with the rest of the statutory health insurance (SHI) population (Dietzel et al., 2015). While this could possibly lead to an overestimation of the potential inefficiencies, the majority of such a bias is controlled for when using risk-adjusted outcomes.

In addition, risk-adjustment based on administrative data, as opposed to clinical data, has limitations with regards to risk factors included (Quan et al., 2011; Pine et al., 2007). Most importantly, administrative data currently does not allow adjustments for disease severity levels and the degree of consciousness at admittance. More general, risk-adjustment only accounts for measurable and reported risk-factors, with many many important risk factors for adverse outcomes not measurable based

[12] A study comparing administrative patient data from the AOK sickness funds and hospital patient data from the HELIOS Group also confirmed comparability and representativeness of the AOK data (WIdO, 2007).

on current methods of administrative data-based risk-adjustment, e.g., preoperative function status, or not consistently reported (e.g., obesity) (AHRQ, 2014).

Furthermore, mortality and readmissions are only two, albeit important, aspects of stroke quality of care. Other factors such as health-related quality of life (Hopman and Verner, 2003) and patient-reported outcomes such as pain, selfcare, mobility, and health gain (Salinas et al., 2016), are also important, and can have even stronger impacts on hospital choice (Gutacker et al., 2016). As our results demonstrate, including multiple outcome quality aspects is important. When we include 30-day readmission only as an inefficiency determinant, an increase in 30-day readmissions reduces inefficiencies and indicates a trade-off between mortality and readmission as two distinct components of outcome quality (Krumholz, 2013).

# 5 Conclusions

In order to estimate a quality-expanded notion of hospital efficiency, we employ an innovative geoadditive SFA model with realized quality as an input and risk-adjusted patient volume as output. Past hospital efficiency research has mostly neglected quality of care, due to the difficulties in and drawbacks of risk-adjustment techniques and data availability. Research has also often neglected the potential for spatial patterns of inefficiency, which is especially problematic in local and regional settings with specific legislation and demand patterns such as hospital markets.

To address these shortcomings, we expand the notion of technical efficiency in three important directions. First, we include measures of realized poor quality as input variables for the production function. Second, we develop a standardized and risk-adjusted patient volume output measure based on the standardized mortality ratio. Third, we include spatial factors as inefficiency determinants. Altogether, these three methodological advances improve the applicability of technical efficiency estimations in a hospital service provision setting and can possibly increase the relevance and adoption of efficiency modelling in health policy making.

Our findings confirm that quality of care as well as spatial structures are highly important in shaping hospital technical efficiency. An overall hospital efficiency mean of 0.730 highlights the efficiency losses for the average hospitals, with regard to both quality and resource usage in stroke care. These inefficiencies have several important determinants. Specialization through a stroke unit improves outcome quality of care and resource efficiency. These effects are further strengthened if the stroke units are certified in accordance with highest standards (DSG certification). Within-hospital stroke care concentration at one medical department as well as regional stroke care concentration can also reduce inefficiencies and annual stroke mortality. Based on the marginal effects on slack resources, we highlight substantial quality of care improvement potentials at both the hospital and the national level.

In general, this work provides clear evidence for the importance of including quality of care in hospital efficiency modelling. The examined setting - a regulated market with a mix of private for-profit, private non-profit and public hospital ownership, with fixed prices and in the short-to-mid term fixed capacities - is common in many health care systems. Operationalizing the increasingly important and widespread concept of risk-adjusted, hospital-level mortality in a quality-enhanced SFA production function enables more comprehensive technical efficiency estimates in all countries, with available measures of standardized mortality ratios at a medical condition level (e.g., Germany, the US, the UK, and the Netherlands). Government agencies, health care regulators, and larger hospital chains (e.g., NHS Care Quality Commission in England, or the Hospital Corporation of America) can adopt the methodology to assess a quality-expanded notion of technical efficiency for their respective markets and hospitals. Best practice hospitals can be identified for different medical conditions providing optimal quality of care given their patient, resource and regional market constraints. Furthermore, medical care can be concentrated at these centers of excellence to increase both the quality of care provided to patients, and optimize resource utilization in health service provision. Additional research can expand the methodology to other treatment areas, such as other emergency conditions like acute myocardinal infarction or elective procedures such as hip and knee replacement.

**Acknowledgements**: We thank the Research Institute of the AOK SHI fund (WIdO) for granting us access to the QSR outcome data. Further, we thank Uwe Jensen, Carsten Schultz and Silvio Daidone for helpful comments and discussions.

# References

AHRQ (2014). Selecting quality and resource use measures: Part II. introduction to measures of quality (continued). Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/perfmeasguide/perfmeaspt2a.html.

Andersen, K. K., T. S. Olsen, C. Dehlendorff, and L. P. Kammersgaard (2009). Hemorrhagic and ischemic strokes compared: stroke severity, mortality, and risk factors. Stroke 40(6), 2068–2072.

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Springer.

Augursky, B. and H. Schmitz (2010). Effizienz von Krankenhäusern in Deutschland im Zeitvergleich: Endbericht.

Augurzky, B., S. Krolop, A. Pilny, C. M. Schmidt, and C. Wuckel (2016). Hospital Rating Report: With tail wind into the future? Heidelberg, Germany.

Barker, I., A. Steventon, and S. R. Deeny (2017). Association between continuity of care in general practice and hospital admissions for ambulatory care sensitive conditions: cross sectional study of routinely collected, person level data. British Medical Journal 356, j84.

Battese, G. E. (1997). A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. Journal of Agricultural Economics 48(1-3), 250–252.

Bayliss, E. A., J. L. Ellis, J. A. Shoup, C. Zeng, D. B. McQuillan, and J. F. Steiner (2015). Effect of continuity of care on hospital utilization for seniors with multiple medical conditions in an integrated health care system. Annals of Family Medicine 13(2), 123–129.

Bech, M. and J. Lauridsen (2008). Exploring the spatial pattern in hospital admissions. Health Policy 87(1), 50–62.

Bechtold, D., G. G. Salvatierra, E. Bulley, A. Cypro, and K. B. Daratha (2015). Geographic variation in treatment and outcomes among patients with AMI: Investigating urban-rural differences among hospitalized patients. The Journal of Rural Health, forthcoming.

Belitz, C., A. Brezger, N. Klein, T. Kneib, S. Lang, and N. Umlauf (2015). BayesX-Software for Bayesian inference in structured additive regression models. Repository version available from http://www.bayesx.org.

Bing, F., G. Jacquin, A. Poppe, D. Roy, J. Raymond, and A. Weill (2013). The cost of materials for intra-arterial thrombectomy. Interventional Neuroradiology 19(1), 83–86.

Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2015). The impact of competition on management quality: evidence from public hospitals. The Review of Economic Studies 82, 457–489.

Bradford, W. D., A. N. Kleit, M. A. Krousel-Wood, and R. N. Re (2001). Stochastic frontier estimation of cost models within the hospital. Review of Economics and Statistics 83(2), 302–309.

Brubakk, K., G. E. Vist, G. Bukholm, P. Barach, and O. Tjomsland (2015). A systematic review of hospital accreditation: the challenges of measuring complex intervention effects. BMC Health Services Research 15, 280.

Busse, O. (2008). Mitteilungen der Schlaganfallgesellschaft. Der Nervenarzt 79(12), 1465–1468.

Cheng, S.-H., C.-C. Chen, and Y.-F. Hou (2010). A longitudinal examination of continuity of care and avoidable hospitalization: evidence from a universal coverage health care system. Archives of Internal Medicine 170(18), 1671–1677.

Chung, S.-C., J. Sundström, C. P. Gale, S. James, J. Deanfield, L. Wallentin, A. Timmis, T. Jernberg, and H. Hemingway (2015). Comparison of hospital variation in acute myocardial infarction care and outcome between Sweden and United Kingdom: population based cohort study using nationwide clinical registries.

CIHI (2005). Health Care in Canada. Ottawa: Canadian Institute for Health information.

Clement, J. P., V. G. Valdmanis, G. J. Bazzoli, M. Zhao, and A. Chukmaitov (2008). Is more better? an analysis of hospital outcomes and efficiency with a DEA model of output congestion. Health Care Management Science 11(1), 67–77.

Deily, M. E. and N. L. McKay (2006). Cost inefficiency and mortality rates in florida hospitals. Health Economics 15(4), 419–431.

Dick, G. P., I. Heras, and M. Casadesús (2008). Shedding light on causation between ISO 9001 and improved business performance. International Journal of Operations & Production Management 28(7), 687–708.

Dietzel, J., K. Neumann, G. Glaeske, and W. Greiner (2015). Support research for the Morbi-RSA (part 1): Criteria, effects and alternatives. IGES Institut.

Donnan, P. T., D. T. Dorward, B. Mutch, and A. D. Morris (2008). Development and validation of a model for predicting emergency admissions over the next year (peony): A UK historical cohort study. Archives of Internal Medicine 168(13), 1416–1422.

DSG (2015). Certification Criteria 2015 - Regionale und Cross- regionale Stroke Units. Germany.

Eastaugh, S. R. (2001). Hospital costs and specialization: benefits of trimming product lines. Journal of Health Care Finance 28(1), 61–71.

Eilers, P. H. and B. D. Marx (1996). Flexible smoothing using B-splines and penalized likelihood. Statistical Science 11, 89–121.

Farsi, M. and M. Filippini (2008). Effects of ownership, subsidization and teaching activities on hospital costs in Switzerland. Health Economics 17(3), 335–350.

Fassbender, K., C. Balucani, S. Walter, S. R. Levine, A. Haass, and J. Grotta (2013). Streamlining of prehospital stroke management: The golden hour. The Lancet Neurology 12(6), 585–596.

Felder, S. and H. Tauchmann (2013). Federal state differentials in the efficiency of health production in germany: an artifact of spatial dependence. European Journal of Health Economics 14, 21–39.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder (2003). The implications of regional variations in medicare spending. part 1: the content, quality, and accessibility of care. Annals of internal medicine 138(4), 273–287.

Fonarow, G. C., E. E. Smith, M. J. Reeves, W. Pan, D. Olson, A. F. Hernandez, E. D. Peterson, and L. H. Schwamm (2011). Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. Stroke 42(1), 159–166.

Ganesalingam, J., E. Pizzo, S. Morris, T. Sunderland, D. Ames, and K. Lobotesis (2015). Cost-utility analysis of mechanical thrombectomy using stent retrievers in acute ischemic stroke. Stroke 46(9), 2591–2598.

Geissler, A., D. Scheller-Kreinsen, and R. Busse (2011). Germany: Understanding G-DRGs. In R. Busse, A. Geissler, W. Quentin, and M. Wiley (Eds.), Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals. Maidenhead: Open University Press and WHO Regional Office for Europe.

Ghaferi, A. A., J. D. Birkmeyer, and J. B. Dimick (2009). Variation in hospital mortality associated with inpatient surgery. The New England Journal of Medicine 361(14), 1368–1375.

Gill, J. M. and A. G. Mainous (1998). The role of provider continuity in preventing hospitalizations. Archives of Family Medicine 7(4), 352–357.

Gladstone, D. J., L. H. Rodan, D. J. Sahlas, L. Lee, B. J. Murray, J. E. Ween, J. R. Perry, J. Chenkin, L. J. Morrison, S. Beck, and S. E. Black (2009). A citywide prehospital protocol increases access to stroke thrombolysis in Toronto. Stroke 40(12), 3841–3844.

Gravelle, H., R. Santos, and L. Siciliania (2014). Does a hospital's quality depend on the quality of other hospitals? A spatial econometrics approach. Regional Science Urban Economics 49, 203–216.

Greene, W. H. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126, 269–303.

Grieco, P. L. and R. C. McDevitt (2016). Productivity and quality in health care: Evidence from the dialysis industry. The Review of Economic Studies, forthcoming.

Gutacker, N., L. Siciliani, G. Moscelli, and H. Gravelle (2016). Choice of hospital: Which type of quality matters? Journal of Health Economics, forthcoming.

Häkkinen, U., G. Rosenqvist, T. Iversen, C. Rehnberg, and T. T. Seppälä (2015). Outcome, use of resources and their relationship in the treatment of ami, stroke and hip fracture at european hospitals. Health Economics 24 Suppl 2, 116–139.

Herr, A. (2008). Cost and technical efficiency of German hospitals: does ownership matter? Health Economics 17(9), 1057–1071.

Herwartz, H. and C. Strumann (2012). On the effect of prospective payment on local hospital competition in germany. Health Care Managment Science 15, 48–62.

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

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. Health Economics 17, 1107–1128.

Hollingsworth, B. (2012). Revolution, evolution, or status quo? guidelines for efficiency measurement in health care. Journal of Productivity Analysis 37(1), 1–5.

Hollingsworth, B. and A. Street (2006). The market for efficiency analysis of health care organisations. Health Economics 15(10), 1055–1059.

Hopman, W. M. and J. Verner (2003). Quality of life during and after inpatient stroke rehabilitation. Stroke 34(3), 801–805.

Hussey, P. S., H. d. Vries, J. Romley, M. C. Wang, S. S. Chen, P. G. Shekelle, and E. A. McGlynn

(2009). A systematic review of health care efficiency measures. Health Services Research 44(3), 784–805.

Jha, A. K., E. J. Orav, A. Dobson, R. A. Book, and A. M. Epstein (2009). Measuring efficiency: the association of hospital costs and quality of care. Health Affairs 28(3), 897–906.

Jondrow, J., C. Lovell, I. Materov, and P. Schmidt (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19, 233–238.

Kalanithi, L., W. Tai, J. Conley, T. Platchek, D. Zulman, and A. Milstein (2014). Better health, less spending: delivery innovation for ischemic cerebrovascular disease. Stroke 45(10), 3105–3111.

Klein, N., H. Herwartz, and T. Kneib (2016). Geoadditive stochastic frontier analysis. Submitted, available from the authors upon request.

Koop, G., J. Osiewalski, and M. F. J. Steel (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. Journal of Econometrics 76, 77–105.

Krumholz, H. M. (2013). Post-hospital syndromean acquired, transient condition of generalized risk. New England Journal of Medicine 368(2), 100–102.

Kumbhakar, S. C. (1991). Estimation of technical inefficiency in panel data models with firm- and time-specific effects. Economics Letters 36, 43–48.

Kumbhakar, S. C., S. Ghosh, and J. T. McGuckin (1991). A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. Journal of Business and Economic Statistics 9, 279–286.

Lahr, M. M. H., G.-J. Luijckx, Vroomen, Patrick C A J, D.-J. van der Zee, and E. Buskens (2012). Proportion of patients treated with thrombolysis in a centralized versus a decentralized acute stroke care setting. Stroke 43(5), 1336–1340.

Lang, S. and A. Brezger (2004). Bayesian P-splines. Journal of Computational and Graphical Statistics 13, 183–212.

Lazar, E. J., P. Fleischut, and B. K. Regan (2013). Quality measurement in healthcare. Annual Review of Medicine 64, 485–496.

Levit, K., C. Smith, C. Cowan, H. Lazenby, and A. Martin (2002). Inflation spurs health spending in 2000. Health Affairs 21(1), 172–181.

Lichtman, J. H., S. B. Jones, Y. Wang, E. Watanabe, E. Leifheit-Limson, and L. B. Goldstein (2011). Outcomes after ischemic stroke for hospitals with and without joint commission-certified primary stroke centers. Neurology 76(23), 1976–1982.

Lim, J.-H. and S.-H. Cheon (2015). Analysis of variation in length of stay (los) after ischemic and hemorrhagic stroke using the Charlson Comorbidity Index (CCI). Journal of Physical Therapy Science 27(3), 799–803.

Lindlbauer, I. and J. Schreyögg (2014). The relationship between hospital specialization and hospital efficiency: do different measures of specialization lead to different results? Health Care Management Science 17(4), 365–378.

Lindlbauer, I., J. Schreyögg, and V. Winter (2016). Changes in technical efficiency after quality management certification: A DEA approach using difference-in-difference estimation with genetic matching in the hospital industry. European Journal of Operational Research 250(3), 1026–1036.

Martin, A. B., M. Hartman, L. Whittle, and A. Catlin (2014). National health spending in 2012: Rate of health spending growth remained low for the fourth consecutive year. Health Affairs 33(1), 67–77.

McGlynn, E. A., P. G. Shekelle, S. Chen, D. P. Goldman, J. A. Romley, P. S. Hussey, H. de Vries, M. C. Wang, M. J. Timmer, J. Carter, et al. (2008). Identifying, categorizing, and evaluating health care efficiency measures. RAND.

McKay, N. L. and M. E. Deily (2008). Cost inefficiency and hospital health outcomes. Health Economics 17(7), 833–848.

Minnerup, J., H. Wersching, A. Teuber, J. Wellmann, J. Eyding, R. Weber, G. Reimann, W. Weber, L. U. Krause, T. Kurth, and K. Berger (2016). Outcome after thrombectomy and intravenous thrombolysis in patients with acute ischemic stroke: A prospective observational study. Stroke 47(6), 1584–1592.

Morris, S., R. M. Hunter, Ramsay, Angus I G, R. Boaden, C. McKevitt, C. Perry, N. Pursani, A. G. Rudd, L. H. Schwamm, S. J. Turner, P. J. Tyrrell, Wolfe, Charles D A, and N. J. Fulop (2014). Impact of centralising acute stroke services in English metropolitan areas on mortality and length of hospital stay: difference-in-differences analysis. British Medical Journal 349, g4757.

Mumford, V., K. Forde, D. Greenfield, R. Hinchcliff, and J. Braithwaite (2013). Health services accreditation: what is the evidence that the benefits justify the costs? International Journal for Quality in Health Care 25(5), 606–620.

Mutter, R. L., M. D. Rosko, and H. S. Wong (2008). Measuring hospital inefficiency: the effects of controlling for quality and patient burden of illness. Health Services Research 43(6), 1992–2013.

Nayar, P., Y. A. Ozcan, F. Yu, and A. T. Nguyen (2013). Benchmarking urban acute care hospitals: efficiency and quality perspectives. Health Care Management Review 38(2), 137–145.

Newhouse, J. P. (1994). Frontier estimation: How useful a tool for health economics? Journal of Health Economics 13(3), 317 – 322.

Nimptsch, U. and T. Mansky (2012). Trends in acute inpatient stroke care in germany–an observational study using administrative hospital data from 2005-2010. Deutsches Ärzteblatt International 109(51-52), 885–892.

Nimptsch, U. and T. Mansky (2014). Stroke unit care and trends of in-hospital mortality for stroke in germany 2005-2010. International Journal of Stroke 9(3), 260–265.

OECD (2015). Focus on health spending. OECD Health Statistics. https://www.oecd.org/health/health-systems/Focus-Health-Spending-2015.pdf.

O'Neill, L., M. Rauner, K. Heidenberger, and M. Kraus (2008). A cross-national comparison and taxonomy of dea-based hospital efficiency studies. Socio-Economic Planning Sciences 42(3), 158–189.

Pauly, M. V. (2011). Analysis & commentary: The trade-off among quality, quantity, and cost: how to make it–if we must. Health Affairs 30(4), 574–580.

Peltola, M., T. T. Seppälä, A. Malmivaara, É. Belicza, D. Numerato, F. Goude, E. Fletcher, and R. Heijink (2015). Individual and regional-level factors contributing to variation in length of stay after cerebral infarction in six european countries. Health Economics 24 Suppl 2, 38–52.

Piacenza, M., G. Turati, and D. Vannoni (2010). Restructuring hospital industry to control public health care expenditure: The role of input substitutability. Economic Modelling 27(4), 881–890.

Picone, G. A., F. A. Sloan, S.-Y. Chou, and D. H. Taylor Jr (2003). Does higher hospital cost imply higher quality of care? Review of Economics and Statistics 85(1), 51–62.

Pine, M., H. S. Jordan, A. Elixhauser, D. E. Fry, D. C. Hoaglin, B. Jones, R. Meimban, D. Warner, and J. Gonzales (2007). Enhancement of claims data to improve risk adjustment of hospital mortality. JAMA 297(1), 71–76.

Porter, M. E. and E. O. Teisberg (2006). Redefining health care: Creating value-based competition on results. Boston, Mass.: Harvard Business School Press.

Prabhakaran, S., K. O'Neill, L. Stein-Spencer, J. Walter, and M. J. Alberts (2013). Prehospital triage to primary stroke centers and rate of stroke thrombolysis. JAMA Neurology 70(9), 1126–1132.

Quan, H., B. Li, C. M. Couris, K. Fushimi, P. Graham, P. Hider, J.-M. Januel, and V. Sundararajan (2011). Updating and validating the charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. American Journal of Epidemiology 173(6), 676–682.

Reistetter, T. A., A. M. Karmarkar, J. E. Graham, K. Eschbach, Y.-F. Kuo, C. V. Granger, J. Freeman, and K. J. Ottenbacher (2014). Regional variation in stroke rehabilitation outcomes. Archives of Physical Medicine and Rehabilitation 95(1), 29–38.

Reistetter, T. A., Y.-F. Kuo, A. M. Karmarkar, K. Eschbach, S. Teppala, J. L. Freeman, and K. J. Ottenbacher (2015). Geographic and facility variation in inpatient stroke rehabilitation: multilevel analysis of functional status. Archives of Physical Medicine and Rehabilitation 96(7), 1248–1254.

Rosko, M. D. and R. L. Mutter (2011). What have we learned from the application of stochastic frontier analysis to u.s. hospitals? Medical care research and review : MCRR 68(1 Suppl), 75S–100S.

Rue, H. and L. Held (2005). Gaussian Markov Random Fields. New York/Boca Raton: Chapman & Hall/CRC.

Salinas, J., S. M. Sprinkhuizen, T. Ackerson, J. Bernhardt, C. Davie, M. G. George, S. Gething, A. G. Kelly, P. Lindsay, L. Liu, S. C. O. Martins, L. Morgan, B. Norrving, G. M. Ribbers, F. L. Silver, E. E. Smith, L. S. Williams, and L. H. Schwamm (2016). An international standard set of patient-centered outcome measures after stroke. Stroke 47(1), 180–186.

Saposnik, G., J. Fang, M. O'Donnell, V. Hachinski, M. K. Kapral, and M. D. Hill (2008). Escalating levels of access to in-hospital care and stroke mortality. Stroke 39(9), 2522–2530.

Schmidt, P. and R. Sickles (1984). Production frontiers and panel data. Journal of Business and Economic Statistics 2, 367–374.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B (Statistical Methodology) 64, 583–639.

Street, A., N. Gutacker, C. Bojke, N. Devlin, and S. Daidone (2014). Variations in outcome and costs among nhs providers for common surgical procedures: econometric analyses of routinely collected data. Health Services and Delivery Research (2.1).

Stroke Unit Trialists' Collaboration (2013). Organised inpatient (stroke unit) care for stroke. The Cochrane Database of Systematic Reviews (9), CD000197.

Turner, S., A. Ramsay, C. Perry, R. Boaden, C. McKevitt, S. Morris, N. Pursani, A. Rudd, P. Tyrrell, C. Wolfe, and N. Fulop (2016). Lessons for major system change: centralization of stroke services in two metropolitan areas of England. Journal of Health Services Research & Policy.

Wang, H.-J. and C.-W. Ho (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. Journal of Econometrics 157, 286–296.

Wang, H.-J. and P. Schmidt (2002). One-step and two-step estimation of the effects of exogneous variables on technical efficiency levels. Journal of Productivity Analysis 18, 129–144.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 11, 3571–3594.

WHO (2014). The top 10 causes of death: Fact sheet n 310.

WIdO (2007). Qualitätssicherung der stationären Versorgung mit Routinedaten: (QSR) ; Abschlussbericht (1. Aufl. ed.). Bonn: Wiss. Inst. der AOK.

WIdO (2015). Regression weights: Reporting year 2015. germany.

Wiedmann, S., S. Hillmann, S. Abilleira, M. Dennis, P. Hermanek, M. Niewada, B. Norrving, K. Asplund, A. G. Rudd, C. D. Wolfe, and P. U. Heuschmann (2015). Variations in acute hospital stroke care and factors influencing adherence to quality indicators in 6 european audits. Stroke 46(2), 579–581.

Wu, V. Y. and Y.-C. Shen (2014). Long-term impact of medicare payment reductions on patient outcomes. Health Services Research 49(5), 1596–1615.

Zuckerman, S., J. Hadley, and L. Iezzoni (1994). Measuring hospital efficiency with frontier cost functions. Journal of Health Economics 13, 255–280.

# A The Bayesian geoadditive stochastic frontier model

In a stylized form, SFA models of technical efficiency consist of a formalized production technique linking an output and input variables, and a random deviation from the production technique that comprises idiosyncratic noise and technical inefficiencies. We rely on a flexible SFA model proposed by Klein et al. (2016) that builds upon earlier contributions, such as Schmidt and Sickles (1984), Kumbhakar (1991), Kumbhakar et al. (1991), Koop et al. (1997), Greene (2005), Wang and Schmidt (2002), Wang and Ho (2010). While SFA has become a standard parametric approach in modelling efficiency and efficiency-enhancing potentials in the provision of goods or services by economic entities such as firms, households, farms, and hospitals, the incorporation of spatial dependence in SFA models and other deterministic approaches to inefficiency modelling is in its infancy.[13]

[13] For a positioning of the geoadditive approach in the rich literature on SFA modelling we refer to Klein et al. (2016). They discuss model representation as well as model selection issues in detail, and illustrate model performance by means of simulated data.

The model addresses the important distinction between (time-invariant) determinants of inefficiency from patterns of unobserved heterogeneity by means of scaling of random inefficiency (Wang and Schmidt, 2002; Wang and Ho, 2010). Formally, the model reads as

$$\begin{aligned} y_{it} &= \eta_{it}^{(y)} + v_{it} - u_{it}, \\ &= \eta_{it}^{(y)} + v_{it} - u_{it}^{*} \exp\left(\eta_{it}^{(u)}\right), \end{aligned} \tag{3}$$

where $y_{it}$ denotes the (log) output of individual (i.e., hospital) $i$ at time $t$, $\eta_{it}^{(y)}$ and $\eta_{it}^{(u)}$ are predictors shaping the deterministic component of production and the distribution of technical inefficiencies, respectively. Concrete specifications of these predictors can be found in equations (1) and (2). The term $v_{it}$ is an idiosyncratic noise term and $u_{it}^{*}$ is a positive random variable invoking a wedge between actual output and its efficient counterpart. Joint with the formalization in (3), the following assumptions constitute the observation model that we use to evaluate quality of health care provision:

1. The predictor $\eta_{it}^{(y)}$ allows to relate the observable output $y_{it}$ to covariates, and individual and time effects. Metric covariates may contribute linearly to $\eta_{it}^{(y)}$ or in a more complex nonlinear manner which is formalized by means of so-called penalised splines (Eilers and Marx, 1996; Lang and Brezger, 2004). In addition, the predictor might process information inherent in categorical variables.[14]
2. To describe random deviations of actual output from the efficient frontier, technical inefficiency $u_{it}$ obeys the following structure

$$u_{it} = u_{it}^{*}\alpha_{it}, \qquad \alpha_{it} = \exp\left(\eta_{it}^{(u)}\right), \qquad u_{it}^{*} \sim \mathcal{N}^{+}\left(\mu_u^{*}, (\sigma_u^{*})^2\right), \tag{4}$$

   where $\mathcal{N}^{+}$ is short for the truncated normal distribution nesting the half normal distribution ($\mu_u^{*} = 0$) as a special case.[15] Scaling technical inefficiency as in (4) has been first proposed by Wang and Schmidt (2002). As an implication, $u_{it}$ exhibits a truncated normal distribution with parameters

$$\mu_{it} = \mu_u^{*}\alpha_{it} \text{ and } (\sigma_u^2)_{it} = (\sigma_u^{*})^2\,\alpha_{it}. \tag{5}$$

3. Economic entities providing goods or services are subject to measurable and unobservable local conditions. Accordingly, the predictor $\eta_{it}^{(u)}$ allows a decomposition as

$$\eta_{it}^{(u)} = \eta_{it}^{(o)} + f_{spat}(dist_i), \tag{6}$$

   where $\eta_{it}^{(o)}$ relates inefficiency to observable covariates, and $f_{spat}(dist_i)$ accounts for unobservable spatial conditions of (in)efficient output provision of the district hosting the $i$-th entity (denoted $dist_i$). As formalized in (6), the 'district' might be considered to provide a third data dimension

[14] The composition of the Bayesian regression design is specific for all these effect types. We refer to Klein et al. (2016) for a more detailed discussion of model representations for geoadditive SFA regression, as well as further methodological references therein to the broad framework of (Bayesian) structured additive regression.

[15] At the empirical side, small means $\mu^{*} > 0$ might be difficult to detect in light of the composite disturbance that describes deviations between actual and efficient output.

apart from the individual hospital and time dimension. Herwartz and Strumann (2014) provide strong evidence for the informational content of random regional effects for the modelling of hospital efficiency in Germany. The rightmost component in (6) allows a further distinction into spatially (i.e., positively) connected effects ($f_{struct}(dist_i)$) and purely random effects ($f_{unstruct}(dist_i)$), i.e.,

$$f_{spat}(dist_i) = f_{struct}(dist_i) + f_{unstruct}(dist_i). \tag{7}$$

The structured part ($f_{struct}$) refers to a specific neighbourhood structure being a (symmetric but possibly weighted) relation based on, e.g., common borders, cross-border care models, and cross-regional academic medical centers. In a Bayesian formulation, such effects can be modelled by means of Gaussian Markov random field priors (Rue and Held, 2005). Similar to temporal autoregressions, the systematic part of spatial dependence shrinks with the distance between districts so that the positive correlation among neighbouring districts captured by means of $f_{struct}(dist_i)$ generalizes similar structures as so-called spatially autoregressive models (Anselin, 1988). Altogether, the complete effect $f_{spat}$ in (7) subjects economic performance to geographical information, which likely invokes performance dependence among neighbouring districts, but also allows the occurrence of pure random effects to capture district-specific differences.

4. As an implication of the nonlinear model structure that results from the scaling of $u^*_{it}$, the predictor structure for $\eta^{(y)}_{it}$ and $\eta^{(u)}_{it}$ may overlap. For identification purposes $\eta^{(u)}_{it}$ does not contain an intercept, as this is represented by the scale of $u^*_{it}$.
5. The idiosyncratic noise $v_{it}$ and innovations to inefficiency $u^*_{it}$ are independent within and across all entities and time.

In its most flexible specification, the SFA approach in Klein et al. (2016) allows the use of predictors for all distribution parameters and hence also for $\sigma_v$, the specification of $\mu^*_u \neq 0$, nonlinear effects for metric variables or different prior structures for all involved unknown parameters. In this analysis, we focus on $f_{spat}$ offering a geoadditive resolution of quality of health care and add the following assumptions to facilitate model identification and interpretation: (i) The predictors for $\sigma^*_u$, $\sigma_v$ contain solely intercepts while $\mu^*_u$ is set to zero, and (ii) metric covariates are modelled linearly to capture effects of such variables by a single slope coefficient. On the implementation side, we (i) assume flat priors for all coefficients of linear effects, while (ii) for the unit and district-specific random effects we specify multivariate Gaussian priors with zero mean vectors and marginal variances that themselves are supposed to have inverse gamma prior distributions. To capture the structured spatial effect by Markov random field priors, (iii) the districts are considered as neighbours if they share common borders. The strength of the spatial structure is controlled by a variance parameter with inverse gamma hyperprior.

# B Marginal effects

In this section we, firstly, provide the formulas to obtain the marginal effects of the explanatory variables describing the inefficiency $u$ on the estimated technical efficiency. Secondly, the determination of slack resources are derived and, thirdly, the marginal effects on the slack resources are given.

## B.1 On technical efficiency

For notational convenience we suppress the indices $t$ and $i$. The technical efficiency ($TE$) of the $i$-th hospital at time $t$ is then obtained as (Jondrow et al., 1982)

$$TE \quad = \quad \mathbb{E}[\exp(-u|\epsilon)] = \frac{\exp\left(-\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2\right)\Phi\left(\frac{\tilde{\mu}}{\tilde{\sigma}} - \tilde{\sigma}\right)}{\Phi\left(\frac{\tilde{\mu}}{\tilde{\sigma}}\right)} = a \cdot b \cdot c, \tag{8}$$

where

$$\begin{aligned}
a &= \exp\left(-\tilde{\mu} + \frac{1}{2}\tilde{\sigma}^2\right) \\
b &= \Phi\left(\frac{\tilde{\mu}}{\tilde{\sigma}} - \tilde{\sigma}\right) \\
c &= \left[\Phi\left(\frac{\tilde{\mu}}{\tilde{\sigma}}\right)\right]^{-1} \\
\tilde{\mu} &= \frac{-\epsilon\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \\
\tilde{\sigma}^2 &= \frac{\sigma_u^2\sigma_v^2}{\sigma_u^2 + \sigma_v^2} \\
\sigma_u^2 &= \left(\sigma_u^*\right)^2 \alpha = \left(\sigma_u^*\right)^2 \exp\left(\eta^{(u)}\right) = \left(\sigma_u^*\right)^2 \exp\left(z'\beta^{(u)}\right) \\
\epsilon &= y - \eta^{(y)}.
\end{aligned}$$

The marginal effect ($ME_k$) of the $k$-th explanatory variable on $TE$ is given by

$$ME_k = \frac{\partial TE}{\partial z_k} \quad = \quad a\left(\frac{1}{2}d - e\right) \cdot b \cdot c \tag{9}$$

$$+\, a \cdot \phi\left(\frac{\tilde{\mu}}{\tilde{\sigma}} - \tilde{\sigma}\right)(g - f) \cdot c \tag{10}$$

$$-\, a \cdot b \cdot c^2 \cdot \phi\left(\frac{\tilde{\mu}}{\tilde{\sigma}}\right) g, \tag{11}$$

respectively, where

$$\begin{aligned}
d &= \beta_k^{(u)} \frac{\tilde{\sigma}^2}{\sigma_u^2} \\
e &= -\epsilon\beta_k^{(u)} \frac{\sigma_u^2\sigma_v^2}{\left(\sigma_u^2 + \sigma_v^2\right)^2} \\
f &= \frac{1}{2}\sigma_u\sigma_v\beta_k^{(u)}\left(\left(\sigma_u^2 + \sigma_v^2\right)^{-\frac{1}{2}} - \left(\sigma_u^2 + \sigma_v^2\right)^{-\frac{3}{2}}\sigma_u^2\right) \\
g &= \frac{\tilde{\sigma}e - \tilde{\mu}f}{\tilde{\sigma}^2}.
\end{aligned}$$

## B.2 On slack resources

To quantify the potential input reduction for a given (risk-adjusted) output level due to improvement of efficiency, we start by considering a simple SFA model in levels

$$\mathbb{E}[Y] = \prod_{k=1}^{K} X_k^{\beta_k} \cdot TE,$$

where $TE = \mathbb{E}[\exp(-u|\epsilon)] \in (0,1)$ (see (8)) denotes the technical efficiency. An increase of $TE$ offers a reduction of the inputs if the output level is not changed. Formally, an increase of $TE$ by $\Delta TE$ relates to a reduction of the deterministic part of the production function, $\prod_{k=1}^{K} X_k^{\beta_k}$, by $\frac{TE}{TE+\Delta TE}$, since

$$\mathbb{E}[Y] = \prod_{k=1}^{K} X_k^{\beta_k} \cdot TE = \prod_{k=1}^{K} X_k^{\beta_k} \frac{TE}{TE + \Delta TE} \cdot (TE + \Delta TE)\,.$$

Defining the reduced part involving the input variable of interest, say $X_k$, as

$$(X_k^*)^{\beta_k} = X_k^{\beta_k} \frac{TE}{TE + \Delta TE} = \frac{X_k^{\beta_k}}{1 + \frac{\Delta TE}{TE}}$$

and solving it for the new level of input $X^*$

$$X^* = \frac{X_k}{\left(1 + \frac{\Delta TE}{TE}\right)^{1/\beta_k}},$$

the reduction of $X_k$ due to an increase of $TE$ by $\Delta TE$ is given by

$$\Delta X_k = X_k^* - X_k = X_k \left( \frac{1}{\left(1 + \frac{\Delta TE}{TE}\right)^{1/\beta_k}} - 1 \right).$$

If for $\Delta TE$ the difference of being efficient, $\Delta TE = 1 - TE$, is considered, the slack resources are determined by

$$slack_k = X_k - X_k^* = X_k \left(1 - TE^{1/\beta_k}\right). \tag{12}$$

By inserting for $\Delta TE$ the marginal effect of an explanatory variable describing the inefficiency $u$ on the technical efficiency $TE$, $\Delta TE = ME$, we quantify the marginal effect of $z$ on the slack resources as

$$\Delta X_k = X_k^* - X_k = X_k \left( \frac{1}{\left(1 + \frac{ME}{TE}\right)^{1/\beta_k}} - 1 \right). \tag{13}$$

# C Tables

| | all | cert. SU | not cert. SU | no SU | others |
|---|---|---|---|---|---|
| number of hospitals | 1883 | 219 | 235 | 937 | 492 |
| *hospital size* | | | | | |
| beds | 270.6 | 658.5 | 435.0 | 226.0 | 104.4 |
| inpatient cases (in k) | 10.1 | 25.9 | 17.5 | 8.7 | 2.2 |
| *stroke patients* | | | | | |
| sum of stroke patients (in k) | 274.5 | 135.5 | 79.9 | 60.0 | - |
| share of stroke patients (in %) | 100 | 49.4 | 29.1 | 21.8 | - |
| ICD diagnoses[a] per hospital | 145.8 | 618.9 | 339.9 | 64.0 | - |
| OPS procedures[b] per hospital | 119.0 | 679.8 | 322.2 | 0.2 | - |
| QSR case volume per hospital | 70.7 | 204.9 | 117.1 | 22.7 | - |
| *QSR quality indicators*[c] | | | | | |
| standardized mortality ratio[d] | 1.0 | 0.9 | 0.9 | 1.0 | - |
| observed mortality rate (in %) | 14.8 | 12.3 | 12.8 | 16.0 | - |
| observed readmission rate (in %) | 13.6 | 11.8 | 12.2 | 14.3 | - |

Table 1: Means and sums of selected variables (2013)

[a]includes ICD codes I61, I63 and I64

[b]OPS procedure volume can be higher than ICD diagnoses volume because OPS procedures are also applied in case of other ICD diagnoses not included in this articles set of ICD stroke diagnoses (e.g., G45 = transient ischemic attack)

[c]Simple average across hospitals without weigthening based on patient volume

[d]The risk-adjusted mortality ratio relates the observed stroke mortality to the expected stroke mortality based on the number of patients treated and their risk-profiles.

| Variable | 2006 | 2008 | 2010 | 2012 | 2013 |
|---|---|---|---|---|---|
| number of hospitals | 1133 | 1121 | 1085 | 1096 | 1087 |
| *output variable* | | | | | |
| $\log(RisAdjPatVol)$ | 3.74 (1.17) | 3.62 (1.31) | 3.39 (1.47) | 3.28 (1.52) | 3.28 (1.59) |
| *input variables* | | | | | |
| $D^{ObsMor}$ | 0.08 (0.27) | 0.11 (0.31) | 0.15 (0.36) | 0.15 (0.36) | 0.16 (0.36) |
| $D^{ObsReadm}$ | 0.09 (0.29) | 0.11 (0.32) | 0.16 (0.36) | 0.18 (0.38) | 0.17 (0.37) |
| $\log(ObsMor^*)$ | 1.80 (1.11) | 1.74 (1.17) | 1.62 (1.25) | 1.56 (1.27) | 1.52 (1.29) |
| $\log(ObsReadm^*)$ | 1.55 (1.07) | 1.53 (1.17) | 1.47 (1.23) | 1.44 (1.24) | 1.46 (1.25) |
| $\log(Phys)$ | −0.58 (1.39) | −0.60 (1.56) | −0.67 (1.69) | −0.65 (1.71) | −0.62 (1.77) |
| $\log(Nurse)$ | 0.64 (1.39) | 0.54 (1.55) | 0.50 (1.73) | 0.49 (1.77) | 0.53 (1.82) |
| *explanatory variables* | | | | | |
| $Spec$ | 1.37 (0.75) | 1.34 (0.74) | 1.46 (0.83) | 1.36 (0.81) | 1.35 (0.81) |
| $MSStrDis$ | 0.31 (0.31) | 0.31 (0.32) | 0.31 (0.33) | 0.29 (0.32) | 0.30 (0.33) |
| $NumStrHospDis$ | 6.69 (8.13) | 6.93 (8.52) | 6.85 (8.57) | 7.52 (9.15) | 7.48 (9.21) |
| $\log(PatShaStroke)$ | −4.42 (1.16) | −4.50 (1.30) | −4.61 (1.44) | −4.62 (1.53) | −4.66 (1.51) |
| $\log(HosBed)$ | 5.52 (0.77) | 5.51 (0.78) | 5.51 (0.80) | 5.50 (0.81) | 5.51 (0.81) |
| $PrivHos$ | 0.17 (0.38) | 0.18 (0.38) | 0.19 (0.39) | 0.21 (0.41) | 0.21 (0.41) |
| $NonProfHos$ | 0.44 (0.50) | 0.44 (0.50) | 0.45 (0.50) | 0.43 (0.50) | 0.43 (0.50) |
| $Teach$ | 0.39 (0.49) | 0.43 (0.49) | 0.47 (0.50) | 0.50 (0.50) | 0.52 (0.50) |
| $UniHos$ | 0.03 (0.18) | 0.03 (0.18) | 0.03 (0.18) | 0.03 (0.18) | 0.03 (0.18) |
| $MedDepCon$ | 0.87 (0.19) | 0.87 (0.19) | 0.85 (0.19) | 0.84 (0.20) | 0.83 (0.20) |
| $ShaI61$ | 0.09 (0.10) | 0.10 (0.10) | 0.11 (0.11) | 0.11 (0.11) | 0.11 (0.10) |
| $ShaI64$ | 0.18 (0.22) | 0.11 (0.18) | 0.10 (0.18) | 0.09 (0.18) | 0.09 (0.17) |
| $DiagCon$ | 0.69 (0.18) | 0.73 (0.17) | 0.73 (0.17) | 0.73 (0.16) | 0.73 (0.16) |
| $SUCert$ | 0.08 (0.26) | 0.07 (0.25) | 0.11 (0.31) | 0.10 (0.30) | 0.11 (0.32) |
| $SUnonCert$ | 0.14 (0.35) | 0.20 (0.40) | 0.25 (0.43) | 0.24 (0.42) | 0.24 (0.43) |
| $\log(GPsPerDis)$ | 3.93 (0.16) | 3.90 (0.16) | 3.88 (0.16) | 3.86 (0.16) | 3.84 (0.17) |

Table 2: Descriptive statistics: mean and standard deviation (in parentheses)

| | | districts | | | NUTS2 | | |
|---|---|---|---|---|---|---|---|
| | no spatial | struct & unstruct | struct | unstruct | struct & unstruct | struct | unstruct |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *DIC* | 2817.6 | **2581.7** | 2591.0 | 2595.7 | 2766.7 | 2768.0 | 2766.2 |
| *WAIC* | 2975.8 | **2742.3** | 2750.8 | 2750.9 | 2933.2 | 2935.2 | 2932.3 |

based on 5522 observations (i.e. 1294 hospitals in 415 district and 38 NUTS2 regions)

Table 3: Model selection criteria for distinct specifications of spatial structures

| | ***Production function*** $(\eta_{it}^{(y)})$ | | | ***Effects on inefficiency*** $(\eta_{it}^{(u)})$ | | |
|---|---|---|---|---|---|---|
| | qual & res | qual only | res only | qual & res | qual only | res only |
| Variables | (2) | (8) | (9) | (2) | (8) | (9) |
| $const$ | 0.470*** | 0.490*** | 0.609*** | | | |
| $D^{ObsMor}$ | −0.329*** | −0.306*** | | | | 0.319*** |
| $D^{ObsReadm}$ | −0.186*** | −0.180*** | | | | 0.246*** |
| $\log(ObsMor^*)$ | 0.397*** | 0.420*** | | | | −0.517*** |
| *stand. coeff.* | *0.339* | *0.358* | | | | |
| $\log(ObsReadm^*)$ | 0.282*** | 0.314*** | | | | −0.287*** |
| *stand. coeff.* | *0.236* | *0.262* | | | | |
| $\log(Phys)$ | 0.056*** | | 0.264*** | | −0.235*** | |
| *stand. coeff.* | *0.063* | | *0.318* | | | |
| $\log(Nurse)$ | 0.015* | | 0.073*** | | 0.129*** | |
| *stand. coeff.* | *0.016* | | *0.077* | | | |
| 2008 | −0.025** | −0.025** | −0.053*** | 0.064 | 0.061 | −0.076 |
| 2010 | −0.085*** | −0.085*** | −0.144*** | 0.235*** | 0.217*** | −0.063 |
| 2012 | −0.107*** | −0.106*** | −0.175*** | 0.253*** | 0.241*** | 0.031 |
| 2013 | −0.056*** | −0.052*** | −0.116*** | 0.362*** | 0.352*** | 0.146** |
| $MSStrDis$ | | | | −0.829*** | −0.743*** | −0.136 |
| $MedDepCon$ | | | | −0.228** | −0.148 | −0.909*** |
| $NumStrHospDis$ | | | | 0.009** | 0.008** | −0.003 |
| $\log(PatShaStroke)$ | | | | −0.252*** | −0.200*** | 0.003 |
| $SUnonCert$ | | | | −0.592*** | −0.584*** | −0.020 |
| $SUCert$ | | | | −1.167*** | −1.295*** | −0.056 |
| $\log(GPsPerDis)$ | | | | −0.004 | −0.019 | 0.129 |
| $Spec$ | | | | −0.022 | −0.001 | 0.156*** |
| $\log(HosBed)$ | | | | −0.555*** | −0.482*** | −0.036 |
| $PrivHos$ | | | | 0.310*** | 0.251*** | 0.090 |
| $NonProfHos$ | | | | 0.158*** | 0.133*** | −0.107** |
| $Teach$ | | | | −0.069* | −0.040 | −0.038 |
| $UniHos$ | | | | 0.175 | 0.137 | 0.455*** |
| $ShaTRHOMB$ | | | | 0.052 | 0.185 | −1.144** |
| $ShaI61$ | | | | 0.619*** | 0.639*** | −0.099 |
| $ShaI64$ | | | | −0.299*** | −0.303*** | −0.258*** |
| $DiagCon$ | | | | 0.394*** | 0.318*** | 0.429*** |
| $\sigma_v^2$ | 0.142 | 0.148 | 0.141 | | | |
| $\sigma_u^2$ | 2.301 | 1.982 | 0.439 | | | |
| $\gamma = \sigma_u^2/(\sigma_v^2+\sigma_u^2)$ | 0.9419 | 0.9305 | 0.7569 | | | |
| $TE$ (in %) | 73.0 | 72.3 | 68.2 | | | |
| $DIC$ | 2581.7 | 2856.3 | 3817.2 | | | |
| $WAIC$ | 2742.3 | 2994.9 | 3956.7 | | | |

Significance levels: *** 1%; ** 5%; * 10%; based on 5522 observations

Table 4: Estimated linear effects

| Variables | Marginal effect on $TE$ (×100) | Marginal effect on slack resources $mort30d$ | $readm30d$ | $phys$ | $nurses$ |
|---|---|---|---|---|---|
| $Spec$ | 0.027 | −0.010 | −0.012 | −0.016 | −0.204 |
| $MSStrDis$ | 1.253*** | −0.453*** | −0.533*** | −0.566*** | −5.361* |
| $NumStrHospDis$ | −0.015** | 0.005** | 0.006** | 0.008** | 0.110 |
| log($PatShaStroke$) | 0.400*** | −0.148*** | −0.175*** | −0.202*** | −2.531* |
| log($HosBed$) | 0.888*** | −0.324*** | −0.382*** | −0.420*** | −4.455* |
| $PrivHos$ | −0.449*** | 0.170*** | 0.202*** | 0.255*** | 4.267* |
| $NonProfHos$ | −0.258*** | 0.097*** | 0.115*** | 0.145*** | 2.227* |
| $Teach$ | 0.092 | −0.034 | −0.041 | −0.049 | −0.664 |
| $UniHos$ | −0.235 | 0.089 | 0.106 | 0.143 | 2.157 |
| $MedDepCon$ | 0.319** | −0.118** | −0.140** | −0.166** | −2.058 |
| $ShaTRHOMB$ | −0.016 | 0.011 | 0.016 | −0.004 | 0.161 |
| $ShaI61$ | −1.000*** | 0.384*** | 0.460*** | 0.624*** | 13.159* |
| $ShaI64$ | 0.482*** | −0.178*** | −0.210*** | −0.241*** | −2.862* |
| $DiagCon$ | −0.627*** | 0.239*** | 0.285*** | 0.368*** | 6.696* |
| $SUCert$ | 2.660*** | −0.927*** | −1.079*** | −1.012*** | −7.016* |
| $SUnonCert$ | 1.105*** | −0.401*** | −0.472*** | −0.508*** | −5.027* |
| log($GPsPerDis$) | 0.108 | −0.040 | −0.047 | −0.064 | −0.841 |
| 2008 | −0.087 | 0.033 | 0.039 | 0.047 | 0.668 |
| 2010 | −0.314*** | 0.118*** | 0.141*** | 0.174*** | 2.779* |
| 2012 | −0.357*** | 0.135*** | 0.160*** | 0.200*** | 3.294* |
| 2013 | −0.499*** | 0.189*** | 0.226*** | 0.288*** | 5.014* |

Significance levels: *** 1%; ** 5%; * 10%

Table 5: Marginal effects on efficiency and slack resources (Model 2)

| | base-line | Translog | rescaling | $SUnonCert$ | | $SUCert$ | health | $f_{str}(dist_i)$& | $f_{str}(NUTS2_i)$& | matching & |
|---|---|---|---|---|---|---|---|---|---|---|
| | model | prod-fun | zero-inputs | (> 5) | (> 20) | age | variables | $f_{unstr}(NUTS2_i)$ | $f_{unstr}(dist_i)$ | DID |
| selected Variables | (2) | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) |
| $MSStrDis$ | $-0.829^{***}$ | $-0.892^{***}$ | $-0.843^{***}$ | $-0.816^{***}$ | $-0.803^{***}$ | $-0.812^{***}$ | $-0.799^{***}$ | $-0.845^{***}$ | $-0.839^{***}$ | |
| $MedDepCon$ | $-0.228^{**}$ | $-0.228^{**}$ | $-0.212^{**}$ | $-0.213^{**}$ | $-0.211^{**}$ | $-0.202^{**}$ | $-0.205^{**}$ | $-0.180^{**}$ | $-0.216^{**}$ | |
| $NumStrHospDis$ | $0.009^{**}$ | $0.009^{*}$ | $0.008^{*}$ | $0.009^{**}$ | $0.009^{**}$ | $0.009^{**}$ | $0.008^{*}$ | 0.004 | $0.011^{**}$ | |
| $\log(PatShaStroke)$ | $-0.252^{***}$ | $-0.207^{***}$ | $-0.245^{***}$ | $-0.258^{***}$ | $-0.256^{***}$ | $-0.257^{***}$ | $-0.258^{***}$ | $-0.253^{***}$ | $-0.262^{***}$ | |
| $SUnonCert$ (> 10) | $-0.592^{***}$ | $-0.669^{***}$ | $-0.603^{***}$ | | | $-0.587^{***}$ | $-0.587^{***}$ | $-0.567^{***}$ | $-0.592^{***}$ | |
| $SUnonCert$ (> 5) | | | | $-0.489^{***}$ | | | | | | |
| $SUnonCert$ (> 20) | | | | | $-0.665^{***}$ | | | | | |
| $D^m \times SUnonCert$ (> 10) | | | | | | | | | | $-1.370^{***}$ |
| $SUCert$ | $-1.167^{***}$ | $-1.265^{***}$ | $-1.182^{***}$ | $-1.098^{***}$ | $-1.197^{***}$ | | $-1.163^{***}$ | $-1.108^{***}$ | $-1.194^{***}$ | |
| $SUCert$ (1-2y) | | | | | | $-1.252^{***}$ | | | | |
| $SUCert$ (3-4y) | | | | | | $-1.070^{***}$ | | | | |
| $D^m \times SUCert$ (DID) | | | | | | | | | | $-0.580^{**}$ |
| $(1 - D^m) \times SUCert$ | | | | | | | | | | $-2.259^{***}$ |
| $\log(GPsPerDis)$ | −0.004 | −0.007 | −0.040 | −0.058 | −0.072 | −0.060 | −0.037 | −0.135 | −0.125 | |
| $\log(MortalityDis)$ | | | | | | | −0.069 | | | |
| $\log(PopageDis)$ | | | | | | | −0.065 | | | |
| $\log(UnemplrateDis)$ | | | | | | | 0.087 | | | |
| $\log(GDPpCapDis)$ | | | | | | | 0.010 | | | |

Significance levels: *** 1%; ** 5%; * 10%; $D^m$ is a dummy variable: 1 for observations of matched hospitals and 0 else

Table 6: Robustness-Checks

# D Figures

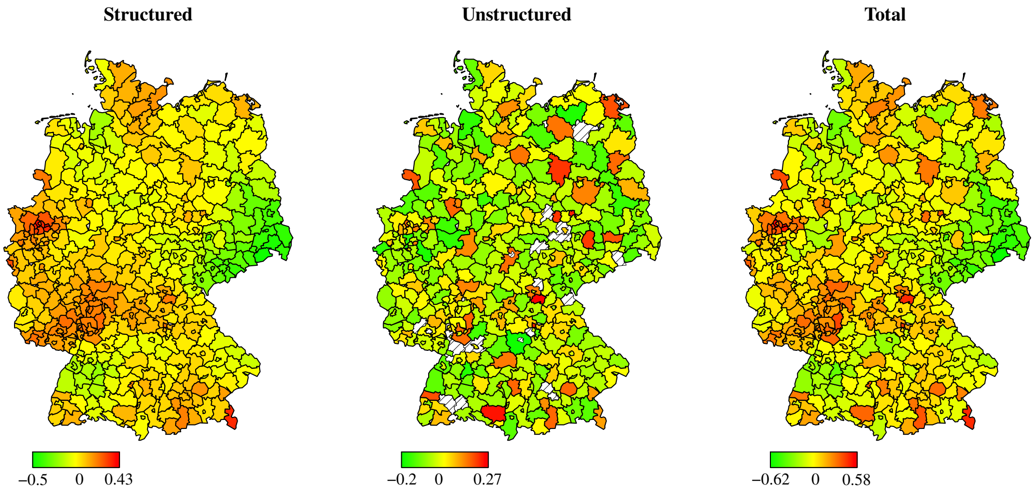



Figure 1: Centered estimated regional random effects of Model 2: quality & resources

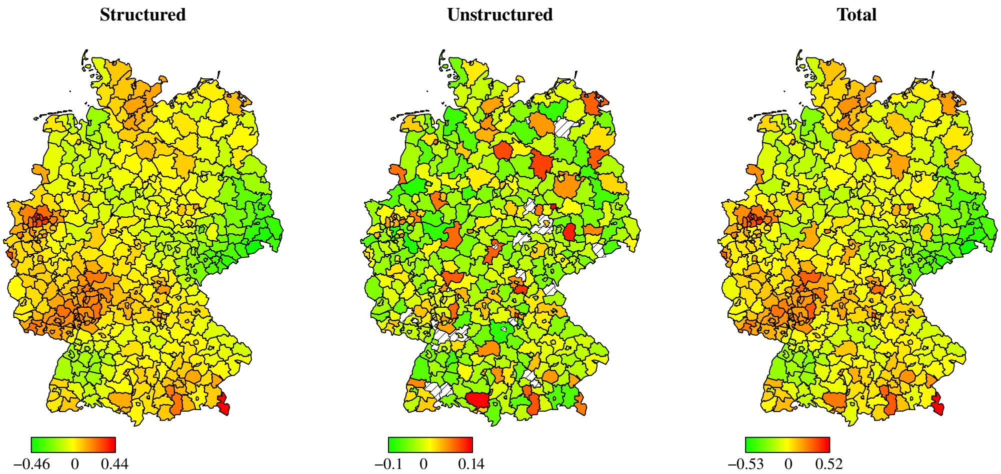



Figure 2: Centered estimated regional random effects of Model 8: quality only

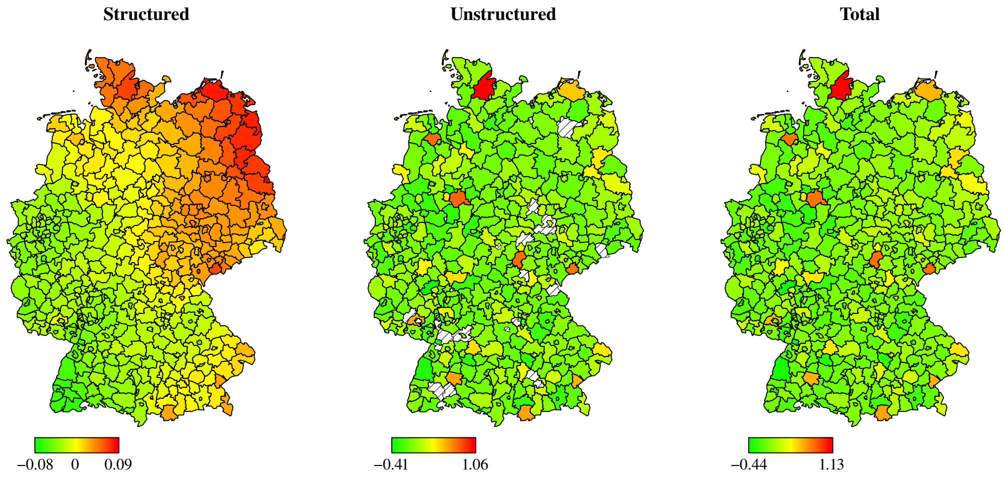



Figure 3: Centered estimated regional random effects of Model 9: resources only

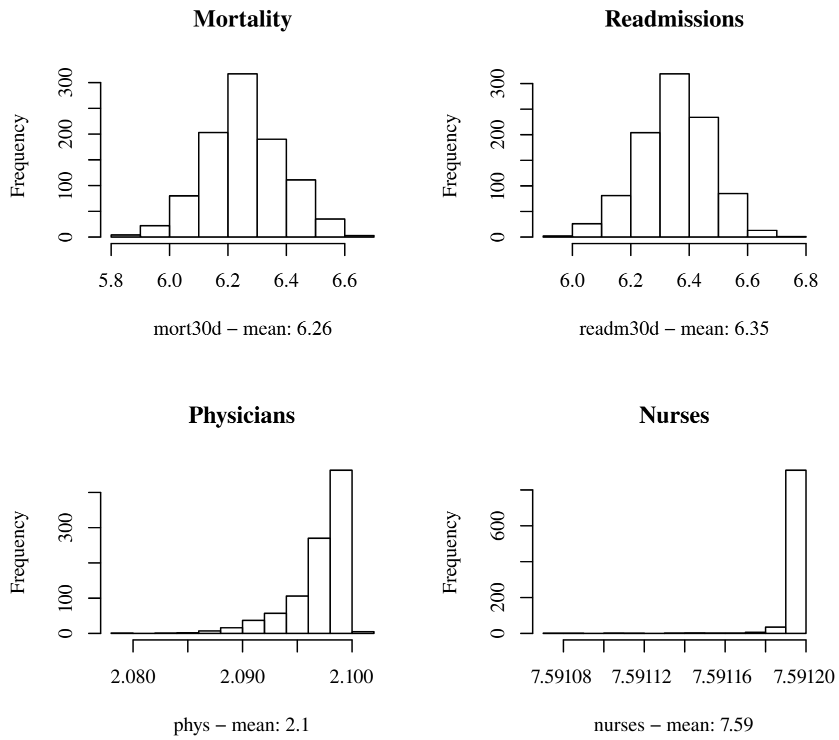



Figure 4: Posterior distribution of slack resources

## Diskussion der Ergebnisse, Limitationen und Ausblick

Die Transparenz hinsichtlich Struktur-, Prozess- und Ergebnisqualität unterstützt den Qualitätswettbewerb zwischen Leistungserbringern und erlaubt Patienten und Einweisern eine faktenbasierte Krankenhauswahl. Sie ermöglicht die Identifikation und Förderung von Best Practice in der Leistungserbringung und die Korrektur von vorher unentdeckten Mängeln.

Deshalb rückt die Qualitätstransparenz in Krankenhäusern stärker in den Fokus von Gesundheitspolitik und -wirtschaft. Jedoch ist der Nachholbedarf in Bezug auf Verfügbarkeit, Vergleichbarkeit und Verständlichkeit von Qualitätsinformationen gegenüber anderen Branchen groß. Die Ergebnisse dieser Arbeit tragen dazu bei, diese Lücke zu schließen. Dazu nimmt Kapitel 1 den Status Quo von Qualitätstransparenz auf und verdeutlicht die mikroökonomischen Limitationen bei der Bereitstellung von Behandlungsqualität, insbesondere auch in dem von Informationsasymmetrien geprägten Prinzipal-Agenten-Verhältnis zwischen Patient und Leistungserbringer. Kapitel 2 identifiziert auf Basis eines internationalen Vergleiches der Messung, Veröffentlichung und Incentivierung von Qualität fünf Maßnahmen zur Steigerung von Qualitätstransparenz. Kapitel 3 findet große Qualitätsunterschiede zwischen deutschen Krankenhäusern und diskutiert den Mangel an regulativen Konsequenzen für schlechte Qualität. Kapitel 4 analysiert das Nutzerverhalten auf dem Public Reporting Portal Weisse Liste.de und entwickelt Ansätze, um Public Reporting nutzerfreundlicher zu gestalten und Einfluss auf Krankenhauswahlentscheidungen zu stärken. Anschließend werden in Kapitel 5 und 6 die Daten des Qualitätstransparenzsystems angewandt, um Strukturfaktoren mit positivem Einfluss auf die Ergebnisqualität zu identifizieren. So können eine SU Spezialisierung sowie eine regionale und organisatorische Zentralisierung innerhalb eines Krankenhauses die Versorgungsqualität und Effizienz von Krankenhäusern verbessern. Ein Qualitätswettbewerb zwischen Krankenhäusern ist dagegen aktuell nur schwach ausgeprägt und es lässt sich kein systematischer Wettbewerbseinfluss auf die

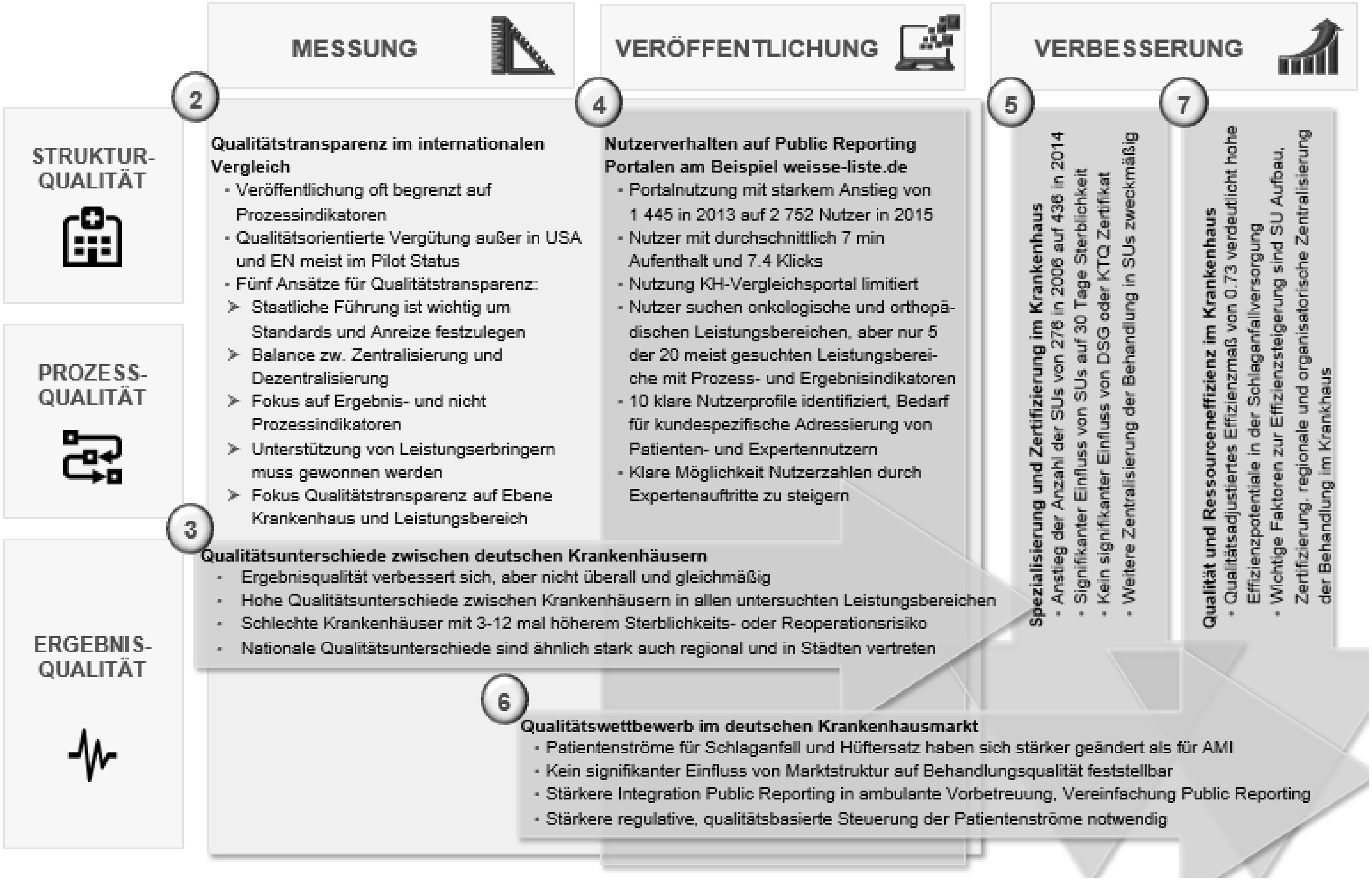
Behandlungsqualität feststellen. Ebenso konnte für Krankenhaus-Zertifizierungen kein eindeutiger positiver Effekt auf die Versorgungsqualität festgestellt werden.

Die Ergebnisse dieser Arbeit werden im Folgenden eingehend diskutiert und Handlungsempfehlungen für eine Verbesserung der Qualitätstransparenz und Steigerung der Behandlungsqualität vorgebracht. Letztlich werden die Limitationen dieser Arbeit aufgezeigt und ein Ausblick zur weiteren Forschungsagenda des Projektes „Qualitätstransparenz in Deutschland und deren Anwendung zur Steigerung der Behandlungsqualität im Krankenhaus" gegeben.

### *Ergebnisse*

Die sechs Kapitel liefern Erkenntnisse zur Steigerung der Qualitätstransparenz und wichtigen Krankenhausstrukturfaktoren entlang der Dimensionen: Messung, Veröffentlichung und Verbesserung von Qualität im Krankenhaus. Zusammen mit den drei Qualitätsdimensionen des Donabedischen Qualitätsmodells lässt sich eine 3x3 Matrix entwickeln (vgl. Abbildung 4), welche als Grundlage für die strukturierte Gesamtbetrachtung der Ergebnisse dieser Arbeit fungiert.

**Abbildung 4**: Gesamtbetrachtung der Forschungsergebnisse (Kapitel 2 bis Kapitel 7)




**Legende:** Kapitel 1 als Aufarbeitung der theoretischen Grundlagen wird in der Abbildung nicht separat aufgeführt.

Um Qualitätstransparenz und deren Anwendung zur Qualitätsverbesserung zu analysieren, wurden zunächst die allgemeinen und in Deutschland spezifischen Grundlagen für die Bereitstellung von Qualität im Gesundheitsmarkt betrachtet (Kapitel 1 – als Aufarbeitung der theoretischen Grundlage nicht aufgeführt in Abbildung 4). Dabei wurde deutlich, dass sowohl auf Nachfrageseite (d.h. Patienten und deren Stellvertreter) als auch auf Angebotsseite (d.h. Leistungserbringer) mehrere wichtige Faktoren die Erreichung des Marktgleichgewichtes in Hinblick auf die Bereitstellung von Qualität behindern. Auf Nachfrageseite sind das insbesondere die Informationsasymmetrie zwischen Patienten als Agenten und Ärzten als Prinzipalen und die Isolation des Konsumenten vom Preissignal. Der Patient kann vor der Behandlung durch den Leistungserbringer nur schwer die Eigenschaften des zu behandelnden Leistungserbringers einschätzen und während und nach der Behandlung den Einsatz des Arztes als auch Umfang und Qualität des Behandlungsergebnisses nicht oder nur minimal einschätzen.

Trotz der oft vorhandenen altruistischen Motivation des medizinischen Personals fehlen auf Krankenhausebene – in einem Zielkonflikt mit Kostendruck und Personalmangel – die Anreize, möglichst optimale Behandlungsqualität zu liefern. Ohne Qualitätstransparenz ist die qualitätsbasierte Krankenhauswahl durch Patienten und ihren Stellvertretern als auch die Integration von Qualität als Vertragsparameter zwischen Leistungserbringern und Kostenerstattern nur limitiert möglich.

Diesem Marktversagen werden in Deutschland und in vielen anderen Ländern staatliche Qualitätssicherung und staatlich oder privatwirtschaftlich organisiertes Public Reporting entgegengesetzt. Die Qualitätssicherung misst Struktur-, Prozess- und Ergebnisqualität und überprüft die Erreichung von ex-ante definierten Zielwerten. Doch werden bisher aus den gewonnenen Erkenntnissen noch keine systematischen Konsequenzen gezogen. So wurden in Deutschland bisher keine Krankenhäuser von den Planungsbehörden aufgrund unzureichender Qualität ausgeschlossen. Im langwierigen, strukturierten Dialog mit schriftlichen

Stellungnahmen gibt es trotz wiederholter Probleme meist keine Konsequenzen. Im Gegensatz zum regulativen Ansatz der Qualitätssicherung setzt Public Reporting auf die Selbstheilungskräfte des Marktes und versucht die Qualität in der Leistungserbringung für Patienten und Leistungserbringer transparent zu machen und so als wichtiges Entscheidungskriterium und Anreiz für Patienten und Leistungserbringer zu stärken. Diese Arbeit analysiert diese Maßnahmen und zeigt Stärken und Schwächen des deutschen Qualitätstransparenzsystems auf Basis eines internationalen Vergleiches, einer retrospektiven Analyse der Qualitätsunterschiede zwischen Krankenhäusern und dem Nutzerverhalten auf dem Public Reporting Portal Weisse Liste.de.

Qualitätstransparenz im internationalen Vergleich

Die Stärken des deutschen Ansatzes liegen im Fokus auf Leistungsbereichs- und Krankenhausebene, das breite Spektrum an Indikatoren insbesondere in Bezug auf risiko-adjustierte Ergebnisindikatoren und die Krankenhaus-Benchmarking-Möglichkeiten auf Portalen wie Weisse Liste.de. Die Qualität im Krankenhaus wird auf Ebene von Leistungsbereichen wie Hüft- und Knie-Implantationen und für einzelne Krankenhäuser dargelegt. Dies erlaubt Patienten und ihren Stellvertretern eine Wahlmöglichkeit zwischen den relevanten Behandlungseinheiten. Die Vielfalt an Indikatoren ermöglicht die Untersuchungen des Einflusses von Strukturfaktoren auf Ergebnisqualität. Risiko-adjustierte Ergebnisindikatoren sind Voraussetzung für eine Vergleichbarkeit von Ergebnisqualität zwischen Krankenhäusern. Die Krankenhaus-Benchmarking-Möglichkeiten schaffen die Voraussetzung für einen direkten Vergleich der für die jeweilige Krankenhauswahl relevanten Krankenhäuser – auf regionaler und auch auf nationaler Ebene.

Entwicklungspotentiale gibt es hingegen bei der Einbeziehung weiterer Leistungsbereiche wie onkologische Indikationen und weiterer orthopädischer Krankheitsbilder [1]. Es findet keine Messung und Veröffentlichung von Patient-Reported-Outcome-Measures (PROM) (z.B.

Oxford-Knee-Score) statt und die Vergleichbarkeit der Qualitätsergebnisse über Portale hinweg ist unzureichend [2, 3]. Bei der qualitätsorientierten Vergütung gibt es in Deutschland aktuell nur kleinere Pilotprojekte im Rahmen der integrierten Versorgung. Alle anderen Länder haben entweder deutlich umfassendere Piloten (z.B. Schweden) oder komplette qualitätsorientierte Vergütungssysteme aufgebaut (z.B. USA). In Deutschland hat das KHSG das Ziel ausgegeben, bis 2018 ein System für qualitätsorientierte Zu- und Abschläge einzuführen [4], doch zeigt sich die Umsetzung dieser Gesetzesinitiative als schwierig und in vollem Umfang unwahrscheinlich.

Qualitätsunterschiede zwischen Krankenhäusern

Auf Basis vorhandener Qualitätsindikatoren wurden Qualitätstrends und Unterschiede zwischen deutschen Krankenhäusern analysiert. Die Behandlungsqualität hat sich von 2006 bis 2014 je nach Bereich und Indikator unterschiedlich entwickelt [5]. Eine Steigerung der Ergebnisqualität konnte für Cholezystektomie, Herzschrittmacher-Implantation und Hüft-Endoprothesen-Implantation festgestellt werden. Die Krankenhaussterblichkeit für PCI hat sich dagegen verschlechtert. Für die Notfallindikatoren Schlaganfall und Herzinfarkt konnte eine Verbesserung der 30-Tage-Sterblichkeit festgestellt werden, gleichzeitig hat sich aber die 90-Tage Wiedereinlieferungsrate verschlechtert.

Es bestehen große Qualitätsunterscheide zwischen Krankenhäusern auf nationaler und regionaler Ebene. So steigt z.B. das Mortalitäts- und Reoperationsrisiko bei Cholezystektomie, Herzschrittmacher-Implantation, Hüft-Endoprothesen-Implantation und PCI durchschnittlich um das 5- bis 12-fache, wenn ein Patient in einem durchschnittlichen Krankenhaus aus dem fünften Quintil anstelle eines durchschnittlichen Krankenhauses aus dem zweiten Quintil behandelt wird. Für Schlaganfall und Herzinfarkt steigt das 30-Tage Sterblichkeitsrisiko um das 3-fache bei Einlieferung in ein Krankenhaus aus dem fünften Quintile relativ zum zweiten Quintile. Diese Qualitätsunterschiede bleiben auch bestehen, wenn man von einer nationalen auf eine regionale Betrachtung herabgeht. Gerade in Städten wie Berlin, München oder

Hamburg könnten Patienten aufgrund der größeren Anzahl erreichbarer Krankenhäuser von einer aktiven, qualitätsbasierten Steuerung ihrer Krankenhauswahl profitieren. Obwohl die identifizierten Qualitätsunterschiede den Mehrwert einer solchen Auswahl verdeutlichen, äußern sich zahlreiche Studien skeptisch bezüglich der tatsächlichen Wirkung von Public Reporting auf die Wahlentscheidungen von Patienten und ihren familiären oder ärztlichen Vertretern.

Neben den weiterhin hohen Qualitätsunterschieden zwischen Krankenhäusern konnten drei weitere methodische Probleme der externen Qualitätssicherung des G-BA identifiziert werden. Beide Probleme verdeutlichen die Komplexität von medizinischer Qualitätsmessung und Veröffentlichung. Auf der einen Seite gibt es viele Krankenhäuser, die aufgrund geringer Fallzahlen aus datenschutzrechtlichen Gründen von Reporting-Verpflichtungen ausgeschlossen werden. Dies betrifft z.B. im Jahr 2014 ungefähr 600 Krankenhäuser im Bereich Cholezystektomie und 400 Krankenhäuser im Bereich PCI [5]. Zusätzlich zu den datenschutzrechtlichen Bedenken ist die statistische Aussagekraft bei geringen Fallzahlen reduziert. Trotz der Nachvollziehbarkeit beider Gründe reduziert dieser Ausschluss die Qualitätstransparenz bei kleineren Krankenhäuser mit geringeren Fallzahlen, die oft eine schlechtere Ergebnisqualität aufzeigen [6–8]. Auf der anderen Seite ist die Vergleichbarkeit der risikoadjustieren Indikatoren zwischen den Jahren nur bedingt gegeben, da die Risikoadjustierung spezifisch für jedes Jahr durchgeführt wird [9, 10]. Hier kann für die Analyse des allgemeinen Trends auf die einfachen Ergebnisraten ausgewichen werden, doch sind für den Vergleich zwischen Krankenhäusern risikoadjustierte Indikatoren unabdingbar. Für den Analyse der risiko-adjustieren Ergebnisqualität von Krankenhäusern über die Zeit müssen die risikoadjustieren Indikatoren auf Basis eines Adjustierungsmodells und der Patientendaten von den verschiedenen Jahren für alle Jahre vergleichbar berechnet werden.

Die Qualitätsmängel vieler Krankenhäuser sind den Akteuren der Selbstverwaltung und der Politik grundsätzlich bewusst [11–13]. Sie sind jedoch im Konsens- und verhandlungsbasierten System der Selbstverwaltung nur schwer zu adressieren. Zusätzlich liegt die Krankenhausplanung und die Qualitätssicherung rechtlich im Aufgabenbereich der Bundesländer, was eine zentrale Steuerung schwer bis unmöglich macht [14, 15]. Zum Beispiel haben die 2004 bzw. 2005 eingeführten Mindestmengen für sechs chirurgische und orthopädische Eingriffe bisher zu keiner Konzentration der Behandlung geführt [16, 17], auch weil Ausnahmen durch die regionalen Planungsbehörden fast immer gewährt wurden.

Als alternativen zu regulativen Maßnahmen auf Landesebene führt das KHSG qualitätsorientierte Vergütung und selektive Qualitätsverträge zwischen Krankenkassen und Leistungserbringern ein. Die Implementierung von qualitätsorientierter Vergütung rückt aktuell im Zuge der schwierigen Diskussion um planungsrelevante Indikatoren in weite Ferne [18–21]. In Bezug auf Selektivverträge hat das KHSG die größere Erprobung von Selektivverträgen in vier Leistungsbereichen erlaubt [4] und einzelne Krankenkassen befinden sich bereits in der Testphase. So setzen die Selektivverträge der AOK Hessen für Versorgung bei krankhafter Adipositas und alloplastischem Gelenkersatz (Knie und Hüfte) Mindestmengenanforderungen und spezifische Qualitätsanforderungen an die Ärzte [22]. Letztere spezifizieren darüber hinaus Minimumwerte für die im Rahmen des QSR Verfahrens errechneten Hüft- und Knieendoprothetik Qualitätsindikatoren [23]. Ein Ausbau der Selektivverträge zu Kosten des Kollektivvertrages ist eine Möglichkeit, eine qualitätsbasierte Versorgungssteuerung einzuführen. Die Schwierigkeiten, Behandlungsqualität als wichtiges Kriterium in der Krankenhausplanung zu etablieren, verdeutlichen die Relevanz des Public Reportings.

<u>Nutzerverhalten auf Public Reporting Portalen am Beispiel weisse-liste.de</u>

Um den Wirkungsmechanismus von Public Reporting besser zu verstehen, untersucht Kapitel 4 das Nutzerverhalten für die Krankenhaussuche auf Deutschlands wichtigstem Public

Reporting Portal Weisse Liste.de. Die Public Reporting Nutzung ist den letzten Jahren von 1 445 Nutzern pro Tag in 2013 auf 2 752 Nutzer in 2015 stark angestiegen. Nutzer verbringen im Durchschnitt 7 Minuten im Portal, mit 7.4 Klicks und 54 Sekunden zwischen den Klicks. Die Nutzung der Vergleichsfunktion zwischen Krankenhäusern ist limitiert. Nutzer suchen Qualitätsinformationen insbesondere für onkologische und orthopädische Leistungsbereiche, doch gibt es nur für 5 der 20 meistgesuchten Indikatoren Prozess- und Ergebnisindikatoren. Auf Basis des Klickverhaltens konnten zehn verschiedene Nutzergruppen identifiziert werden. Die Ergebnisse lassen einen klaren Bedarf zum Ausbau der nutzerorientierten Inhalte und Darstellungsformen erkennen (z.B. verschiedene Subportale für unerfahrene Patientennutzer und ärztliches Fachpersonal). Dies ist insbesondere wichtig, um die hohen Abbruchraten zu reduzieren und die Nutzung der Vergleichsfunktion auszubauen [1, 24, 25]. So werden die Informationen, wie sie in den Qualitätsberichten erfasst und ohne große Vereinfachung auf Weisse Liste.de präsentiert werden, von Patienten als komplex und in Aufbereitung wenig ansprechend gesehen [26]. Die Elastizität der Nutzerzahlen auf Weisse Liste.de relativ zu Weisse Liste.de/Public Reporting Experteninhalten in Print und Fernsehen erlauben die Möglichkeit, die Relevanz der Krankenhauswahl und Public Reporting Portale als Medium dafür in der Bevölkerung zu stärken [1].

Spezialisierung und Zertifizierung im Krankenhaus

Die Kapitel 5 bis 7 wenden die durch Qualitätstransparenz gewonnenen Daten zu Struktur- und Ergebnisqualität an, um bestimme Strukturfaktoren als Möglichkeit zur Verbesserung von Ergebnisqualität und Ressourceneffizienz zu evaluieren. Kapitel 5 untersucht am Beispiel Schlaganfall den Einfluss von Infrastruktur- und Prozessspezialisierung sowie von Zertifizierung auf Ergebnisqualität. Auf der einen Seite wird eine signifikante Verbesserung der Schlaganfall 30-Tage Sterblichkeit durch Behandlung in einer SU festgestellt [27]. Auf der anderen Seite kann kein signifikanter Einfluss durch das DSG SU Zertifikat und das KTQ

Qualitätsmanagementzertifikat festgestellt werden. Durch eine Verschiebung der Behandlung von Schlaganfallpatienten aus Krankenhäuser ohne SU in Krankenhäuser mit SU und damit verbundene Zentralisierung kann die deutschlandweite Schlaganfallsterblichkeit um bis zu 460 Fälle reduziert werden. Die SU Anzahl hat sich in Deutschland seit 2006 von 276 auf 436 in 2014 stark erhöht. In einigen Regionen bedarf es vielleicht noch einer zusätzlichen SU. Dennoch sollten Schlaganfallpatienten, die aktuell noch in Krankenhäusern ohne SU behandelt werden, grundsätzlich in Krankenhäusern mit SU umgelenkt werden, um unnötigen zusätzlichen Kapazitätsaufbau zu vermeiden. Studien haben gezeigt, dass trotz der teilweise längeren Notfallanfahrtswege zu Krankenhäusern mit SU bessere Outcomes bei einer überregionalen Konzentration der Schlaganfallbehandlung in Krankenhäusern mit SUs zu erwarten sind [28].

Qualitätswettbewerb im deutschen Krankenhausmarkt

Kapitel 6 prüft den Einfluss von Krankenhauswettbewerb auf die Qualität der Versorgung in Deutschland am Beispiel der Notfallleistungsbereiche Schlaganfall und Herzinfarkt und des planbaren Leistungsbereichs Hüft-Total-Endoprothesenersatz. Während die Marktkonzentration für Schlaganfall zwischen 2006 und 2014 zugenommen hat, ist sie für Herzinfarkt relativ konstant geblieben und hat leicht abgenommen für Hüft-Total-Endoprothesenersatz. Für Schlaganfall und Hüft-Total-Endoprothesenersatz ist die Ergebnisqualität am schlechtesten für Krankenhäuser in hochkonzentrierten Märkten, für Herzinfarkt ist sie am schlechtesten für Krankenhäuser in mäßig konzentrierten Märkten. In den unterschiedlichen Regressionsmodellen konnte kein signifikanter Effekt von Wettbewerb auf Ergebnisqualität festgestellt werden. Dies mag auf der einen Seite an dem fehlenden geringen Bewusstsein für Qualitätsunterschiede zwischen Krankenhäusern und den dazu verfügbaren Public Reporting Daten liegen. Auf der anderen Seite können sich auch gegenläufig Effekte – Konzentration und Spezialisierung mit positiver Qualitätswirkung bei geringeren Fallzahlen und Wettbewerbsreduktion mit negativer Qualitätswirkung – gegenseitig aufheben. Die Arbeit verdeutlicht die Notwendigkeit des Ausbaus und der Optimierung des Public Reportings (Kapitel 4) und schafft die Grundlage für weitere Untersuchungen des Wirkungsmechanismus zwischen Wettbewerb und Behandlungsqualität auf Basis von Patientenlevel-Daten.

Qualität und Ressourceneffizienz im Krankenhaus

Abschließend wird in Kapitel 7 eine neue Methode zur Schätzung qualitätsadjustierter, technischer Krankenhauseffizienz entwickelt und auf die Schlaganfallversorgung angewandt. Neben den Produktionsfaktoren Qualität und Ressourceneinsatz fließen räumliche Faktoren wie Wettbewerb zwischen Krankenhäusern sowie Faktoren der demografischen Entwicklung und ambulanten Versorgung in die Bestimmung der Ineffizienz mit ein [29]. Die errechnete durchschnittliche technische Krankenhauseffizienz liegt bei einem Effizienzmaß von 0.73 relativ zum Optimum 1.0. Diese verdeutlicht hohe Potentiale zur Steigerung der qualitätsadjustierten Effizienz. Durch die Betrachtung der marginalen Effekte auf die Effizienz für das Durchschnittskrankenhaus können wichtige Faktoren zur Steigerung der Effizienz identifiziert werden. Hierzu gehören insbesondere eine Spezialisierung durch eine Stroke Unit, eine Fachabteilungszertifizierung, regionale Zentralisierung und organisatorische Konsolidierung der Behandlung innerhalb des Krankenhauses. Die Methodik erlaubt zusätzlich die Identifikation von einzelnen Krankenhäusern, welche die Versorgung von Schlaganfallpatienten mit einem besonders hohen Maß an qualitätsadjustierter Effizienz leisten. Diese Krankenhäuser können im Rahmen von Fallstudien untersucht werden, um Best Practices zu identifizieren und diese auch in anderen Krankenhäusern mit Nachholbedarf einzuführen. Durch die Verwendung der standardisierten Mortalitätsrate (SMR), welche auch für andere Leistungsbereiche wie Herzinfarkt und orthopädische Eingriffe verfügbar ist, kann die Methodik auch zur Schätzung der qualitätsadjustierten Effizienz in anderen Leistungsbereichen angewandt werden. Da die SMR auch in anderen Ländern (z.B. England) zur Messung von Qualität im Krankenhaus verwendet wird, kann die Methodik auch internationale Anwendung finden.

Kapitel 7 verdeutlicht darüber hinaus einen Trade-off zwischen Behandlungsqualität und Kostenreduktion. Effizienzsteigerung durch Spezialisierung, Zertifizierung oder Konzentration der Behandlung können sowohl zur Steigerung der Qualität als auch zur Reduktion der eingesetzten Ressourcen verwenden werden. Letztlich ist eine Annäherung an die Effizienzgrenze im ökonomischen Sinne nur durch beide Strategien möglich. Qualitätstransparenz ist eine notwendige Bedingung, um die Effizienz der eingesetzten Ressourcen zu erhöhen bzw. den Einsatz von Ressourcen dort zu erhöhen, wo die Behandlungsqualität möglichst hoch ist und dort zu reduzieren, wo die Behandlungsqualität gering ist.

Der Einfluss von SU Zertifikaten auf die Ergebnisqualität in der Schlaganfallversorgung ist unklar bzw. die Ergebnisse dazu unterscheiden sich. In Kapitel 5 kann auf Basis eines Fixed-Effects-Modells keine signifikante bessere Behandlung für SUs mit DSG SU Zertifikat festgestellt werden. Dahingegen wird in Kapitel 7 ein signifikanter qualitäts- und effizienzsteigernder Zusatzeffekt für SU Zertifikate festgestellt. Die genauen Gründe für die unterschiedlichen Ergebnisse lassen sich aufgrund der unterschiedlichen statistischen Methoden (SFA vs. Fixed-Effekt Regression) nicht beschreiben.

Insgesamt verdeutlicht Abbildung 4 die zielorientierte Ausrichtung in einem Qualitätstransparenzsystem. Strukturfaktoren sind die Grundbausteine der Krankenhausversorgung und jeglicher dazugehörigen Prozesse. Aus der Kombination von Strukturfaktoren und Prozessfaktoren entsteht die Ergebnisqualität, welche als Endpunkt das ultimative Kriterium für die Effizienz und die Qualität der Behandlung ist. Gleichzeitig gilt Qualitätsmessung als Startpunkt für Qualitätstransparenz. Die über alle Leistungserbringer hinweg standardisiert gemessenen Struktur-, Prozess- und Ergebnisindikatoren fungieren als Grundlage für das Public Reporting und die Information von Patienten und ihren Stellvertretern. Auch fungieren sie als Grundlage für interne Verbesserungsprozesse im Krankenhaus (i.e. Qualitätsmanagement) und die systematische Identifikation von Best Practice Krankenhäusern und Strukturfaktoren mit hohen Einfluss auf gute Ergebnisqualität. So zielt das gesamte System auf die Verbesserung von Ergebnisqualität für den Patienten.

### *Handlungsempfehlungen*

Aus der Analyse ergeben sich wichtige Empfehlungen zur Verbesserung des deutschen Qualitätstransparenzsystems und zur Steigerung der Behandlungsqualität in deutschen Krankenhäusern. Die verpflichtende Qualitätssicherung des G-BA und das Public Reporting durch Weisse Liste.de stehen dabei im Zentrum. Aber auch für andere Qualitätstransparenzakteure sind die Empfehlungen relevant. Zusammengenommen können sie die Aussage- und Wirkungskraft des deutschen Qualitätstransparenzsystems stärken und einen wirksameren Qualitätswettbewerb zwischen Leistungserbringern entfachen. Eine Stärkung der Strukturfaktoren mit positiver Wirkung auf die Ergebnisqualität kann die Qualität in der Leistungserbringung fördern und mit den entwickelten Methoden können weitere relevante Strukturfaktoren identifiziert werden. Im Folgenden werden die Empfehlungen entlang der Dimensionen Messung, Veröffentlichung und Verbesserung aus Abbildung 4 dargestellt.

#### Messung von Qualität

Für eine Steigerung der Qualitätstransparenz sollte das Indikatoren-Set sowohl in bestehenden Leistungsbereichen als auch für neue Leistungsbereiche ausgebaut werden. Innerhalb bestehender Leistungsbereiche sollten die Anzahl Betten auf Fachabteilungsebene und Informationen zur Zentrenbildung (z.B. Prostata- oder Darmzentrum der DKG) als zusätzliche Strukturfaktoren aufgenommen werden, da sie einen wichtigen Einfluss auf Qualität und Effizienz in die Datenspezifikation der externen Qualitätssicherung haben [30–32].

Wichtiger jedoch ist die Einführung von PROM-Indikatoren. Hier hat Deutschland im internationalen Vergleich großen Nachholbedarf [2]. Neben „harten Endpunkten“, wie Mortalität oder Komplikation, zählen auch weichere Behandlungsergebnisse wie Lebensqualität, Beweglichkeit und motorische Fähigkeiten als wichtige Gradmesser einer guten Krankenhausqualität [33–35] und hoher Relevanz für Patienten [36]. Das International Consortium for Health Outcome Measurement (ICHOM) hat bereits mehr als 20 PROM Indikatoren-Sets entwickelt und bietet eine anerkannte und in vielen Ländern bereits angewandte Möglichkeit PROM Indikatoren zu messen [37]. Die ICHOM Indikatoren-Sets können als Grundlage für eine PROM Indikator Offensive in Deutschland fungieren. Die Integration von PROM ist eine stärkere Ausrichtung des Gesundheitssystems am Patienten. Grundsätzlich sollte die Nachfrage nach Informationen als stärkeres Kriterium für die Weiterentwicklung herangezogen werden, um die Nutzerorientierung des Qualitätstransparenzsystems zu stärken [1].

Der Ausbau der Qualitätstransparenz für weitere Leistungsbereiche sollte darüber hinaus höchste Priorität haben. Aktuell deckt die externe Qualitätssicherung mit ihren 27 Leistungsbereichen weniger als 30% aller stationären Aufnahmen ab [2, 10]. Andere Länder, wie die USA oder England, haben hier einen deutlich umfassenderen Ansatz [2]. Prozess- oder Ergebnisindikatoren sollten insbesondere für onkologische (z.B. Darm- oder Brustkrebs) als auch weitere orthopädische Indikatoren (z.B. Bandscheibenvorfall) ergänzt werden [1]. Das Qualitätstransparenzsystem sollte sich grundsätzlich stärker an der Patientennachfrage und Interesse orientieren.

Zusätzlich sollte der Ausbau der sektorenübergreifenden Qualitätsmessung vorangetrieben werden. Aktuell gibt es in der gesetzlichen Qualitätssicherung mit PCI und postoperativer Wundinfektion nur zwei Leistungsbereiche mit sektorenübergreifenden Indikatoren. Die Einbeziehung von Ereignissen nach der Krankenhausentlassung ist essentiell, um ein vollständiges Bild über die geleistete Behandlungsqualität zu erlangen. Ein zu frühes Entlassen im Zuge des finanziellen Drucks durch Fallpauschalen kann sich negativ auf längerfristige Gesundheitsergebnisse auswirken. Zudem sind auch längere Zeiträume als nur 30 Tage von Bedeutung, da sich die Ergebnisqualität in kürzeren und längeren Zeiträumen gegensätzlich entwickeln kann [5]. Die Entwicklung von sektorenübergreifenden Indikatoren sollte, wie bei den QSR Indikatoren, auf Basis von Routinedaten geschehen, um den Dokumentationsaufwand für Krankenhäuser nicht weiter zu erhöhen.

Methodisch sollte darüber hinaus der Ausschluss von kleineren Krankenhäusern mit geringen Fallzahlen aufgrund von Datenschutz überdacht werden. Qualitätstransparenz sollte für alle Krankenhäuser gleichermaßen gelten, insbesondere auch für solche Krankenhäuser, die bestimmte kritische Eingriffe nur selten und in kleinen Fallzahlen vornehmen. Zwei mögliche Ansätze um Datenschutzbedenken vorzubeugen, sind das Gruppieren der Fallzahlen und Ergebnisse über 2 Jahre oder die Veröffentlichung von relativen Ergebnissen oder Aussagen wie unterdurchschnittlich, durchschnittlich und überdurchschnittlich für Krankenhäuser mit geringen Fallzahlen.

Veröffentlichung von Qualitätsinformation

Aktuell werden die gesetzlichen Qualitätsindikatoren erst mehr als 14 Monate nach Ende des Berichtsjahres veröffentlicht [10]. Dieser Zeitraum sollte deutlich verkürzt werden. Eine qualitätsbasierte Steuerung der Versorgung sowohl durch regulatorische Eingriffe als auch durch Patientenwahl bedarf möglichst aktuellen Daten. Nur so ist Validität der Entscheidungsgrundlage gewährleistet. Möglich wäre sogar ein Qualitätsreporting in Echtzeit. So bekommen Chefärzte der HELIOS Kliniken monatliche Qualitätsleistungskennzahlen und treffen auf Basis dieser aktuellen Kennzahlen Managemententscheidungen [38].

Online-Kommunikation durch Public Reporting Portale sollten sowohl kundenspezifischer als auch nutzerfreundlicher gestaltet werden. Denn die Ergebnisse machen deutlich, dass die

Fähigkeiten, Qualitätsinformationen zu verstehen und richtig einzusetzen, stark zwischen den einzelnen Nutzergruppen variieren. Die Krankenhaussuche auf Weisse Liste.de könnte mit der Eingabe des Leistungsbereiches und der Postleitzahl die Option zur freiwilligen Selbstidentifikation als unerfahrener Patient, erfahrener Patient oder Mediziner erlauben. Dadurch könnten nutzergruppenspezifische Informationen angeboten werden. Zusätzlich sollte die Option eines Log-Ins die Speicherung von Profil-Informationen (z.B. Nutzertyp, Suchstandort) erlauben. Innerhalb der Portale sollte die Vergleichsfunktion als zentraler Bestandteil gestärkt werden. Nur in dieser Funktion findet wirklich eine Gegenüberstellung und Auswahl zwischen verschiedenen Leistungserbringern statt. Ein Vergleich der Krankenhäuser mit der höchsten Behandlungsqualität sollte einfach und schnell auf eine geringe Zahl an relevanten (d.h. Leistungsbereiche) und guten (d.h. Ergebnisqualität) Krankenhäusern zu reduzieren sein. Dies kann zum Beispiel über einen Qualitätsfilter funktionieren. Dieser wird auf Basis von vordefinierten Kriterien wie Minimumfallzahl und Zugehörigkeit zum qualitativ besten Drittel der Krankenhäuser die Auswahl an Krankenhäusern auch in Ballungsräumen schnell auf die relevanten Krankenhäuser reduzieren [1].

Einfache und intuitive Onlinevergleichsfunktionen sind das Kernstück von erfolgreichen Vergleichsportalen wie *Booking.com* oder *Hotel Tonight*. Es fällt u.a. auf, dass jedes Vergleichsportal eine responsive Landkarte (z.B. *Google Maps*) mit den ausgewählten Objekten (z.B. Hotels) anbietet. Einfaches An- und Ausschalten von Suchkriterien ist auch weitverbreitete Anwendung auf Hotel -oder Flugsuchportalen. Diese Funktionen und viele andere einfache und intuitive Funktionen von Vergleichsportalen aus anderen Wirtschaftszweigen sollten nach entsprechender Prüfung auch in Public Reporting Portalen integriert werden.

Auch die Offline-Kommunikation von Qualitätsunterschieden an Patienten und ihre Stellvertreter sollte ausgeweitet werden. Nach Feststellung des stationären Behandlungsbedarfs könnten die Krankenkassen einfach aufbereitete Public Reporting Informationen zur Unterstützung der Krankenhauswahl zum Patienten nach Hause schicken. Auch könnten regelmäßige, einfache Beilagen mit Qualitätsinformationen zu regionalen Krankenhäusern in Tageszeitungen die Bekanntheit und das Verständnis des Public Reporting stärken und wichtige Qualitätsinformationen zur Krankenhauswahl an Patienten, vor allem auch älteren Patienten mit weniger Onlineaffinität, vermitteln. Noch besser wäre eine Integration des Public Reporting in die Arzt-Patienten Gespräche im ambulanten Sektor. Einweisende Ärzte könnten die Krankenhauswahl zusammen mit Ihren Patienten am Desktop- oder Tablet-PC besprechen und so das richtige Krankenhaus identifizieren. Ärzte könnten dabei hilfreichen medizinischen Kontext zur korrekten Interpretation der Public Reporting Daten beitragen. Schließlich bietet sich auch eine Kombination aus Print und Online Angebot an. Ein erfolgreiches Beispiel bietet der Klinikführer Rhein-Ruhr: Schon 2005 war dessen Erstauflage mit 10 000 Exemplaren in weniger als 24 Stunden ausverkauft. Innerhalb von wenigen Monaten hatten zwei Million Nutzer die Onlineausgabe besucht [39]. In den Folgejahren konnte der Klinikführer seine Auflage weiter steigern.

Diese Verbesserungsvorschläge zeigen: Das Public Reporting sollte deutlich einfacher werden, näher an den Patienten als Kunden herangeführt werden und zusätzlich zur Weiterentwicklung und Individualisierung der Online-Plattformen auch eine stärkere Print- bzw. Offline-Komponente bekommen. Ohne eine deutliche Steigerung der Bekanntheit und des Nutzens von Public Reporting ist die Übersetzungsfunktion der Veröffentlichung von Krankenhausqualität zwischen Messung und Verbesserung begrenzt. Doch letztlich ist Qualitätstransparenz kein Selbstzweck, sondern sollte neben der Stärkung des Krankenhausqualitätswettbewerbs und der Durchführung des krankenhausinternen Qualitätsmanagements auch aktiv zur Verbesserung der Krankenhausqualität auf Systemebene genutzt werden.

Verbesserung von Qualität

Auf Systemebene können die durch Qualitätstransparenz gesammelten Struktur- und Ergebnisindikatoren zur retrospektiven Analyse von Strukturdeterminanten von Ergebnisqualität genutzt werden. So können wichtige Strukturfaktoren identifiziert, deren positiver Effekt auf die Ergebnisqualität verifiziert und regulative und Vergütungsanreize für diese Strukturfaktoren gestärkt werden. Zum Beispiel konnte diese Arbeit die bessere Behandlungsqualität in SUs und Nutzen von weiterer Konzentration der Schlaganfallbehandlung in SU aufzeigen.

Qualitätswettbewerb zwischen Krankenhäusern kann sich positiv auf die Behandlungsqualität auswirken. Doch ist die Wirkung in Deutschland noch unklar, auch weil der Qualitätswettbewerb noch unzureichend ausgeprägt ist. Daher gilt es den Qualitätswettbewerb zu stärken. Möglich wäre dies anhand von drei parallel zu verfolgenden Ansätzen. Die oben beschriebenen Verbesserungsvorschläge für das Public Reporting können qualitätsbasierte Krankenhauswahlentscheidungen durch Patienten stärken. Zusätzlich können durch qualitätsorientierte Vergütung finanzielle Anreize für besser Qualität geschaffen werden. Letztlich können qualitätsbasierte Selektivverträge den Qualitätswettbewerb zwischen Krankenhäusern stärken [5].

Qualitätsorientierte Vergütung schafft einen zusätzlichen Anreiz zur Investition in qualitätssteigernde Infrastruktur und Prozesse (vgl. Kapitel 1). In Deutschland wurden qualitätsorientierte Vergütungssysteme bisher nur vereinzelt und in begrenzten Piloten innerhalb der integrierten Versorgung erprobt. Andere Länder haben dort bereits weitere Schritte unternommen (vgl. Kapitel 2). Doch ist die wissenschaftliche Fundierung des positiven Effektes bisher nur schwach, auch aufgrund von methodischen Limitationen [40, 41]. Daher sollte das System der qualitätsorientierten Vergütung in ein bis zwei Leistungsbereichen auf nationaler Ebene pilotiert werden, ohne weitere Änderungen am Vergütungssystem, um verzerrende Einflüsse zu minimeren. Als mögliche Piloten für die qualitätsorientierte Vergütung bieten sich elektive und standardisierte Eingriffe wie die Implantation von Knie- oder Hüft-Endoprothesen an. Das bestehende Qualitätstransparenzsystem sollte dann angewandt werden, um die Wirkung der qualitätsorientierten Vergütung zu evaluieren.

Mit Hilfe von qualitätsbasierten Selektivverträgen können Krankenkassen als Kostenerstatter die Zentralisierung und Spezialisierung in der Leistungserbringung stärken. Sie können für elektive Leistungsbereiche wie Hüft- oder Knie-Endroprothesen-Implantationen oder Herzschrittmacher-Implantationen regionale Versorgungsverträge mit den qualitativ besten Kliniken in einer Region abschließen, um so Hochleistungszentren zu stärken oder aufzubauen. Auf Basis von Qualitätstransparenz können die Krankenkassen Krankenhäuser und Fachabteilungen mit hoher Ergebnisqualität identifizieren (Kapitel 3) und gleichzeitig auch die qualitätsadjustierte Versorgungseffizienz (Kapitel 7) in die Auswahl der relevanten Kliniken mit einfließen lassen. Zusätzlich können auf Basis der Qualitätstransparenzdaten, angereichert mit Patientendaten wie Wohnortentfernung zum Krankenhaus, die Auswirkungen von regulativen Eingriffen in die Versorgungslandschaft auf zu erwartende Ergebnisqualität, Leistungsdichte, Fahrtzeiten von Patienten und Notfallversorgung simuliert werden [13, 29].

Eine stärkere Rolle der Krankenkassen in der Auswahl von Krankenhäusern mit Qualitätsführerschaft würde die Verhandlungsmacht der Käufer im Business-to-Business (B2B) Verhältnis zwischen Leistungserbringern und Kostenerstattern stärken. Im Vergleich zum Business-to-Consumer (B2C) Verhältnis zwischen Leistungserbringern und Patienten sind die Informationsasymmetrien in einem B2B Verhältnis zwischen Leistungserbringern und Krankenkassen geringer. Die AOK Hessen verfolgt diese Strategie bereits für die Leistungsbereiche Hüft- und Knieersatz und Adipositaschirurgie [22, 23].

Zur systematischen Analyse der Verbesserungspotentiale sollte das IQTIG eine Abteilung aufbauen, die dem G-BA durch Qualitätstransparenzdaten informierte Vorschläge zur

Verbesserung der Versorgungslandschaft unterbreitet. Die vorliegende Arbeit nutzt die Qualitätstransparenzdaten des G-BA und anderer Institute für genau diesen versorgungswissenschaftlichen Ansatz. Für eine breitere Nutzung der durch Qualitätstransparenz generierten Daten wären eine standardisierte Aufbereitung und ein einfacher Zugang erforderlich. Aktuell werden die Qualitätsberichte der deutschen Krankenhäuser als einzelne XML Text Dateien vom G-BA zur Verfügung gestellt. Technischer und zeitlicher Aufwand für die Datenvalidierung sind hoch. Um diese Hürde zu reduzieren, könnte die Integration bereits beim G-BA oder IQTIG erfolgen, wo sowohl Expertise für die Daten als auch das grundsätzliche Interesse an einer entsprechenden Nutzung der Daten hoch ist. Der G-BA könnte eine standardisierte Datenbank zur Verfügung stellen, in die auch zusätzliche Organisationen ihre Daten einspielen (z.B. Zertifizierungsagenturen). Der Zugang zu dieser Datenbank für Forschungszwecke könnte dann über ein zentrales G-BA Datenportal gemanagt werden.

### *Limitationen*

Spezifische Limitationen der einzelnen Studien wurden bereits im Zuge der jeweiligen Kapitel 2 – 7 diskutiert. Darüber hinaus gibt es weitere Limitationen, die in der Gesamtbetrachtung der Arbeit anzuführen sind und thematische Einschränkungen sowie methodische Aspekte betreffen. Die Limitationen können teilweise in weiteren Forschungsprojekten adressiert werden. Doch bedarf es für einzelne Aspekte auch einer Weiterentwicklung des Qualitätstransparenzsystems und einem Ausbau der Datenverfügbarkeit.

Die vorliegende Arbeit konzentriert sich auf die Qualität im Krankenhaus. Sie klammert die Qualitätsmessung und Darstellung auf regionaler Ebene und auf Ebene des einzelnen Arztes aus. Die Messung regionaler Versorgungsqualität hat möglicherweise einen Nutzen für die Versorgungsforschung und die Politik. Der Nutzen für Patienten ist jedoch limitiert, weil sie nur wenig Relevanz für die Wahl von Leistungserbringern hat. Viele Konsumenten wünschen sich

Informationen zur Behandlungsqualität einzelner Ärzte [42, 43], was sich in der Popularität von Arztbewertungsportalen wie Jameda.de ausdrückt. Doch sprechen zahlreiche Gründe gegen die Messung und Veröffentlichung von Qualitätsdaten auf Arztebene: Methodische Gründe gegen arztbasierte Indikatoren sind die geringe Validität von Indikatoren auf Basis von kleinen Patientengruppen und die oft wenig standardisierte Ergebnismessung [2, 44]. Politisch als auch aus Datenschutzgründen ist eine Messung und Veröffentlichung auf Arztebene noch deutlich kontroverser als die Veröffentlichung auf Krankenhausebene. Aus diesen Gründen werden in Deutschland Qualitätsindikatoren im stationären Sektor ausschließlich auf der Ebene des Leistungsbereiches im Krankenhaus gemessen und eine umfangreiche empirische Arbeit ist nur mit diesen möglich.

Zudem konzentriert sich die vorliegende Arbeit primär auf die stationäre Versorgung; sektorenübergreifende Betrachtungen werden nur im Rahmen der QSR Indikatoren unternommen. Obwohl die sektorenübergreifende Betrachtung der Behandlungsqualität von Wissenschaft und Politik gefordert wird, ist die Entwicklung und Messung von sektorenübergreifenden Indikatoren für die verpflichtende Qualitätssicherung nach Paragraph 137 SGB V bisher nicht über die Pilotphase hinausgegangen. Die Ausnahme in Deutschland sind die QSR Daten auf Basis von Routinedaten der AOK Patienten. Für die Leistungsbereiche Schlaganfall (Kapitel 3, 5, 6 und 7) und Herzinfarkt (Kapitel 3 und 6) werden auch diese Indikatoren verwendet.

Weitere, empirische Limitationen ergeben sich durch die beschränkte Datenverfügbarkeit. So sind PROM Ergebnisindikatoren nicht zugänglich, obwohl diese wichtigen Aspekte der Behandlung, neben den Endpunkten Mortalität und Komplikation, betrachten. Des Weiteren spielt die Risikoadjustierung für die Vergleichbarkeit von Qualitätsindikatoren zwischen Krankenhäusern eine entscheidende Rolle. Da die Arbeit ausschließlich auf Sekundärdaten beruht, wird die bereits durchgeführte Risikoadjustierung bei den entsprechenden Indikatoren als gegeben betrachtet. Die Komplexität der Risikoadjustierung konnte in diesem Rahmen ansatzweise diskutiert werden, da diese eine ganz entscheidende Rolle für die Validität und Akzeptanz von Ergebnisindikatoren spielt.

Die Nutzung von Sekundärdaten in der vorliegenden Arbeit unterliegt, wie alle Sekundärdatenanalysen der Krankenhausangaben, grundsätzliche Limitationen. Qualitätsberichte werden von den Krankenhäusern, auf Basis von detaillierten G-BA Richtlinien, selbst erstellt [45]. Bei diesem Prozess können ungewollte Fehler als auch bewusste Falschangaben auftreten, da Krankenhäuser im Wettbewerb miteinander stehen und sich daher um Ihre Reputation sorgen [5]. So haben stichprobenartige Validitätsprüfungen auf Basis von Patientenakten ergeben, dass bei 5% der Krankenhäuser Fehler im Reporting für 15-60% der untersuchten Datenaspekte zu verzeichnen sind [46, 47]. Dennoch lässt sich aufgrund der detaillierten Berichtregelwerke des G-BA, der unterstützenden Software und der standardisierten Datenvalidierungsverfahren von einer hohen Datenqualität ausgehen [10]. Die QSR Indikatoren werden auf Basis von Patientendaten zentral berechnet. Sie sind daher weniger anfällig für krankenhausspezifische Probleme bei Dokumentation und Reporting.

Die Arbeit nutzt ausschließlich sekundäre Qualitätsdaten, welche zuvor von Patientenebene auf die Ebene des Krankenhauses aggregiert wurden. Dies verhindert die Nutzung von Patientenwahlmodellen, welche in der Literatur verwendet werden, um die Krankenhauswahl unter Einfluss von Qualitätstransparenz zu modellieren [36, 48–50]. So werden zum Beispiel Patientenwahlmodelle, auf Basis von Patienteninformation wie Wohnort, Alter, Krankheitsbild zusätzlich zu den Krankenhaus-Charakteristika, in vielen Analysen zum Effekt von Wettbewerb auf Krankenhausqualität genutzt [48, 50–53]. Dies ist in der vorliegenden Arbeit nicht möglich.

Da Qualitätstransparenz für bestimmte Leistungsbereiche und in allen Krankenhäuser gleichzeitig eingeführt wurde, besteht nicht die Möglichkeit Kontrollgruppen zu bilden. Gleichzeitig sind Qualitätsverbesserungsdynamiken in verschiedenen Leistungsbereichen zu unterschiedlich, um Leistungsbereiche mit und ohne Qualitätstransparenz miteinander zu vergleichen. Diese beiden Aspekte erschweren den Aufbau eines ökonometrischen Ansatzes zur Isolierung des kausalen Effektes von Qualitätstransparenz auf die Behandlungsqualität.

Letztlich können in dieser Arbeit nur einzelne ausgewählte Aspekte von Qualitätstransparenz und darauf basierender Qualitätsverbesserung analysiert werden. Das System Qualitätstransparenz und die Anwendung zur Qualitätssteigerung sind zu umfassend, um in einer Arbeit vollumfänglich betrachtet zu werden. Doch wurden in Bezug auf Messung und Veröffentlichung von Qualität wichtige Probleme identifiziert und konkrete Verbesserungsvorschläge entwickelt. Das System Qualitätstransparenz wurde dann angewandt, um ausgewählte Strukturfaktoren für Qualitätssteigerung zu untersuchen und neue Analysemethoden auf Basis der Qualitätstransparenzdaten zu entwickeln.

### *Ausblick Forschungsagenda*

Die mikroökonomischen Limitationen in der Bereitstellung von Qualität als auch die staatlichen und privatwirtschaftlichen Lösungsansätze um diese Limitationen zu überkommen wurden im Detail und aus verschiedenen Blickwinkel betrachtet. Wichtige Empfehlungen für eine Steigerung der Qualitätstransparenz und einer Stärkung von wichtigen Strukturfaktoren mit positiven Einfluss auf Ergebnisqualität wurden formuliert. Die gewonnenen Erkenntnisse erlauben den Entwurf einer weiterführenden Forschungsagenda für das Projekt „Qualitätstransparenz in Deutschland und deren Anwendung zur Steigerung der Behandlungsqualität im Krankenhaus“. Auf der einen Seite knüpfen die weiterführenden Forschungsansätze direkt an die in diese Dissertation eingebrachten wissenschaftlichen Artikel an, auf der anderen Seite handelt es sich um weiterführende Projektideen zur Steigerung von Qualitätstransparenz und Evaluation von Strukturfaktoren zur Qualitätsverbesserung.

Forschungsansätze mit direktem Bezug zu dieser Arbeit

Der internationale Vergleich der Qualitätstransparenzsysteme in Kapitel 2 sollte um zwei Dimensionen erweitert werden. Um die Weiterentwicklung der Qualitätsindikatoren zu unterstützen, bietet sich eine Vergleichsarbeit in Bezug auf Anwendung und genaue Gestaltung von Qualitätsindikatoren für wichtige Leistungsbereiche wie Kardiologie und Orthopädie an. Hier wäre zu untersuchen welche Indikatoren, vor allem Ergebnisindikatoren und PROMs, zum Einsatz kommen, auf welchen Daten diese basieren und welche Art von Risikoadjustierung durchgeführt wird. In einem weiteren Schritt wäre es für die Diskussion innerhalb des deutschen Gesundheitssystems zur Anwendung der Qualitätsindikatoren hilfreich, international Beispiele zu identifizieren, bei denen Qualitätsindikatoren eine konkrete Auswirkung auf die Versorgungslandschaft haben. Diese Auswirkung kann durch Restriktion von Leistungen für qualitativ schlechte Krankenhäuser oder eine Reduktion der Vergütung geschehen. Auf Basis der für den deutschen Versorgungkontext und Datenlandschaft geeignetsten Indikatoren und verbundenen Maßnahmen könnten dann Szenarien durchgespielt werden, wie eine striktere Anwendung der Indikatoren mit wirklichen Konsequenzen aussehen könnte und welche Auswirkungen diese hätte.

Die Untersuchung der Qualitätsentwicklung und Unterschiede zwischen Krankenhäusern in Kapitel 3 kann insbesondere durch eine Analyse weiterer Leistungsbereiche und Indikatoren ergänzt werden. Hier bieten sich vor allem eine stärkere Verwendung der QSR Indikatoren an, die es neben Schlaganfall und Herzinfarkt auch für 16 weitere Leistungsbereiche wie Herzinsuffizienz, Sectio und Vaginale Entbindung gibt [54]. In einer weiteren Kooperation mit dem WidO könnten die Indikatoren auf Basis eines spezifischen Jahresmodells zur Risikoadjustierung über alle Jahre vergleichbar errechnet werden.

Die Public Reporting Untersuchung im Kapitel 4 wird durch ein bereits initiiertes Folgeprojekt zusammen mit Weisse Liste.de vertieft. Hier werden für ausgewählte Leistungsbereiche (z.B.

Hüft- und Knieersatz, Bypass-Operationen oder Geburten) Fragebögen auf der Webseite Weisse Liste.de geschaltet, welche Intention, Hintergrund und Datenverständnis spezifischer Nutzer eruieren. Die Umfrageantworten werden für jeden Teilnehmer mit seinem tatsächlichen Nutzerverhalten verbunden, um so als erste Studie Informationen zu Hintergrund, Intention, Verhalten und Verständnis von Public Reporting Nutzern zu erfassen. So können noch spezifischere Empfehlungen für die Weiterentwicklung des Public Reporting hin zu einem kundenspezifischen und wirksamen Motor des Qualitätswettbewerbs entwickelt werden.

Auf Basis eines umfangreichen DRG Patientenlevel-Datensatzes der Forschungsdatenzentren der Länder und des Bundes für zwei elektive Leistungsbereiche, wie Hüftimplantation und Bypass Operation, sollten Patientenwahlmodelle entwickelt und geschätzt werden, um die in Kapitel 6 untersuchten Auswirkungen von Qualitätswettbewerb zwischen Krankenhäusern auf die Versorgungsqualität und Marktanteile genauer zu untersuchen. Die Patientenwahlmodelle erlauben eine Berechnung des Konzentrationsmaßes auf Basis eines flexiblen Radius Ansatzes. Auch erlauben diese Patientendaten eine empirische Überprüfung der Frage, ob Krankenhäuser mit besserer Qualität mehr Patienten anlocken können und somit relativ einen höheren Marktanteil erreichen.

Das geoadditive SFA Modell zur Schätzung von qualitätsadjustierter Effizienz aus Kapitel 7 kann auf der einen Seite durch die Integration weiterer Ressourcen oder Kostendaten, wie Bettenkapazität in den jeweiligen Abteilungen, weiterentwickelt werden. Auf der anderen Seite kann das Modell zur Schätzung der qualitätsadjustieren Effizienz in anderen Leistungsbereichen, wie die Notfallindikation Herzinfarkt oder die planbaren Prozeduren Hüft- oder Knieersatz, eingesetzt werden. Neben diesen Forschungsansätzen mit direktem Bezug zu dieser Arbeit sind drei weitere Forschungsprojekte angedacht.

Weiterführende Forschungsansätze

Zuerst sollte der Zusammenhang zwischen den drei in Deutschland geführten Indikatortypen zur Krankenhausbewertung untersucht werden. Dazu zählen die Qualitätsindikatoren der gesetzlichen Qualitätssicherung, die Ranking Daten der Focus Klinikliste (insbesondere Reputation) und die Patientenerfahrung aus den Patient Experience Questionaire (PEQ) Daten der Weisse Liste. Partnerschaften für die letzteren beiden Datenblöcke konnten bereits abgeschlossen werden. Einerseits kann die Korrelation zwischen den verschiedenen Datenwelten untersucht werden. Anderseits kann untersucht werden, wie aussagekräftig das Focus Klinik Ranking und die Patientenerfahrungen für die Ergebnisqualität sind. Dies kann helfen, die Zusammenhänge zwischen den verschiedenen Datenwelten zu klären und Hierarchieebenen bzw. eine Priorisierung für die verschiedenen Informationstypen und Darstellungsformen einzuführen.

Zusätzlich kann das Zusammenspiel zwischen gesteigerter Erfahrung und Ergebnisqualität betrachtet werden. International zeigt sich ein klares Bild für den positiven Zusammenhang zwischen Fallzahlsteigerung und Ergebnisqualität, doch variiert dies in Stärke und Signifikanz je nach Leistungsbereich [6, 55–58]. Im deutschen Kontext kann das Mengen-Qualität Zusammenspiel für Notfall- und planbare Leistungsbereiche untersucht werden. Die Erkenntnisse können insbesondere für die Ausgestaltung und Umsetzung von Mindestmengen (z.B. Identifikation von Schwellwerten) Verwendung finden. Zusätzliche Evidenz für die Qualitätssteigerungspotentiale von Mindestmengen scheinen vor dem Hintergrund der Aushöhlung bestehender Mindestmengenregelungen angebracht [16, 17]. Diese Forschungsrichtung kann um einen internationalen Vergleich zur Anwendung und Ausgestaltung von Mindestmengen erweitert werden.

Letztlich kann die Beziehungen zwischen Behandlungsqualität und dem Verhältnis von Belegärzten zu angestellten Ärzten untersucht werden. Das Belegarztsystem ist in Deutschland weitverbreitet und mögliche Qualitätsunterschiede zwischen den zwei Formen der Arztintegration in das Krankenhaus wurden in Deutschland bisher nicht untersucht. Die Frage ob Krankenhäuser und Fachbereiche mit höherer oder niedriger Belegarztquote bessere Versorgungsqualität leisten hat für die Versorgungsplanung und Regulierung des Belegarztsystems, insbesondere auch im ländlichen Raum, hohe Relevanz. Scott et al. (2017) untersuchen diese Beziehung für häufige Krankheiten und haben keinen signifikanten Einfluss identifiziert [59]. Eine ähnliche Studie könnte auf Basis der in der Projektdatenbank enthaltenen Variablen für Leistungsbereiche wie Orthopädie oder Kardiologie durchgeführt werden.

Durch die großzügige finanzielle Unterstützung des Forschungsprojektes „Qualitätstransparenz in Deutschland und deren Anwendung zur Steigerung der Behandlungsqualität im Krankenhaus“ durch die Deutsche Forschungsgemeinschaft kann das Forschungsprojekt und die hier skizzierte Forschungsagenda in den nächsten drei Jahren weiterverfolgt werden.

## References


1. Pross C, Averdunk L-H, Stjepanovic J, Busse R, Geissler A. Health care public reporting utilization: User clusters, web trails, and usage barriers on Germany's public reporting portal Weisse Liste.de. BMC Medical Informatics and Decision Making (in press). 2017.
2. Pross C, Geissler A, Busse R. Measuring, Reporting, and Rewarding Quality of Care in 5 Nations: 5 Policy Levers to Enhance Hospital Quality Accountability. Milbank Q. 2017;95:136–83. doi:10.1111/1468-0009.12248.
3. Thielscher C, Antoni B, Driedger J, Jacobi S, Krol B. Geringe Korrelation von Krankenhausführern kann zu verwirrenden Ergebnissen führen. Gesundheitsökonomie & Qualitätsmanagement. 2014;19:65–9. doi:10.1055/s-0033-1335362.
4. BT-Drs. 18/5372. Drucksache des Deutschen Bundestages 18/5372 vom 30. Juni 2015:: Entwurf eines Gesetzes zur Reform der Strukturen der Krankenhausversorgung (Krankenhausstrukturgesetz – KHSG). Bundestagdrucksache. 2015.
5. Pross C, Busse R, Geissler A. Hospital choice matters: A time-trend and cross-section analysis of outcome quality performance and variation in German hospitals from 2006-2014. Health Policy (under review). 2017.
6. Birkmeyer JD, Siewers AE, Finlayson, Emily V A, Stukel TA, Lucas FL, Batista I, et al. Hospital volume and surgical mortality in the United States. The New England Journal of Medicine. 2002;346:1128–37. doi:10.1056/NEJMsa012337.
7. Hannan EL, Wu C, Walford G, King SB3, Holmes DR, JR, Ambrose JA, et al. Volume-outcome relationships for percutaneous coronary interventions in the stent era. Circulation. 2005;112:1171–9. doi:10.1161/CIRCULATIONAHA.104.528455.
8. Hillner BE, Smith TJ, Desch CE. Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care. J Clin Oncol. 2000;18:2327–40. doi:10.1200/jco.2000.18.11.2327.
9. AQUA. Generelle Methoden 3.0. 2013. https://www.sqg.de/sqg/upload/CONTENT/Hintergrund/Methodenpapier/AQUA_AllgemeineMethoden_Version_3-0.pdf.
10. IQTIG. Qualitätsreport 2015. 2016. https://www.iqtig.org/ergebnisse/qualitaetsreport/. Accessed 17 Oct 2016.
11. GKV Spitzenverband. Qualitätsorientierte Versorgungssteuerung und Vergütung. 2014. https://www.gkv-spitzenverband.de/krankenversicherung/qualitaetssicherung_2/qualitaetssicherung_2.jsp. Accessed 23 Dec 2016.
12. Schreyögg J, Bäuml M, Krämer J, Dette T, Busse R, Geissler A. Forschungsauftrag zur Mengenentwicklung nach § 17b Abs. 9 KHG: Endbericht; 2014.
13. Busse R, Ganten D, Huster S, Reinhardt E, Sutturp N, Wiesing U. Zum Verhältnis von Medizin und Ökonomie im deutschen Gesundheitssystem. Halle: Nationale Akademie der Wissenschaften; 2016.
14. Ebsen I. Perspektiven der Krankenhausplanung in einem gewandelten Markt und einem förderalen Gefüge. In: Schellschmidt H, Robra B-P, Klauber J, editors. Krankenhaus-Report 2006: Schwerpunkt: Krankenhausmarkt im Umbruch. Stuttgart, New York: Schattauer; 2007. p. 117–131.
15. DPA. Krankenhausgesellschaft: Planung muss Ländersache bleiben. 2016. http://aok-bv.de/presse/dpa-ticker/index_17411.html. Accessed 2 Nov 2016.
16. Nimptsch U, Peschke D, Mansky T. Mindestmengen und Krankenhaussterblichkeit - Beobachtungsstudie mit deutschlandweiten Krankenhausabrechnungsdaten von 2006 bis 2013. Gesundheitswesen 2016. doi:10.1055/s-0042-100731.

17. Cruppe W de, Malik M, Geraedts M. Minimum volume standards in German hospitals: do they get along with procedure centralization? A retrospective longitudinal data analysis. BMC Health Serv Res. 2015;15:279. doi:10.1186/s12913-015-0944-7.
18. KMA. Bundesärztekammer kritisiert IQTIG-Qualitätsindikatoren. KMA. 19.08.2016.
19. DNVF. Stellungnahme des Deutschen Netzwerk Versorgungsforschung (DNVF) e.V. zum Vorbericht Planungsrelevante Qualitätsindikatoren des Institutes für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) in der Fassung vom 18.7.2016. 2016. http://www.netzwerk-versorgungsforschung.de/uploads/Stellungnahmen/DNVF-Stellungnahme_IQTIG_PlanQI.pdf. Accessed 19 Dec 2016.
20. IQTIG. Planungsrelevante Qualitätsindikatoren. Abschlussbericht zur Auswahl und Umsetzung: Stand 31. August 2016. Berlin; 2016.
21. Beerheide R, Osterloh F. Länder kritisieren G-BA-Richtlinie. Deutsches Ärzteblatt. 2016;113:2353.
22. AOK Hessen. Aktuelle Vertragsabsichten nach § 140 SGB V durch die AOK Hessen; 2017.
23. AOK Hessen. Planung von Selektivverträgen nach § 140 SGB V: Alloplastischer Gelenkersatz - Hüftendoprothetik -; 2016.
24. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health. 2003;24:413–33. doi:10.1146/annurev.publhealth.24.100901.141005.
25. Hibbard JH, Peters E, Dixon A, Tusler M. Consumer competencies and the use of comparative quality information: it isn't just about literacy. Med Care Res Rev. 2007;64:379–94. doi:10.1177/1077558707301630.
26. Geraedts M, Auras S, Hermeling P, de Cruppé W. Abschlussbericht zum Forschungsauftrag zur Verbesserung der gesetzlichen Qualitätsberichte auf der Basis einer Krankenhaus-, Patienten- und Einweiserbefragung. Witten; 2010.
27. Pross C, Berger E, Geissler A, Martin S, Reinhard B. Stroke Units, Certification, and Stroke Outcomes in German Hospitals: A Fixed Effect Model for Stroke Mortality from 2006 - 2014. Health Economics (submitted). 2017.
28. Fassbender K, Balucani C, Walter S, Levine SR, Haass A, Grotta J. Streamlining of prehospital stroke management: The golden hour. The Lancet Neurology. 2013;12:585–96. doi:10.1016/S1474-4422(13)70100-5.
29. Pross C, Strumann C, Geissler A, Herwartz H, Klein N. Quality and Efficiency in Hospital Service Provision: A Geoadditive Stochastic Frontier Analysis of Stroke Quality of Care in Germany. Journal of Health Economics (submitted);2017.
30. Proudlove NC. Can good bed management solve the overcrowding in accident and emergency departments? Emergency Medicine Journal. 2003;20:149–55. doi:10.1136/emj.20.2.149.
31. Liu J, Griesman J, Nisenbaum R, Bell CM. Quality of care of hospitalized internal medicine patients bedspaced to non-internal medicine inpatient units. PLoS One. 2014;9:e106763. doi:10.1371/journal.pone.0106763.
32. Schrodi S, Tillack A, Niedostatek A, Werner C, Schubert-Fritschle G, Engel J. No Survival Benefit for Patients with Treatment in Certified Breast Centers-A Population-based Evaluation of German Cancer Registry Data. Breast J. 2015;21:490–500. doi:10.1111/tbj.12444.
33. Brettschneider C, Luhmann D, Raspe H. Informative value of Patient Reported Outcomes (PRO) in Health Technology Assessment (HTA). GMS Health Technol Assess. 2011;7:Doc01. doi:10.3205/hta000092.
34. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: A new era in clinical research. Perspect Clin Res. 2011;2:137–44. doi:10.4103/2229-3485.86879.
35. Donabedian A. Evaluating the quality of medical care. Milbank Q. 2005;83:691–729. doi:10.1111/j.1468-0009.2005.00397.x.

36. Gutacker N, Siciliani L, Moscelli G, Gravelle H. Choice of hospital: Which type of quality matters? J Health Econ. 2016;50:230–46. doi:10.1016/j.jhealeco.2016.08.001.
37. Porter ME, Larsson S, Lee TH. Standardizing Patient Outcomes Measurement. The New England Journal of Medicine. 2016;374:504–6. doi:10.1056/NEJMp1511701.
38. HELIOS. Qualitätsergebnisse. 2017. http://www.helios-kliniken.de/medizin/qualitaetsmanagement/qualitaetsergebnisse.html. Accessed 18 Mar 2017.
39. Lamprecht P. Punktwertung für Kliniken. Welt am Sonntag. 16.10.2005.
40. Milstein R, Schreyoegg J. Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. Health Policy. 2016;120:1125–40. doi:10.1016/j.healthpol.2016.08.009.
41. Damberg C, Sorbero M, Lovejoy S, Martsolf G, Raaen L, Mandel D. Measuring Success in Health Care Value-Based Purchasing Programs: Findings from an Environmental Scan, Literature Review, and Expert Panel Discussions; 2014.
42. Yegian JM, Dardess P, Shannon M, Carman KL. Engaged patients will need comparative physician-level quality data and information about their out-of-pocket costs. Health Aff (Millwood). 2013;32:328–37. doi:10.1377/hlthaff.2012.1077.
43. Dixon A, Robertson R, Appleby J, Burge P, Devlin NJ. Patient choice: How patients choose and how providers respond. London; 2010.
44. AHRQ. The Challenges of Measuring Physician Quality. 2015. https://www.ahrq.gov/professionals/quality-patient-safety/talkingquality/create/physician/challenges.html. Accessed 8 Feb 2017.
45. G-BA. Regelungen zum Qualitätsbericht 2012; 2013.
46. AQUA. Bericht zur Datenvalidierung 2013. 2014.
47. AQUA. Qualitätsreport 2013. Göttingen: AQUA; 2014.
48. Chandra A, Finkelstein A, Sacarny A, Syverson C. Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector. Am Econ Rev. 2016;106:2110–44. doi:10.1257/aer.20151080.
49. Pope DG. Reacting to rankings: evidence from "America's Best Hospitals". J Health Econ. 2009;28:1154–65. doi:10.1016/j.jhealeco.2009.08.006.
50. Propper C, Burgess S, Gossage D. Competition and Quality: Evidence from the NHS Internal Market 1991-9*. The Economic Journal. 2008;118:138–70. doi:10.1111/j.1468-0297.2007.02107.x.
51. Chou S-Y, Deily ME, Li S, Lu Y. Competition and the impact of online hospital report cards. J Health Econ. 2014;34:42–58. doi:10.1016/j.jhealeco.2013.12.004.
52. Gaynor M, Moreno-Serra R, Propper C. Can competition improve outcomes in UK health care? Lessons from the past two decades. J Health Serv Res Policy. 2012;17 Suppl 1:49–54. doi:10.1258/jhsrp.2011.011019.
53. Moscelli G, Gravelle H, Siciliani L. Market Structure, Patient Choice and Hospital Quality for Elective Patients. 2016. https://www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP139_market_structure_patient_choice_hospital_quality.pdf. Accessed 12 Feb 2017.
54. WIdO. QSR Leistungsbereiche. 2017. http://www.qualitaetssicherung-mit-routinedaten.de/methoden/bereiche/index.html. Accessed 10 Feb 2017.
55. Lee KCL, Sethuraman K, Yong J. On the Hospital Volume and Outcome Relationship: Does Specialization Matter More Than Volume? Health Serv Res. 2015;50:2019–36. doi:10.1111/1475-6773.12302.
56. Lin X, Tao H, Cai M, Liao A, Cheng Z, Lin H. A Systematic Review and Meta-Analysis of the Relationship Between Hospital Volume and the Outcomes of Percutaneous Coronary Intervention. Medicine (Baltimore). 2016;95:e2687. doi:10.1097/MD.0000000000002687.

57. Wouters MW, Siesling S, Jansen-Landheer ML, Elferink, M A G, Belderbos J, Coebergh JW, Schramel, F M N H. Variation in treatment and outcome in patients with non-small cell lung cancer by region, hospital type and volume in the Netherlands. European Journal of Surgical Oncology. 2010;36 Suppl 1:S83-92. doi:10.1016/j.ejso.2010.06.020.
58. Urbach DR, Baxter NN. Does it matter what a hospital is "high volume" for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. Qual Saf Health Care. 2004;13:379–83. doi:10.1136/qhc.13.5.379.
59. Scott KW, Orav EJ, Cutler DM, Jha AK. Changes in Hospital-Physician Affiliations in U.S. Hospitals and Their Effect on Quality of Care. Ann Intern Med. 2017;166:1–8. doi:10.7326/M16-0125.