

# Entwicklung und Anwendung eines Composite Measures für die Qualitätsdarstellung von Krankenhäusern

Machbarkeitsanalyse

# Entwicklung und Anwendung eines Composite Measures für die Qualitätsdarstellung von Krankenhäusern

## Machbarkeitsanalyse

### **Verantwortlich**

Uwe Schwenk  
Director Programm Gesundheit  
Bertelsmann Stiftung  
Telefon +49 5241 81-81418  
uwe.schwenk@bertelsmann-stiftung.de  
www.bertelsmann-stiftung.de

### **Kontakt**

Johannes Strotbek, Hannah Wehling  
Senior Project Manager  
Telefon +49 30 275788-320, -326  
johannes.strotbek@weisse-liste.de  
hannah.wehling@weisse-liste.de

### **Verfasser**

Prof. Dr. Martin Emmert, MHMM (Universität Bayreuth) [1]  
Prof. Dr. Florian Meier (Wilhelm Löhe Hochschule Fürth) [2]  
Juli 2023

Titelbild: Adobe Stock #569593464s

# Inhalt

<b>1</b>	<b>Hintergrund für die Entwicklung eines Composite Measures .....</b>	<b>4</b>
<b>2</b>	<b>Methodik .....</b>	<b>5</b>
<b>3</b>	<b>Ergebnisse .....</b>	<b>9</b>
3.1	Definition eines Qualitätsmodells und Festlegung der Qualitätskategorien .....	9
3.2	Auswahl von relevanten Kennzahlen für die festgelegten Qualitätskategorien .....	13
3.3	Berechnung von Zwischenergebnissen für die festgelegten Qualitätskategorien .....	15
3.4	Standardisierung der Zwischenergebnisse (z-Transformation) .....	16
3.5	Begrenzung negativer Ausreißerwerte .....	17
3.6	Berechnung eines ungewichteten Composite Measures .....	18
3.7	Zuordnung zu den Sternengruppen .....	19
3.8	Setzen von Filterkriterien für Top-KH führt zu Empfehlung .....	20
3.9	Setzen von Filterkriterien für Low-Performer-KH führt zu Warnung .....	21
<b>4</b>	<b>Diskussion der Ergebnisse .....</b>	<b>22</b>
<b>5</b>	<b>Zusammenfassung und Ausblick .....</b>	<b>26</b>
<b>6</b>	<b>Literaturverzeichnis .....</b>	<b>27</b>

# 1 Hintergrund für die Entwicklung eines Composite Measures

Transparenzinitiativen (Public Reporting Initiativen) berichten über die Qualität von medizinischen Leistungserbringern (z.B. Krankenhäuser, niedergelassene Ärzte). Die Initiativen ziehen dabei unterschiedliche Qualitätsdaten heran und bereiten diese so auf, dass Nutzer hierauf basierend eine Entscheidung für oder gegen bestimmte Leistungserbringer treffen können.[1] Welche Qualitätsinformationen dies im Einzelnen sind, hängt unter anderem an der Intention und inhaltlichen Ausrichtung des Anbieters der Transparenzinitiative ab. Untersuchungen haben diesbezüglich gezeigt, dass sich vorhandene Onlineportale auf verschiedene Qualitätsaspekte fokussieren bzw. diese abdecken. So zeigten mehrere Studien, dass es Unterschiede bei der Einbindung von Aspekten der Struktur-, Prozess- und Ergebnisqualität sowie in der Darstellung von Ergebnissen von Patientenbefragungen und weiterer (Qualitäts-)Aspekte gibt.[1–5] Ebenso unterscheiden sich Transparenzinitiativen hinsichtlich der Anzahl der gezeigten Qualitätsinformationen. Anfänglich ging der Trend von Public Reporting Bemühungen in Richtung detaillierter Präsentation vieler unterschiedlicher Informationen. Mittlerweile konnte jedoch beobachtet werden, dass sich einige Transparenzinitiativen im internationalen Kontext darum bemühen, so genannte Composite Measures zu präsentieren.[6–8]

Im Gegensatz zur Darstellung vieler einzelner Qualitätsinformationen werden bei einem Composite Measure zwar auch viele Informationen berücksichtigt, diese jedoch zu einer einzigen Kennzahl zusammengefasst und aggregiert dargestellt. Schwartz und Kollegen schreiben hierzu: „Composite measures of health care provider performance aggregate individual performance measures into an overall score, thus providing a useful summary of performance.“[9] Der nützliche Aspekt des Composite Measures soll dadurch entstehen, dass durch die reduzierte Menge an bereitgestellten Informationen die Komplexität der Entscheidungsfindung reduziert wird.[7] Werner und Kollegen führen diesbezüglich an: „[...] patient-oriented report cards are increasingly moving toward using summary measures to display providers' overall quality information, which combine multiple measures into one or more summary measures.“[7] Des Weiteren konnte gezeigt werden, dass Composite Measures im Gegensatz zu einzelnen Indikatoren besser dazu geeignet sind, künftige Qualitätsergebnisse sowie schlecht und gut abschneidende Leistungserbringer vorherzusagen.[10] Erfahrungen aus den USA haben zudem gezeigt, dass sich das Wahlverhalten von Konsumenten verbessert [7]. Durch die reduzierte Menge an präsentierten Informationen werden signifikant häufiger exzellente und weniger häufig schlecht abschneidende Leistungserbringer gewählt („The nursing home star rating system significantly affected consumer demand for high- and low-rated nursing homes“[7]).

Im Verlauf der letzten Jahre haben sich einige Transparenzinitiativen im internationalen Kontext der Report Cards angenommen und entsprechende Composite Measures entwickelt und veröffentlicht. Zu den prominentesten Beispielen hierbei zählen:

- Hospital Compare
- Nursing Home Compare
- NHS Choices
- US News Best Hospital
- Home Health star rating system

## 2 Methodik

Im Folgenden wird die Methodik zur Entwicklung und Anwendung eines Composite Measures für die Qualitätsbewertung von Krankenhäusern in Deutschland beschrieben (Machbarkeitsanalyse). Die Methodik wurde mit dem Ziel entwickelt, die Qualität von Krankenhäusern in Deutschland übersichtlich und in einem finalen Composite Measure Wert darzustellen. Es ist angedacht, dass sie als neues Bewertungsverfahren im Rahmen der Weißen Liste zum Einsatz kommt und die bisher verwendete Systematik ergänzt bzw. ersetzt. In Absprache mit Vertretern der Weisse Liste wurde die Methodik anhand der oben erwähnten Leistungsbereiche (1) Hüftendoprothesenversorgung sowie (2) Mammachirurgie entwickelt und veranschaulicht. Verwendet wurden dabei (insbesondere) die öffentlich zugänglichen Daten der externen stationären Qualitätssicherung (im Folgenden Gesetzliche Qualitätssicherung, Gesetzliche QS).

Es sollte eine Methodik entwickelt werden, die es ermöglicht, neben den beiden Fallbeispielen auch weitere Leistungsbereiche, die beispielsweise im Rahmen der Gesetzlichen QS betrachtet werden, zu analysieren. Zum anderen sollte die Vorgehensweise so flexibel sein, dass neben den aktuell vorliegenden Daten der Gesetzlichen QS auch weitere Qualitätsdaten eingebunden werden können. Somit sollte die Methodik generell sein, um möglichst viele medizinische Leistungsbereiche berücksichtigen zu können, aber auch flexibel genug, um weitere Kennzahlen zu integrieren. Insgesamt wurden 9 Schritte durchgeführt, um einen spezifischen Composite Measure für jedes Krankenhaus zu ermitteln, die in der folgenden 1 dargestellt und anschließend sukzessive erläutert werden. Darüber hinaus soll eine Empfehlung oder eine Warnung für einzelne Krankenhäuser ausgesprochen werden können, sofern weitere definierte Kriterien erfüllt werden bzw. vorliegen.

<p>Modelldefinition:</p> <ol style="list-style-type: none"> <li>1. Definition eines Qualitätsmodells und Festlegung der Qualitätskategorien</li> <li>2. Auswahl relevanter Kennzahlen für die festgelegten Qualitätskategorien</li> </ol> <p>Quantitativer Teil – Berechnung eines Composite Measures:</p> <ol style="list-style-type: none"> <li>3. Berechnung von Zwischenergebnissen für die festgelegten Qualitätskategorien</li> <li>4. Standardisierung der Zwischenergebnisse (Z-Transformation)</li> <li>5. Begrenzung negativer Ausreiserwerte</li> <li>6. Berechnung eines ungewichteten Composite Measure (Gesamtsumme der z-Werte dividiert durch die Anzahl an Kategorien) *</li> <li>7. Zuordnung von Kliniken zu drei Sternengruppen</li> </ol> <p>Qualitativer Teil – Identifizierung besonderer Krankenhäuser:</p> <ol style="list-style-type: none"> <li>8. Setzen von Filterkriterien für Top-Performer führt zu Empfehlung</li> <li>9. Setzen von Filterkriterien für Low-Performer führt zu Warnung</li> </ol> <p>* Keine Berücksichtigung von Kliniken mit fehlendem Ergebniswert bei einer der Qualitätskategorien</p>
---

Abbildung 1: Vorgehensweise für die Entwicklung und Anwendung eines Composite Measures

Nachfolgend werden die 9 Schritte allgemein beschrieben. Eine ausführliche Erläuterung bzw. Darstellung wird anhand der beiden Fallbeispiele in Kapitel 3 vorgenommen.

## Modelldefinition

### **Schritt 1.** Definition eines Qualitätsmodells und Festlegung der Qualitätskategorien

Zu Beginn muss ein Qualitätsmodell definiert werden, das für die weiteren Schritte zur Ermittlung des Composite Measures grundlegend ist. Der Composite Measure wird final für jedes Krankenhaus ermittelt und eine reduzierte bzw. komprimierte Information für die Nutzer darstellen. Demnach ist für den Nutzer auf den ersten Blick unklar, was sich genau hinter der Kennzahl im Detail verbirgt. Vor diesem Hintergrund muss darauf geachtet werden, dass Vertrauen zu der Kennzahl aufgebaut wird, indem klar und transparent kommuniziert wird, welche „Werte“ sich im Composite Measure widerspiegeln und wie diese verarbeitet werden. Bereits zu Beginn der Entwicklung ist daher eine klare Festlegung notwendig, was sich im Composite Measure schlussendlich widerspiegeln soll. Im Grunde geht es hierbei um die Frage, was unter Qualität verstanden wird, da der Composite Measure dem Nutzer einen Eindruck von der Qualität einer medizinischen Einrichtung für die Durchführung einer bestimmten Operation bzw. Leistung vermitteln soll.

### **Schritt 2.** Auswahl relevanter Kennzahlen für die Qualitätskategorien

Nach der Festlegung der Qualitätskategorien müssen diese durch relevante Kennzahlen operationalisiert werden. D.h. es müssen bestimmte Kennzahlen herangezogen werden, um die Qualität in den jeweiligen Kategorien zu messen. Im Folgenden wird daher beschrieben, welche Indikatoren bzw. Kennzahlen zur Messung der einzelnen Qualitätsaspekte herangezogen werden.

## Quantitativer Teil – Berechnung eines Composite Measures

### **Schritt 3.** Berechnung von Zwischenergebnissen für die festgelegten Qualitätskategorien

Für die einzelnen Qualitätskategorien wird ein numerischer Wert (Zwischenergebnis) ermittelt. Die in dem beschriebenen Qualitätsmodell vorliegenden Daten ermöglichen es, für jede Qualitätskategorie ein solches Zwischenergebnis zu ermitteln. Hierfür sind in Abhängigkeit von der jeweiligen Qualitätskategorie unterschiedliche Schritte und Daten notwendig.

### **Schritt 4.** Standardisierung der Zwischenergebnisse (z-Transformation)

Im nächsten Schritt wird eine bessere Vergleichbarkeit der Zwischenergebnisse für die festgelegten Qualitätskategorien angestrebt. Damit sollen die Daten, die unterschiedliche Einheiten aufweisen, auf eine einheitliche Skala

transformiert werden. Dieser Schritt ist insbesondere deshalb sinnvoll, da die entwickelte Methodik auch für andere Anwendungsgebiete bzw. Leistungsbereiche anwendbar sein soll. Durch die Standardisierung wird es möglich, auch Daten aus anderen Bereichen (mit unterschiedlichen Einheiten) in das Modell zu integrieren. Bei der z-Transformation wird jeder Ausgangswert in einen neuen, standardisierten Wert überführt. Dabei werden Mittelwert und Standardabweichung berücksichtigt. Der durch die z-Transformation standardisierte Wert ist nun mit anderen standardisierten Werten vergleichbar.

#### **Schritt 5.** Begrenzung negativer Ausreiserwerte

In Anlehnung an Hospital Compare [11] werden negative Ausreiser eliminiert. Dieser Schritt wird vollzogen, um potenziell falsche Angaben bei der Berichtserstellung („potentially inaccurate reporting“ [11]) zu berücksichtigen. Im Gegensatz zu CMS Hospital Compare werden allerdings nicht alle negativen und positiven Extremwerte auf -3 / +3 begrenzt (0.125th percentile (Z= -3), 99.875 percentile (Z=3)), sondern es werden im Basismodell der Auswertungen die negativen Werte, die sich im unteren 1%-Perzentil befinden, begrenzt und auf diesen Wert gesetzt.

#### **Schritt 6.** Berechnung eines ungewichteten Composite Measure

Im nächsten Schritt erfolgt die Kalkulation eines ungewichteten Composite Measures; dieser ergibt sich aus der Gesamtsumme der z-Werte der Qualitätskategorien dividiert durch die Anzahl an Kategorien. Dieser Gesamtwert ist insofern ungewichtet, als dass nun alle Qualitätskategorien mit dem gleichen Gewicht in den Endwert eingehen. Hierbei muss angemerkt werden, dass kein Composite Measure für ein Krankenhaus ermittelt wird, wenn bei einer oder bei mehreren Qualitätskategorie(n) ein fehlender Wert vorliegt (d.h. eine Qualitätskategorie weist keine Ergebnisse auf; z.B. wegen fehlender oder falscher Berichterstattung). Der Gesamtwert (d.h. der ungewichtete Composite Measure) wird nur berechnet, sofern für jede Qualitätskategorie ein z-Wert vorliegt.

#### **Schritt 7.** Zuordnung von Kliniken zu drei Sternengruppen

Im nachfolgenden Schritt wird eine Zuordnung der Kliniken zu drei Gruppen mit unterschiedlichem Qualitätsniveau je nach Ausprägung des Gesamtwertes vorgenommen. Die drei Gruppen werden in Form von Sternen dargestellt. Ein Stern entspricht geringer, zwei Sterne mittlerer und drei Sterne hoher Qualität. Wie in der Literatur dargestellt [11], existieren für die Zuordnung der Kliniken zu den drei Gruppen unterschiedliche Möglichkeiten („categorizing hospitals by percentile, setting statistical significance cutoffs, and using a clustering algorithm.“ [11]). Die Zuordnung erfolgt über den berechneten Composite Measure. Ein Krankenhaus erhält einen Stern, sofern der Composite Measure mehr als 0,5 der Standardabweichung um den Mittelwert nach unten abweicht. Sollte der Composite Measure mehr als 0,5 um den Mittelwert nach oben abweichen, erhält das Krankenhaus hingegen drei Sterne. Zwei Sterne erhält ein Krankenhaus, sofern der Composite Measure des Krankenhauses zwischen den beiden Grenzen liegt (Abbildung 2).

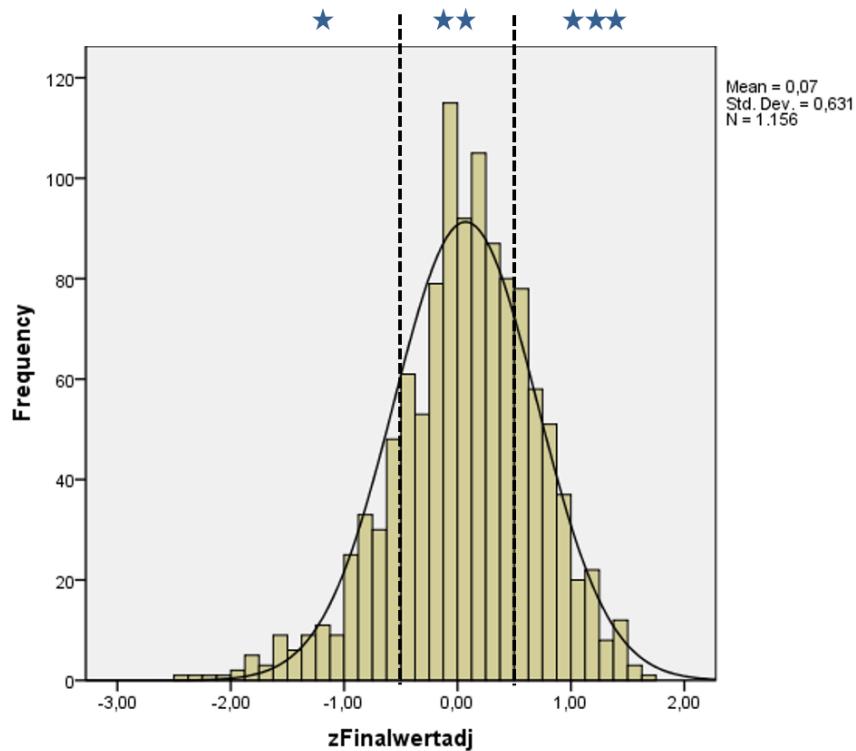


Abbildung 2: Einteilung der Kliniken in die drei Gruppen

## Qualitativer Teil

### Schritt 8. Setzen von Filterkriterien für Top-Performer-Kliniken [Empfehlung]

Im nächsten Schritt werden weitere Elemente in die Betrachtung aufgenommen. Wie oben beschrieben, werden besonders gute Krankenhäuser mit einem hohen Composite Measure Ergebnis zu der Gruppe mit drei Sternen zugeordnet. Diese haben in den einzelnen Qualitätskategorien einen überdurchschnittlichen Wert und damit einen besonders hohen Composite Measure. Sie befinden sich oberhalb der Grenze mit  $+0,5$  Standardabweichungen vom Mittelwert. In diesem Schritt können nun auch weitere Elemente in der Bewertung berücksichtigt werden. Erfüllen die Krankenhäuser bestimmte Anforderungen (je nach Leistungsbereich können unterschiedliche Anforderungen existieren), erhalten sie zusätzlich eine besondere Empfehlung. Durch die Empfehlung der Top-Performer sollen den Nutzern auf den ersten Blick exzellente Spitzenkliniken aufgezeigt werden, die eine herausragende Versorgungsqualität in allen Kategorien leisten.

Im Bereich der Hüftendoprothesenversorgung wird beispielsweise vorausgesetzt, dass ein Krankenhaus (1) 3 Sterne hat, (2) zu den Top 10% Performern beim Gesamt-z-Wert gehört, (3) eine vollständig unauffällige Versorgungsqualität für die externe Qualitätssicherung (d.h. eine Quote 100%) aufweist und (4) oberhalb des Medians bei den Fallzahlen liegen muss. Durch diesen Schritt ist es möglich, beispielsweise auch die Fallzahlen der Kliniken im Rahmen der Bewertung zu berücksichtigen. Fallzahlen sind zwar ein metrischer Wert und könnten vom statistischen Procedere her auch im quantitativen Teil Berücksichtigung finden, jedoch gibt es bei dem Aspekt der



Fallzahl einen inhaltlichen Ausschluss. Fallzahlen stellen an sich kein Qualitätskriterium dar, können aber zumindest Rückschluss auf die Erfahrung einer Klinik mit einem Eingriff (repräsentiert durch die Fallzahl) geben, die wiederum positive Qualitätseffekte haben können [12,13]. Bislang wurde jedoch kein linearer Zusammenhang zwischen Qualität bzw. Erfahrung und der Fallzahl nachgewiesen. Daher kann die Fallzahl nicht als primäres Qualitätskriterium, aber als qualitative Größe berücksichtigt werden.

### **Schritt 9.** Setzen von Filterkriterien für Low-Performer-Kliniken [Warnung]

Im neunten und letzten Schritt werden weitere Kriterien hinzugezogen, um vor besonders auffälligen Krankenhäusern zu „warnen“. Diese haben in den Qualitätskategorien einen unterdurchschnittlichen Wert und damit einen besonders niedrigen Composite Measure. Sie befinden sich unterhalb der Grenze mit 0,5 Standardabweichungen vom Mittelwert.

Im Bereich der Hüftendoprothesenversorgung wird beispielsweise vorausgesetzt, dass ein Krankenhaus (1) 1 Stern hat, (2) zu den Low 10% Performern beim Gesamt-z-Wert gehört, (3) keine vollständig unauffällige Versorgungsqualität für die externe Qualitätssicherung (d.h. eine Quote <100%) aufweist und (4) unterhalb des Medians bei den Fallzahlen liegen muss. Durch das Aussprechen einer „Warnung“ sollen den Nutzern auf den ersten Blick Kliniken aufgezeigt werden, die eine von den Zahlen her geringe Versorgungsqualität aufweisen.

## **3 Ergebnisse**

### **3.1 Definition eines Qualitätsmodells und Festlegung der Qualitätskategorien**

Zu Beginn muss ein Qualitätsmodell definiert werden, das grundlegend für die weiteren Schritte zur Ermittlung eines Composite Measures ist. Dieses Qualitätsmodell soll übergreifend für alle Qualitätsdarstellungen (d.h. für alle Leistungsbereiche) Gültigkeit besitzen. Ziel des Ansatzes ist die Berechnung eines Composite Measures für jedes Krankenhaus, der alle relevanten Qualitätsinformationen so zusammenfasst, dass die Nutzer eine reduzierte bzw. komprimierte Information erhalten. Demnach soll für den Nutzer klar und transparent kommuniziert werden, welche „Werte“ sich im Composite Measure widerspiegeln und wie diese verarbeitet werden. Bereits zu Beginn der Entwicklung ist eine klare Festlegung notwendig, was sich im Composite Measure schlussendlich widerspiegeln soll. Im Grunde geht es hierbei um die Frage, was unter Qualität verstanden wird, da der Composite Measure dem Nutzer einen Eindruck von der Qualität einer medizinischen Einrichtung für die Durchführung einer bestimmten Operation bzw. Leistung vermitteln soll.

An dieser Stelle der Ausführungen soll auf die Ausführung vom 28.5.2019 verwiesen werden, in der ein erstes Qualitätsmodell, basierend auf der aktuellen Darstellung der Weissen Liste, erstellt wurde. Demnach wird unter Qualität unter Berücksichtigung der Konzepte von Donabedian [14–17] und weiterer Literatur [18,19] ein gleich gewichtetes Zusammenspiel der vier folgenden Elemente verstanden werden:

- Die *Qualitätsbeschreibung* nach *Donabedian* betrachtet strukturelle Voraussetzungen von Krankenhäusern, den Behandlungsprozess sowie die medizinischen Ergebnisse der Behandlung. Demnach wird hierbei in die strukturellen Voraussetzungen bzw. Gegebenheiten (Ausstattung) eines Krankenhauses sowie leistungsbereichsspezifische Qualitätsparameter (Prozess- und Ergebnisqualität) unterschieden.
- Die *Nutzererfahrung* umfasst die persönliche Einschätzung relevanter Personengruppen. Dies sind zu meist Patienten, können aber auch einweisende Ärzte, Angehörige oder andere Experten sein.
- Informationen zur *Patientensicherheit & Hygiene* umfassen allgemeine und besondere Maßnahmen, Maßnahmen beim Operieren sowie Maßnahmen zur Hygiene und Infektionsschutz.
- Die von Patienten selbst zu tragenden Kosten für die Behandlung, Unterbringung oder bestimmte medizinische Leistungen.

Derzeit gibt es im deutschen Gesundheitswesen allerdings keine ausreichend detaillierten Informationen, um ausgewogen über die vier Komponenten berichten zu können. Die folgende Abbildung 3 zeigt durch eine farbliche Hervorhebung an, welche Informationen aktuell verfügbar sind (die dunkelblauen Bereiche sind verfügbar) und damit Gegenstand der Berichterstattung sein können. Im Krankenhaussektor sind neben den Patientenerfahrungen auch Informationen zur medizinischen Qualität (Struktur-, Prozess- und Ergebnisqualität) sowie Patientensicherheit und Hygiene verfügbar, über die berichtet werden kann und im Folgenden berücksichtigt werden.

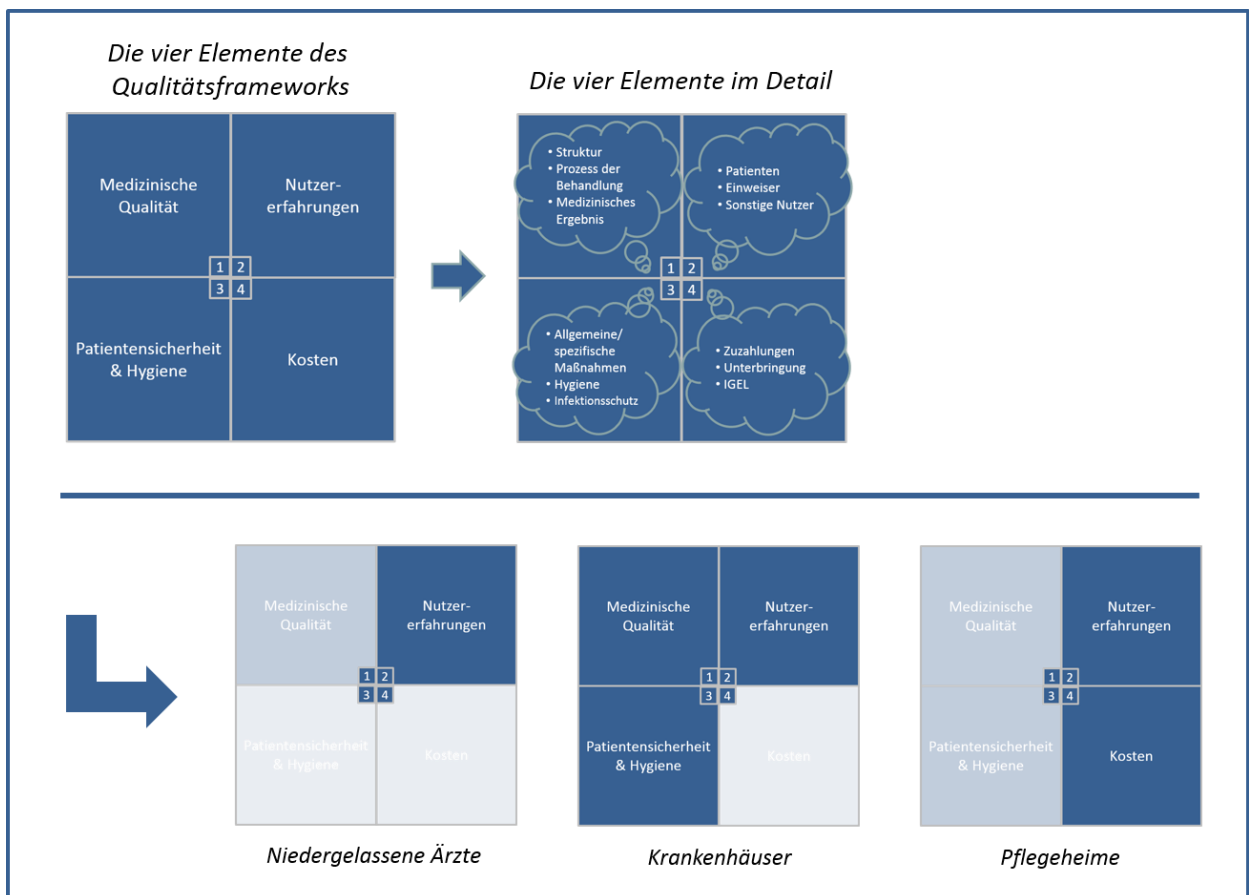


Abbildung 3: Das Qualitätsframework der Weissen Liste

### Exkurs: Zusammenhang zwischen Fallzahlen, Erfahrung und Qualität

Über die vier bisher gezeigten Bereiche soll im Qualitätsmodell darüber hinaus dem Aspekt der Erfahrung (Fallzahl) Rechnung getragen werden. Zahlreiche Studien haben gezeigt, dass für bestimmte Operationen eine mindestens durchschnittliche Fallzahl und damit gewisse Erfahrung vorliegen sollte, um ein medizinisch gutes Ergebnis zu erzielen [12,13,20,21]. Allerdings zeigen andere Studien einen eher heterogenen Evidenzstand zu diesem Zusammenhang [22,23]. Insgesamt zeigt die Literatur allerdings zumeist, dass insbesondere Leistungserbringer mit hohen Fallzahlen bessere Ergebnisse erzielen als Leistungserbringer mit eher geringe(re)n Fallzahlen. Somit scheint sich demnach eine gewisse Menge an Erfahrung positiv auf die Behandlungsergebnisse auszuwirken.

Es besteht weitestgehend Konsens darüber, dass die Fallzahl per se kein Qualitätskriterium ist, es aber durch die zunehmende Erfahrung einer Klinik mit einem Eingriff (repräsentiert durch die Fallzahl) zu positiven Qualitätseffekten kommen kann. Die folgende Abbildung 4 stellt einen Scatterplot zum Zusammenhang der Fallzahl für die Hüftendoprothesenversorgung bzw. Mammachirurgie und den Ergebnissen der externen Qualitätssicherung dar. Gut erkennbar ist, dass bei niedrigen Fallzahlen eine größere Bandbreite von Ergebnissen beobachtet werden kann, bei hohen Fallzahlen die Ergebnisse in beiden Leistungsbereichen hingegen tendenziell besser ausfallen (pearson correlation Hüft-TEP-OP: 0,125;  $p < 0.001$ ; Mammakarzinom: 0,175;  $p < 0.001$ ).

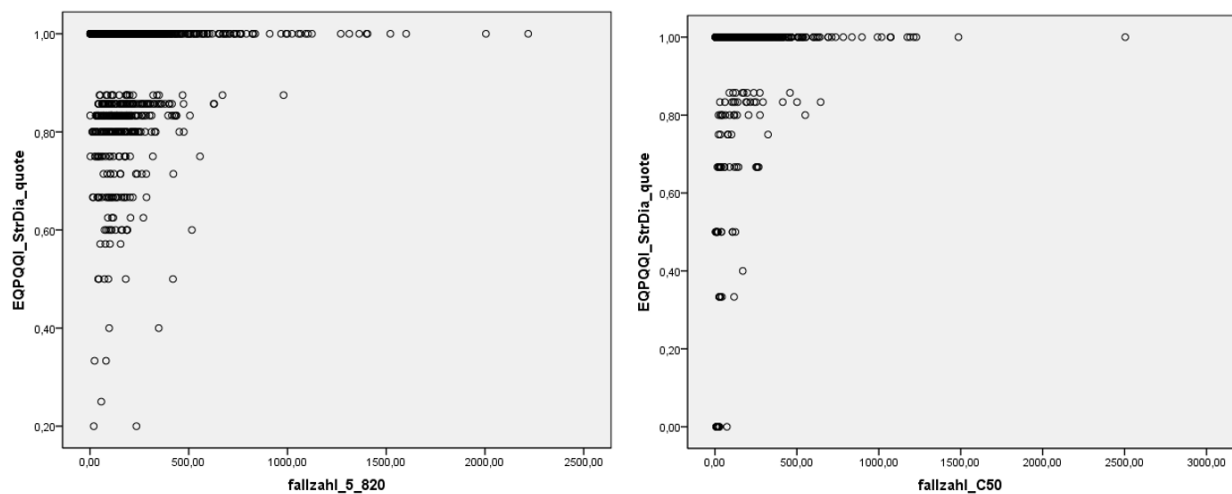


Abbildung 4: Scatterplot zum Zusammenhang der Fallzahl für die Hüftendoprothesenversorgung / Mammachirurgie und den Ergebnissen der externen Qualitätssicherung (N=1.175/606 Krankenhäuser)

Eine einwandfreie Qualität ist aufgrund des kleineren Konfidenzintervalls bei hohen Fallzahlen mit einer höheren Sicherheit zu bewerten als eine einwandfreie Qualität bei niedrigen Fallzahlen. Des Weiteren gibt es, wie oben angemerkt, keinen linearen Zusammenhang zwischen beiden Variablen, d. h. es kann nicht geschlussfolgert werden, dass ein Krankenhaus mit einer höheren Fallzahl immer auch immer eine bessere Qualität erbringt.

Inhaltlich gesehen kann die Qualität bei sehr hohen Fallzahlen auch wieder abnehmen. In Abbildung 5 wird durch ein Streudiagramm der Zusammenhang zwischen Fallzahl und Composite Measure dargestellt. Da der Composite Measure eine Kennzahl darstellen soll, der die Qualität eines Krankenhauses wiedergibt, kann aus der Grafik

Folgendes abgelesen werden: Mit der Zunahme der Fallzahl steigt auch die erbrachte Qualität eines Krankenhauses. Dass es sich nicht um einen linearen Zusammenhang handelt, ist erkennbar. Auch bei hohen Fallzahlen gibt es eine Streuung und die Werte befinden sich nicht auf einer Linie. Darüber hinaus zeigt der  $R^2$ -Wert, eine statistische Kenngröße, wie gut die Streuung durch zugrunde gelegte Kurven erklärt werden kann. Hier wurde einmal ein linearer Zusammenhang ( $R^2=0,186$ ) und einmal ein quadratischer ( $R^2=0,223$ ) angenommen. Da die Berechnung bei einem quadratischen Zusammenhang zu einem höheren  $R^2$ -Wert führt, ist diese Annahme besser geeignet den Zusammenhang zu erklären. Insgesamt kann zwischen Fallzahl und Qualität kein linearer Zusammenhang angenommen werden. Diesem Aspekt soll im Folgenden Rechnung getragen werden.

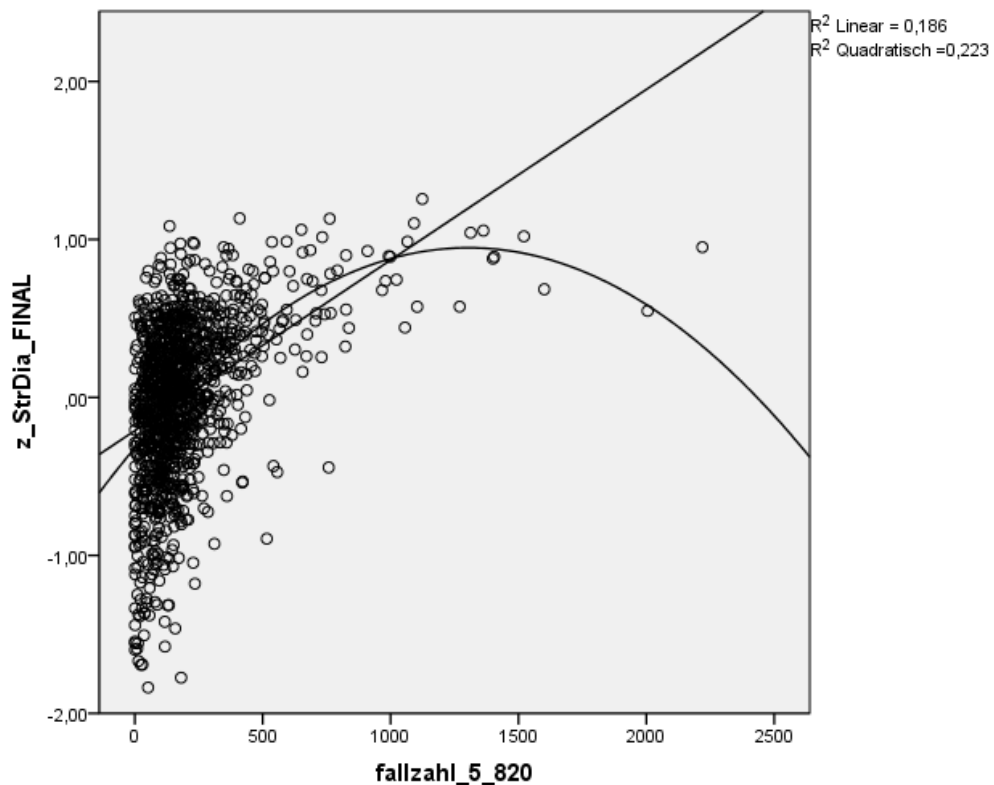


Abbildung 5: Darstellung des Zusammenhangs zwischen Fallzahl und Composite Measure

Beachtung soll ebenfalls die Tatsache finden, dass es keine Mindestmengengrenze gibt, die erreicht werden muss, um eine gute Qualität zu „garantieren“ (analog zur Mindestmengenregelung des G-BA [24]). Ebenso kann es aber auch dazu kommen (wie oben gesagt), dass trotz hoher Fallzahlen eine nicht vollständig unauffällige Qualität vorliegt. Die folgende Tabelle 1 gibt einen Überblick über die Fallzahlverteilung für die Hüftendoprothesenversorgung bzw. Mammachirurgie über die beteiligten Kliniken.

Deskriptive Statistik		Hüftendoprothesenversorgung	Mammachirurgie
Mean		195,01	102,90
Median		148,00	20,00
Std. Deviation		203,437	181,18
Range		2218,00	2503,00
Minimum		1,00	1,00
Maximum		2219,00	2504,00
Percentiles	25	84,00	3,00
	75	229,75	151,00

Tabelle 1: Deskriptive Statistik zur Fallzahl für die beiden Fallbeispiele 5-820 bzw. C50

Basierend auf den obigen Ausführungen soll im Rahmen der Ermittlung des Composite Measures zusammenfassend keine „einfache“ Berücksichtigung der Fallzahl als primäres Kriterium erfolgen, sondern ausschließlich im Zusammenhang mit der dokumentierten Versorgungsqualität anhand der Ergebnisse des strukturierten Dialogs der externen stationären Qualitätssicherung.

Das dem Composite Measure zugrunde liegende Qualitätsmodell umfasst neben der Qualitätsbeschreibung von Donabedian (Struktur-, Prozess- und Ergebniskriterien) und Informationen zur Patientensicherheit & Hygiene auch Patientenerfahrungen als primäre Qualitätskategorien. Mit sekundärer Bedeutung wird der Aspekt der Erfahrung (Fallzahl) berücksichtigt.

### 3.2 Auswahl von relevanten Kennzahlen für die festgelegten Qualitätskategorien

Nach der Definition des Qualitätsmodells müssen relevante Kennzahlen für die festgelten Qualitätskategorien benannt werden.

1. Für den Bereich medizinische Qualität und hier im Besonderen leistungsbereichsspezifische Qualitätsparameter (Prozess- und Ergebnisqualität) wird auf Kennzahlen zurückgegriffen, die im Rahmen der externen stationären Qualitätssicherung erhoben und veröffentlicht werden (Kennzahlen anderer Qualitätssicherungssysteme, wie beispielsweise das QSR-Verfahren der AOK, sind nicht für alle Krankenhäuser öffentlich verfügbar und finden daher an der Stelle keine Berücksichtigung). Die Kennzahlen müssen als veröffentlichungspflichtig eingestuft worden sein und einen definierten Referenzbereich aufweisen. Es wurden die Kennzahlen der Bundesauswertungen des Qualitätsreports 2019 herangezogen und die

relevanten Kennzahlen festgelegt. Demnach wurden für den Bereich der Hüftendoprothesenversorgung insgesamt 8 veröffentlichungspflichtige Indikatoren mit einem definierten Referenzbereich und für den Bereich der Mammachirurgie insgesamt 7 veröffentlichungspflichtige Indikatoren mit einem definierten Referenzbereich in diese Untersuchung eingeschlossen (Tabelle 2).

Hüftendoprothesenversorgung	Mammachirurgie
<ol style="list-style-type: none"> <li>1. 54001 Indikation zur elektiven Hüftendoprothesen-Erstimplantation</li> <li>2. 54003 Präoperative Verweildauer</li> <li>3. 54004 Sturzprophylaxe</li> <li>4. 54016 Allgemeine Komplikationen bei elektiver Hüftendoprothesen-Erstimplantation</li> <li>5. 54018 Spezifische Komplikationen bei endoprothetischer Versorgung einer hüftgelenknahen Femurfraktur</li> <li>6. 54019 Spezifische Komplikationen bei elektiver Hüftendoprothesen-Erstimplantation</li> <li>7. 54012 Verhältnis der beobachteten zur erwarteten Rate (O/E) an Patientinnen und Patienten mit Gehunfähigkeit bei Entlassung</li> <li>8. 54013 Todesfälle während des akut-stationären Aufenthaltes bei geringer Sterbewahrscheinlichkeit</li> </ol> <p>Indikatoren werden ausgeschlossen, sofern sie sich auf den Hüftendoprothesen- bzw. Komponentenwechsel beziehen.</p>	<ol style="list-style-type: none"> <li>1. 51846 Prätherapeutische histologische Diagnosesicherung</li> <li>2. 52330 Intraoperative Präparatradiografie oder intraoperative Präparatsonografie bei mammografischer Drahtmarkierung</li> <li>3. 52279 Intraoperative Präparatradiografie oder intraoperative Präparatsonografie bei sonografischer Drahtmarkierung</li> <li>4. 2163 Primäre Axilladisektion bei DCIS</li> <li>5. 50719 Lymphknotenentnahme bei DCIS und brusterhaltender Therapie</li> <li>6. 51847 Indikation zur Sentinel-Lymphknoten-Biopsie</li> <li>7. 51370 Zeitlicher Abstand von unter 7 Tagen zwischen Diagnose und Operation</li> </ol>

Tabelle 2: Qualitätsindikatoren der externen stationären Qualitätssicherung (2019) für die beiden Fallbeispiele

2. Für den medizinischen Bereich und hier im Besonderen der leistungsbereichsspezifischen Struktur-Qualitätsparameter behandlungsrelevante Ausstattung wird auf Kennzahlen zurückgegriffen, die ebenfalls im Rahmen der externen stationären Qualitätssicherung erhoben und veröffentlicht werden. Hierbei wurde sich auf die bereits auf der Weissen Liste gezeigten Informationen bezogen und diese übernommen. Diese umfassen 11 Ausstattungsmerkmale bei der Hüftendoprothesenversorgung bzw. 19 Ausstattungsmerkmale bei der Mammachirurgie (siehe Tabelle 3).

Hüftendoprothesenversorgung	Mammachirurgie
<ol style="list-style-type: none"> <li>1. Ärztliche Qualifikation (4 Kriterien)</li> <li>2. Spezielles therapeutisches Personal (3 Kriterien)</li> <li>3. Medizinisches Leistungsangebot der Fachabteilungen (4 Kriterien)</li> </ol>	<ol style="list-style-type: none"> <li>1. Besondere apparative Ausstattung (3 Kriterien)</li> <li>2. Ärztliche Qualifikation (10 Kriterien)</li> <li>3. Medizinisches Leistungsangebot der Fachabteilungen (6 Kriterien)</li> </ol>

Tabelle 3: Behandlungsrelevante Ausstattungsparameter für die beiden Fallbeispiele

3. Für den medizinischen Bereich der Patientensicherheit und Hygiene werden insgesamt 55 Kriterien berücksichtigt, welche von der Weissen Liste bereits auf der Krankenhaussuche dargestellt werden. Diese sind unabhängig vom Leistungsbereich und daher für die beiden Fallbeispiele identisch.
  - Allgemeine Maßnahmen (14 Kriterien)
  - Spezifische Maßnahmen (8 Kriterien)
  - Maßnahmen beim Operieren (6 Kriterien)
  - Hygiene & Infektionsschutz (14 Kriterien)
  - Arzneimitteltherapiesicherheit (13 Kriterien)
  
4. Für den Bereich Nutzerfahrung wurde festgelegt, dass die Patientenweiterempfehlungsrate, ein allgemeiner Indikator aus der PEQ-Patientenbefragung, herangezogen wird. Dieser stellt die Weiterempfehlung des Krankenhauses dar und zeigt den Prozentwert der Patienten, die ihrem besten Freund bzw. ihrer besten Freundin das Krankenhaus weiterempfehlen würden.

### 3.3 Berechnung von Zwischenergebnissen für die festgelegten Qualitätskategorien

Im nächsten Schritt wurde für die vier Qualitätskategorien jeweils ein numerischer Wert ermittelt. Dies ist ein Zwischenschritt auf dem Weg zur Berechnung des Composite Measures. Für die vier Bereiche wurde jeweils ein Zwischenergebnis in prozentualer Form berechnet. Für die einzelnen Qualitätskategorien waren hierfür unterschiedliche Schritte notwendig.

- Im Bereich der leistungsbereichsspezifischen Qualitätsparameter der externen stationären Qualitätssicherung wurde der prozentuale Anteil der Qualitätsindikatoren ermittelt, für die nach dem Strukturierten Dialog [25] eine unauffällige Qualität bestätigt wurde. Hierbei wurde sich an die Klassifizierungssystematik der Qualitätsberichterstattung der Nürnberger Transparenzinitiative angelehnt [26]. Dies bedeutet beispielsweise, dass Dokumentationsfehler in der Qualitätssicherung als eine nicht unauffällig bestätigte Qualität gewertet wurden. Für eine Klinik jeweils nicht zutreffende Qualitätsindikatoren wurden bei der klinikspezifischen Bewertung entsprechend nicht berücksichtigt.  
Beispiel: Von 8 Qualitätsindikatoren im Bereich der Hüftendoprothesenversorgung wurden für ein Krankenhaus nach dem Strukturierten Dialog 7 Qualitätsindikatoren als unauffällig eingestuft, 1 Qualitätsindikator hingegen als nicht unauffällig. Dies ergibt ein Ergebnis in Höhe von 87,5%.
- Für den Bereich der leistungsbereichsspezifischen Struktur-Qualitätsparameter behandlungsrelevante Ausstattung wurde ermittelt, welchen prozentualen Anteil an Ausstattungsmerkmalen eine Klinik jeweils erfüllt (11 Ausstattungsmerkmale bei der Hüftendoprothesenversorgung bzw. 19 Ausstattungsmerkmale bei der Mammachirurgie).  
Beispiel: Bei 19 Ausstattungsmerkmalen im Bereich der Mammachirurgie erfüllt eine Klinik insgesamt 17 Merkmale. Dies ergibt ein Ergebnis in Höhe von 89,5%.

- Im Bereich der *Patientensicherheit und Hygiene* wurde ermittelt, wie hoch der Anteil an Kriterien ist, den eine Klinik jeweils erfüllt. Insgesamt gibt es 55 leistungsbereichsunspezifische Kriterien.  
Beispiel: Eine Klinik erfüllt insgesamt 53 der 55 Merkmale. Dies ergibt ein Ergebnis in Höhe von 96,4%.
- Für den Bereich *Patientenweiterempfehlungsrates* wird ein allgemeiner Indikator aus der PEQ-Patientenbefragung herangezogen, der sich auf die Weiterempfehlung des Krankenhauses bezieht. Dieser Indikator zeigt den Prozentwert der Patienten, die ihrem besten Freund bzw. ihrer besten Freundin das Krankenhaus weiterempfehlen würden. Hier ist keine weitere Berechnung notwendig. Die vorliegende Kennzahl stellt bereits das benötigte Zwischenergebnis in Form einer Prozentzahl dar.  
Beispiel: 84% der befragten Patienten würden das Krankenhaus ihrem besten Freund bzw. ihrer besten Freundin weiterempfehlen.

### 3.4 Standardisierung der Zwischenergebnisse (z-Transformation)

Wie oben beschrieben, wird bei der z-Transformation jeder Ausgangswert in einen neuen Wert überführt. Dabei werden Mittelwert und Standardabweichung berücksichtigt, sodass auch Kennzahlen mit vorher unterschiedlichen Einheiten standardisiert und damit vergleichbar werden. Der Mittelwert der standardisierten Skala beträgt 0 und die Standardabweichung 1. In der folgenden Tabelle 4 wird die z-Transformation für die Kennzahl PEQ beispielhaft dargestellt.

	PEQ-Quote	Z-Wert (PEQ-Quote)
Mean	,8048	,0000000
Std. Deviation	,06483	1,00000000
Minimum	,60	-3,15958
Maximum	,97	2,54800

Tabelle 4: z-Transformation der PEQ-Werte in standardisierte Werte (N=1188 Kliniken)

Die folgende Abbildung 6 veranschaulicht den Sachverhalt der z-Transformation der PEQ-Werte in standardisierte Werte grafisch.

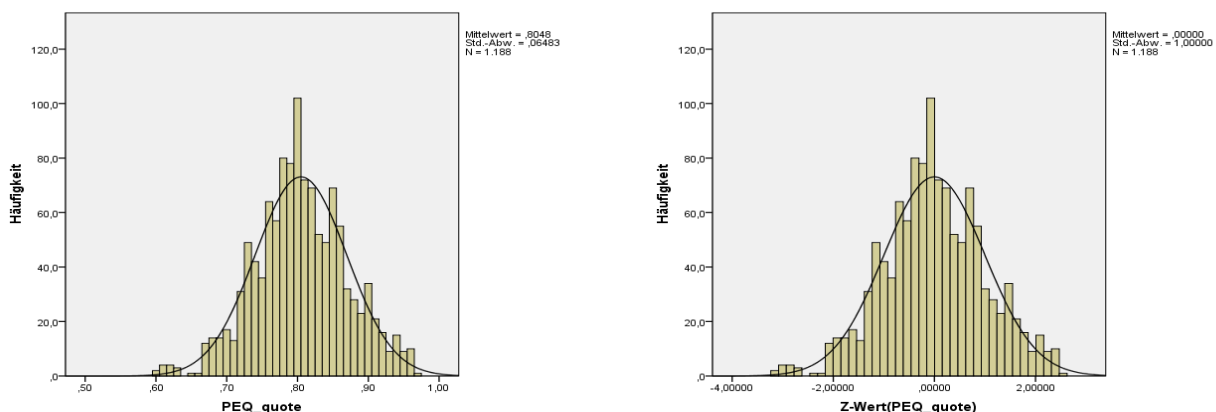


Abbildung 6: z-Transformation der PEQ-Werte in standardisierte Werte (N=1.188 Kliniken)



Würde man die Werte aus den festgelegten Qualitätskategorien nicht standardisieren, würden sich daraus folgende Probleme ergeben. Zum einen könnte kein Vergleich stattfinden, wenn die Qualitätsdaten unterschiedliche Einheiten hätten. Zum anderen würde man bei einem reinen Vergleich von Prozentwerten nicht den Mittelwert und die Standardabweichung berücksichtigen. Ob nun ein bestimmter Wert besonders hoch oder besonders niedrig wäre, würde nicht beachtet werden. Dies soll an dem folgenden Beispiel erläutert werden. In Tabelle 5 sind als Ergebnis sowohl die Prozentwerte als auch die z-Werte für zwei Krankenhäuser aufgeführt. Würde man für die vier Qualitätsbereiche lediglich einen Durchschnittswert bilden, wäre Krankenhaus A (0,96 vs. 0,94) das bessere Krankenhaus. Hierbei werden allerdings nur Quoten berücksichtigt, ohne einen Bezug zu den anderen Krankenhäusern herzustellen; dies kann über die Transformation in einen z-Wert vollzogen werden. Dadurch, dass der eigene Wert ins Verhältnis zu allen anderen Werten gesetzt wird, kann beurteilt werden, was hoch oder niedrig ist.

	PEQ_Quote	BRA_Quote	PuH_Quote	esQS_Quote	
KH A	0,88	<b>1</b>	<b>0,95</b>	1,0	<b>0,96</b>
KH B	<b>0,94</b>	0,91	0,91	1,0	0,94
	PEQz_Final	BRAz_Final	PuHz_Final	esQSz_Final	
KH A	1,15967	<b>1,81895</b>	<b>1,14077</b>	0,57	1,17
KH B	<b>2,08522</b>	1,29378	0,84597	0,57	<b>1,20</b>

*Tabelle 5: Gegenüberstellung der %-Werte und der z-Werte bei zwei ausgewählten Krankenhäusern (Legende: PEQ Patients' Experience Questionnaire, esQSz Externe stationäre Qualitätssicherung, BRA Behandlungsrelevante Ausstattung, PuH Patientensicherheit und Hygiene)*

Krankenhaus B ist nur in einem einzigen Bereich besser als Krankenhaus A (PEQ). Hier ist der Wert in Höhe von 0,94 aber so besonders einzuschätzen, weil Krankenhaus B nicht nur deutlich über dem Mittelwert, sondern in der Spitzengruppe liegt, und dadurch einen sehr hohen z-Wert erhält. Dieses Krankenhaus B ist im Vergleich zu allen anderen Krankenhäusern in diesem Bereich so gut, dass der z-Wert aus dieser Kategorie die Werte aus den anderen Bereichen in dem Sinne ausgleicht, als dass nun Krankenhaus B bei einem standardisierten Vergleich besser abschneidet als Krankenhaus A.

### 3.5 Begrenzung negativer Ausreißerwerte

In Anlehnung an Hospital Compare [11] werden negative Ausreißer eliminiert. Dieser Schritt wird vollzogen, um der (Un)Sicherheit der zugrunde liegenden Daten Rechnung („potentially inaccurate reporting“ [11]) zu tragen. Im Gegensatz zu CMS Hospital Compare werden allerdings nicht alle negativen und positiven Extremwerte auf -3 / +3 des z-Wertes begrenzt (0.125th Perzentil bzw. 99.875 Perzentil), sondern es werden alle negativen Werte, die sich innerhalb des 1%-Perzentils befinden, auf exakt den Wert festgesetzt, welchen das Krankenhaus hat, das dieses letzte 1%-Perzentil einleitet. Daneben wurden in alternativen Szenarien eine Begrenzung beim 2,5%-Perzentil bzw. keine Berücksichtigung der negativen Ausreißer vorgenommen. Da es hier zu unwesentlich anderen Ergebnissen kam, ist das 1%-Perzentil eine konservative Annahme, um mögliche fehlerhafte Werte auszuschließen. Bei der Anwendung der Begrenzung beim 1%-Perzentil ergeben sich die folgenden deskriptiven

Ergebnisse für die z-Werte der vier Qualitätskategorien für den Leistungsbereich der Hüftendoprothesenversorgung (siehe Tabelle 6).

	Hüftendoprothesenversorgung				Mammachirurgie			
	PEQz_Final	esQsz_Final	BRAz_Final	PuHz_Final	PEQz_Final	esQsz_Final	BRAz_Final	PuHz_Final
Valid	1188	1175	1208	1208	1183	606	1259	1259
Missing	20	33	0	0	76	653	0	0
Mean	,00	,00	,00	,00	,00	,00	,00	,00
Std. Deviation	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Minimum*	-3,16	-6,33	-3,96	-4,61	-3,18	-5,52	-1,81	-4,57
Maximum	2,55	,57	1,82	1,58	2,75	,33	1,92	1,52

\* Extrem negative Werte wurden bei dem Wert des 1%-Perzentils begrenzt

PEQ Patients' Experience Questionnaire, esQS-Externe stationäre Qualitätssicherung, BRA Behandlungsrelevante Ausstattung, PuH Patientensicherheit und Hygiene

Tabelle 6: Deskriptive Ergebnisse für die z-Werte der vier Qualitätskategorien

Die Begrenzung auf das unter 1%-Perzentil berücksichtigt das mögliche Problem einer ungenauen Berichterstattung, ohne aber zu weit einzugreifen. Das Setzen der Grenze auf das 2,5%-Perzentil hätte nur wenig starke Auswirkung auf die Sternenvergabe. Zwar würde sich bei dieser Grenze bei 84 von 1.188 Krankenhäusern der finale Composite Measure ändern (7,07%), aber lediglich bei 6 Krankenhäusern (0,51%) würde sich die Gruppe mit den Sternen verändern. Hierbei handelt es sich um Krankenhäuser, die knapp an der Grenze zu den 2 Sternen stehen. Mit der 1%-Grenze berücksichtigt man die Gefahr der ungenauen Berichterstattung, ohne diesem aber zu viel Raum zu geben und möglicherweise wirklich qualitativ auffällig arbeitende Krankenhäuser nicht ungerechtfertigterweise zu schützen.

Zuletzt muss angemerkt werden, dass nur eine Begrenzung bei den negativen Werten durchgeführt wurde, weil Krankenhäuser, bei denen möglicherweise eine fehlerhafte Berichterstattung aufgetreten ist, vor einer ungerechtfertigten Schlechterstellung bewahrt werden sollen. Eine Grenze bei den positiven Werten wurde nicht gezogen, da die Krankenhäuser, die tatsächlich Top-Performer sind und keine fehlerhafte Berichterstattung aufweisen, ungerechtfertigter Weise durch eine Begrenzung schlechter gestellt worden wären. Dies sollte vermieden werden.

### 3.6 Berechnung eines ungewichteten Composite Measures

Im nächsten Schritt erfolgt die Berechnung eines ungewichteten Composite Measures. Dieser ergibt sich aus der Gesamtsumme der z-Werte der vier Qualitätskategorien, dividiert durch die Anzahl an Kategorien. Hierbei muss angemerkt werden, dass bei einem fehlenden Wert (d.h. eine Qualitätskategorie weist keine Ergebnisse auf) kein Composite Measure für ein Krankenhaus ermittelt wird. Eine Berechnung findet nur dann statt, wenn alle vier Werte vollständig vorliegen. Bei 20 Krankenhäusern lagen keine PEQ-Ergebnisse vor, bei 33 Krankenhäusern keine Informationen für die externe stationäre Qualitätssicherung, so dass insgesamt 52 Krankenhäuser bei der Ermittlung des Composite Measures nicht berücksichtigt werden konnten.

Tabelle 7 zeigt die deskriptiven Ergebnisse für den ungewichteten z-Wert für die 1.156 Krankenhäuser im Bereich Hüftendoprothesenversorgung und für die 593 Krankenhäuser im Bereich der Mammachirurgie, für die jeweils ein Composite Measure ermittelt wurde.

Deskriptive Statistik zum z-Wert	Hüftendoprothesenversorgung (N=1.156)	Mammachirurgie (N=593)
Mean	,0272	,2228
Std. Deviation	,52478	,48825
Minimum	-2,13	-2,47
Maximum	1,37	1,30

Tabelle 7: Deskriptive Ergebnisse für den ungewichteten z-Wert (N=1.156 Krankenhäuser)

### 3.7 Zuordnung zu den Sternengruppen

Auf Basis des ermittelten Composite Measures findet eine Zuordnung eines Krankenhauses zu einer von drei Qualitätskategorien statt. Es wurde eine Systematik festgelegt, in der es drei Gruppen mit jeweils unterschiedlicher Qualität entsprechend des oben beschriebenen Qualitätsmodells gibt. Ein Stern bedeutet, dass ein Krankenhaus zu den Krankenhäusern mit niedriger Qualität gehört, bei zwei Sternen zu mittlerer und bei drei Sternen zu hoher Qualität. Je nachdem welche Ausprägung der Composite Measure hat, wird ein Krankenhaus zu einer der drei Sternengruppen zugeordnet. In der Gruppe mit den zwei Sternen und damit der mittleren Qualität wird man zugeordnet, wenn sich der Composite Measure innerhalb einer halben Standardabweichung um den Mittelwert befindet. Ein Stern erhält ein Krankenhaus, sofern der ungewichtete Composite Measure mehr als 0,5 der Standardabweichung vom Mittelwert nach unten abweicht. Drei Sterne erhält ein Krankenhaus, wenn der Composite Measure über 0,5 der Standardabweichung liegt. Wie in Tabelle 8 dargestellt, sind im Bereich Hüftendoprothesenversorgung insgesamt 16,8% der Kliniken in die Performancegruppe mit einem Stern eingeordnet worden, 58,1% der Kliniken zur Gruppe mit zwei Sternen und 25,1% aller Kliniken in die Gruppe mit drei Sternen.

Sterne	Hüftendoprothesenversorgung		Mammachirurgie	
	Häufigkeit (absolut)	Häufigkeit (relativ)	Häufigkeit (absolut)	Häufigkeit (relativ)
1	194	16,8	33	5,6%
2	672	58,1	291	49,1%
3	290	25,1	269	45,4%
Gesamt	1.156	100,0	593	100,0%

Tabelle 8: Deskriptive Ergebnisse zur Einteilung der Kliniken in drei Performancegruppen

### 3.8 Setzen von Filterkriterien für Top-KH führt zu Empfehlung

Bisher wurde mit statistischen Methoden ein Composite Measure berechnet, der die Qualität eines Krankenhauses ausdrücken soll. Basis dieser Berechnung sind Kennzahlen aus den festgelegten Qualitätskategorien. Neben diesen Kennzahlen gibt es allerdings noch weitere Merkmale, die ebenfalls bei der Bewertung von Qualität berücksichtigt werden können, aber auf Grund ihrer Skalierung nicht mit in die Berechnung aufgenommen werden können. Bisher wurden im quantitativen Teil metrisch skalierte Variablen verwendet. Darüber hinaus könnten ordinal und nominal skalierte Variablen berücksichtigt werden. Da diese, wie oben angeführt, nicht in die Berechnung eines Composite Measures aufgenommen werden, erfolgt die Integration auf alternative Art und Weise (qualitativer Teil der Bewertung).

Ist ein Krankenhaus den drei Sternen zugeordnet worden, wird eine hohe Qualität zugeschrieben. Nun können innerhalb dieser Gruppe besonders herausragende Krankenhäuser durch eine Empfehlung noch stärker hervorgehoben werden. Erfüllen die Krankenhäuser besondere Anforderungen (je nach Leistungsbereiche können unterschiedliche Anforderungen existieren), erhalten sie zusätzlich eine Empfehlung. Eine Empfehlung wird für ein Krankenhaus ausgesprochen, wenn beispielsweise im Bereich der Hüftendoprothesenversorgung folgende Anforderungen erfüllt werden:

3 Sterne,

Top 10% beim Gesamt-z-Wert,

Vollständig unauffällige Versorgungsqualität für die externe Qualitätssicherung (d.h. eine Quote von 100%) und Fallzahl oberhalb des Medians.

Durch diesen Schritt ist es möglich, beispielsweise auch Fallzahlen mit in die Bewertung aufzunehmen. Fallzahlen sind ein metrischer Wert und könnten vom statistischen Procedere her auch im quantitativen Teil verwendet werden, jedoch gibt es, wie oben beschrieben, bei den Fallzahlen einen inhaltlichen Ausschluss. In der nachfolgenden Tabelle 9 werden Krankenhäuser gelistet, für die in den beiden Leistungsbereichen Hüftendoprothesenversorgung und Mammachirurgie eine Empfehlung ausgesprochen wird.

Hüftendoprothesenversorgung (n=18)	Mammachirurgie (n=15)
BG Klinik Ludwigshafen	
Breitenbrunn	Diakonie-Klinikum Stuttgart
Diakonie-Klinikum Stuttgart	Ev. Diakoniekrankenhaus
Helios ENDO-Klinik Hamburg	Kath. Marien-Krankenhaus Lübeck gemeinnützige GmbH
Klinikum Garmisch-Partenkirchen - Haupthaus	KEM   Evang. Kliniken Essen-Mitte gGmbH, Standort
Krankenhaus Martha-Maria Nürnberg	Evang. HuysSENS-Stiftung Essen-Huttrop
Krankenhaus St. Marienstift Magdeburg GmbH	Klinikum Garmisch-Partenkirchen - Haupthaus
Oberlinklinik Orthopädische Fachklinik	Krankenhaus Barmherzige Brüder Regensburg
Orthopädische Klinik Markgröningen	Krankenhaus St. Joseph-Stift Dresden GmbH
Parkklinik Manhagen GmbH & Co. KG	Krankenhaus St. Marienstift Magdeburg GmbH
Sana Kliniken Solln Sendling GmbH Umfirmierung im September 2018 in Sana Klinik München GmbH	Marienhospital Stuttgart
Sana Kliniken Sommerfeld	St. Elisabeth-Krankenhaus Leipzig
Schön Klinik Harthausen	Städtisches Klinikum Görlitz gGmbH
St. Franziskus-Hospital GmbH Münster	Universitätsklinikum Carl Gustav Carus Dresden an der Technischen Universität Dresden, Anstalt des öffentlichen Rechts des Freistaates Sachsen
St. Josef-Stift Sendenhorst	Universitätsklinikum Erlangen
Universitätsklinikum Carl Gustav Carus Dresden an der Technischen Universität Dresden, Anstalt des öffentlichen Rechts des Freistaates Sachsen	Universitätsklinikum Hamburg-Eppendorf
Universitätsklinikum Heidelberg	Universitätsklinikum Heidelberg
Waldkliniken Eisenberg	

*Tabelle 9: Übersicht über die Kliniken mit einer „Empfehlung“ für die Fallbeispiele Hüftendoprothesenversorgung und Mammachirurgie*

### 3.9 Setzen von Filterkriterien für Low-Performer-KH führt zu Warnung

Im letzten Schritt werden die Krankenhäuser hervorgehoben, die aufgrund des Composite Measures und weiterer besonderer Merkmale negativ hervorstechen. Ist ein Krankenhaus einem Stern zugeordnet worden, wird diesem eine tendenziell niedrige Qualität zugesprochen. Nun können innerhalb dieser Gruppe besonders auffällige Krankenhäuser durch eine Warnung hervorgehoben werden. Haben die Krankenhäuser besondere Auffälligkeiten (je nach Leistungsbereich können unterschiedliche Auffälligkeiten existieren) erhalten sie zusätzlich eine Warnung. Im Bereich der Hüftendoprothesenversorgung wird vorausgesetzt, dass ein Krankenhaus

- 1 Stern hat,
- zu den Low-Performern (untere 10% des z-Wertes) gehört,
- <100 % bei QS und
- eine Fallzahl unterhalb des 20. Perzentils aufweist.

In der nachfolgenden Tabelle 10 werden die Krankenhäuser in den Leistungsbereichen Hüftendoprothesenversorgung und Mammachirurgie gelistet, für eine Warnung ausgesprochen wird.

Hüftendoprothesenversorgung (n=9)	Mammachirurgie (n=2)
<ul style="list-style-type: none"> <li>• AMEOS Klinikum Staßfurt</li> <li>• Dalberg Klinik Fulda</li> <li>• HELIOS Klinik Bad Schwalbach</li> <li>• Hohenloher Krankenhaus Künzelsau</li> <li>• KMG Klinikum Luckenwalde</li> <li>• Krankenhaus Maria Hilf</li> <li>• Paracelsus-Klinik Bad Ems</li> <li>• Sana Kliniken des Landkreises Cham - Krankenhaus Cham</li> <li>• St. Elisabeth Krankenhaus Salzgitter gGmbH</li> </ul>	<ul style="list-style-type: none"> <li>• HELIOS St. Marienberg Klinik Helmstedt</li> <li>• Krankenhaus Maria Hilf</li> </ul>

*Tabelle 10: Übersicht über die Kliniken mit einer „Warnung“ für die Fallbeispiele Hüftendoprothesenversorgung und Mammachirurgie*

## 4 Diskussion der Ergebnisse

Im Folgenden wird die oben vorgestellte Methodik kritisch betrachtet, um einen umfassenden Eindruck von den jeweiligen Stärken und Schwächen zu erhalten.

Zum einen muss bezüglich der Leistungsbereiche erwähnt werden, dass deren Auswahl jeweils klar und zielgerichtet erfolgen muss. Dies trifft insbesondere für solche Leistungsbereiche zu, für die im Rahmen der externen stationären Qualitätssicherung mehrere Unter-Leistungsbereiche vorhanden sind; so zum Beispiel im Bereich der Hüftendoprothesenversorgung, für den drei Unter-Leistungsbereiche vorhanden sind (Erstimplantation, Hüft- und Komponentenwechsel, Femurfraktur). Hier muss jeweils festgelegt werden, für welchen der Bereiche bzw. für welche Bereiche der Composite Measure ermittelt werden soll. Im Rahmen der hier vorgestellten Ausführungen wurde der Wechseleingriff ausgeschlossen, die anderen beiden Unter-Leistungsbereiche sind berücksichtigt worden. Hier könnte theoretisch weiter fokussiert werden, beispielsweise ausschließlich auf die Erstimplantation. Dies hätte insbesondere Auswirkungen auf die entsprechende Berücksichtigung der Qualitätsindikatoren aus der externen stationären Qualitätssicherung.

Des Weiteren muss insbesondere das dem Composite Measure zugrunde Konzept des Qualitätsmodells Gegenstand der Diskussion sein. Das zugrunde liegende Qualitätsmodell ist von elementarer Bedeutung für die Ausgestaltung und Kalkulation des Composite Measures. Dieses gibt beispielsweise die zu berücksichtigenden Qualitätskategorien vor. Am Ende der Kalkulation steht ein einzelner Wert, der „für etwas stehen muss“.

Genau diesen Wert gilt es inhaltlich zu definieren und festzulegen. In dem hier vorgestellten Qualitätsmodell werden vier Qualitätskategorien auf der ersten Ebene und der Aspekt der Erfahrung über die Fallzahl auf der zweiten Prioritätenebene betrachtet; diese beziehen sich auf alle Bereiche der Qualitätstriade nach Donabedian (Struktur, Prozess, Ergebnis). In den Vereinigten Staaten hat Hospital Compare beispielsweise ein anderes Qualitätsmodell zugrunde gelegt. Hier sind Kennzahlen mit Prozess- und Ergebnischarakter integriert und in Prozentwerten (metrisch skaliert) hinterlegt. Insgesamt sind hierbei sieben Qualitätskategorien definiert worden, die auch in der folgenden Abbildung 7 (Step 2) aufgezeigt sind. Nursing Home Compare hingegen kalkuliert seinen Composite

Measure basierend auf insgesamt drei Qualitätskategorien (Health Inspections, Staffing, Quality Measures). Die Kennzahlen der Kategorie „Staffing - Measures based on nursing home staffing levels“ beziehen sich hierbei auf die Struktur der Leistungserbringung [27]. Das Qualitätsmodell von US News besteht hingegen aus 5 Qualitätskategorien, die zusammenfassend betrachtet werden (Strukturqualität, Prozessqualität, Ergebnisqualität, Patientenerfahrung, Öffentliche Transparenz) [28]. Abschließend soll an dieser Stelle Erwähnung finden, dass alternative Qualitätsmodelle ebenfalls durch die hier angewandte Methode umgesetzt werden können. Sollten zusätzliche Qualitätskategorien als relevant erachtet werden, könnten diese in dem Modell berücksichtigt werden.

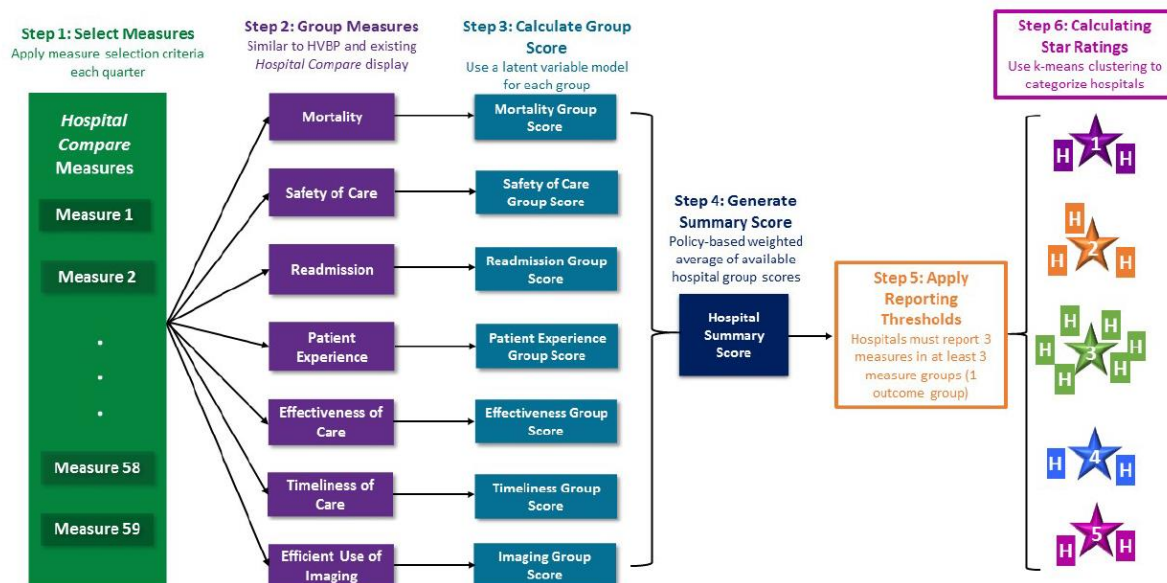


Abbildung 7: The Six Steps of the Overall Star Rating Methodology (Hospital Compare)

Zur Definition des Qualitätsmodells gehört in der nächsten Stufe die Festlegung von Kennzahlen der einzelnen Qualitätskategorien. Diese ist von zentraler Bedeutung, so wurden hier beispielsweise die aktuell auf der Weissen Liste verwendeten Kriterien zur Messung der Qualitätskategorien mit aufgenommen (z.B. 55 Kriterien für Patientensicherheit und Hygiene, 19 Ausstattungsmerkmale für die behandlungsrelevante Ausstattung bei Mammachirurgie, Patientenweiterempfehlungsrates der PEQ-Patientenbefragung). Die Qualitätskategorien hätten allesamt noch einmal empirisch überprüft bzw. nach alternativen Messkonzepten hin untersucht werden können. Im Rahmen der ersten Machbarkeitsanalyse wurde hierauf allerdings verzichtet. Ebenso kann diskutiert werden, anhand welcher Kriterien die Zuordnung des Ergebnisses des Strukturierten Dialogs [25] als unauffällig bzw. nicht unauffällig gewertet wird. So sind beispielsweise Dokumentationsfehler nicht als nachgewiesene Auffälligkeit gewertet, allerdings auch nicht als nachgewiesene Unauffälligkeit.

Für den Patienten ergibt sich hierbei eine Unsicherheit, da der jeweilige Qualitätsindikator nicht nachgewiesen als unauffällig eingestuft werden kann. Im Rahmen des hier entwickelten Composite Measures wurde sich an die Methodik der Qualitätsberichterstattung der Metropolregion Nürnberg angelehnt [26]; hierbei werden lediglich solche Qualitätsergebnisse als unauffällig eingestuft, die durch den Strukturierten Dialog auch als solche gekennzeichnet wurden (nicht relevante Qualitätsindikatoren wurden jeweils nicht berücksichtigt).

Ebenfalls kann diskutiert werden, ob sich die ermittelten Composite Measures auf das gesamte Krankenhaus oder einen speziellen Leistungsbereich beziehen sollen. Auf den ersten Blick würde man tendenziell eine spezifische Ausrichtung vermuten; diese erfordert allerdings auch relevante Daten, die hier zugrunde gelegt werden müssen. In den beiden Fallbeispielen (Hüftendoprothetik, Mammakarzinom) liegen jeweils relevante Qualitätsinformationen aus der externen stationären Qualitätssicherung vor, die eine spezifische Betrachtung ermöglichen. Wie sollte aber beispielsweise mit Leistungsbereichen umgegangen werden, zu denen keine spezifischen Qualitätsdaten vorliegen? Andere Transparenzinitiativen wählen für das Public Reporting unterschiedliche Ansätze. So weist Hospital Compare beispielsweise einen nicht-krankheitsspezifischen Composite Measure aus, US News hingegen schon (siehe Abbildung 8). Alternativ könnte auch der Weg über Routinedaten weiter angedacht werden; beispielsweise würde die Einbindung der QSR-Daten der AOK (bzw. Initiative Qualitätsmedizin) sowohl zusätzliche Leistungsbereiche mit Qualitätsergebnissen hervorbringen (z.B. Leistenhernie, Prostatakarzinom) als auch die Datenbasis zu bereits abgebildeten Leistungsbereichen erweitern (z.B. Hüftendoprothetik). Wettbewerbsrelevante Aspekte könnten die Möglichkeit der Integration zusätzlicher Daten allerdings auch limitieren. Es sollte jeweils diskutiert werden, welche Daten für einen Leistungsbereich vorhanden sind und ob diese sinnvoll und aussagekräftig zusammengefasst werden können.


The image displays two screenshots of hospital comparison websites. The left screenshot is from Medicare.gov's Hospital Compare tool, showing a list of hospitals in New York with columns for 'Hospital information', 'Overall rating', 'Emergency Services', and 'Hospital Type'. A red circle highlights the 'Overall rating' column, which shows star ratings and the number of reviews. The right screenshot is from US News Best Hospitals for Orthopedics, showing a list of hospitals with columns for 'Hospital Name', 'Location', and 'Specialty'. A red circle highlights the 'Orthopedics Score' column, which shows a score out of 100 and a link to the 'Orthopedics Scorecard'.

Abbildung 8: Composite Measures auf Hospital Compare und US News


Im Rahmen der hier vorgestellten Methodik wurde bei der Zusammenfassung der vier Qualitätskategorien (externe stationäre Qualitätssicherung, Patientensicherheit & Hygiene, Patientenerfahrungen, Behandlungsrelevante Ausstattung) eine Gleichgewichtung vorgenommen; d.h. alle vier Qualitätskategorien gehen mit 25% Gewicht in die Kalkulation des finalen Composite Measures ein. Dieser Ansatz wurde in einem ersten Schritt gewählt, da eine alternative Gewichtung (d.h. die einen Qualitätskategorien mehr Gewicht, die anderen Qualitätskategorien weniger Gewicht) eine inhaltliche Begründung notwendig(er) erscheinen lässt als dies bei einer Gleichgewichtung der Fall wäre. Hierfür könnten insbesondere Patientenpräferenzen ausschlaggebend sein (z.B. durch ein DCE-



Experiment ermittelt). Allerdings wurde an dieser Stelle aufgrund der Heterogenität der Literatur hierauf verzichtet, diese als Basis für eine inhaltliche Gewichtung heranzuziehen. So existieren zwar durchaus Studien zur Patientenpräferenz bei der Krankenhauswahl [29–43], allerdings sind diese nur limitiert dazu geeignet, das hier zugrunde gelegte Qualitätsmodell zu gewichten. Eine Studie aus dem Jahr 2019 könnte hierfür möglicherweise herangezogen werden [33], allerdings wurde hier ein generelles, d.h. kein krankheitsspezifisches, Sample befragt. Einem krankheitsspezifischem Composite Measure sollte allerdings auch eine krankheitsspezifische Präferenzstruktur zugrunde gelegt werden. Internationale Erfahrungen sollen an dieser Stelle ebenfalls Erwähnung finden. So zeigt die folgende Abbildung 9, dass den Composite Measures von US News und Hospital Compare unterschiedliche Gewichtungen der einzelnen Qualitätskategorien zugrunde liegen. Während die Patientenerfahrung bei US News beispielsweise lediglich ein Gewicht 5% aufweist, nimmt sie im Rahmen von Hospital Compare mit 22% einen deutlich größeren Stellenwert ein.



Component	Cardiology & Heart Surgery Weight (%)	Weight, All Other Specialties (%)
Outcomes	37.5%	37.5%
Structure	30.0%	30.0%
Process/expert opinion	24.5%	27.5%
Patient experience	5.0%	5.0%
Public transparency	3.0%	0.0%

Group	Star Ratings Weight
Mortality	22%
Safety of Care	22%
Readmission	22%
Patient Experience	22%
Effectiveness of Care	4%
Timeliness of Care	4%
Efficient Use of Medical Imaging	4%

Abbildung 9: Gewichtungsfaktoren für die Ermittlung des Composite Measures bei US-Transparenzinitiativen

Im Einklang mit Hospital Compare [11] wurde im Rahmen des hier entwickelten Composite Measure ebenfalls eine Begrenzung extremer Werte vorgenommen, um der (Un)Sicherheit der zugrunde liegenden Daten Rechnung zu tragen, allerdings wurden die Grenzen hierzu anders gewählt. Im Gegensatz zu CMS Hospital Compare wurden einerseits nur negative Werte begrenzt, andererseits wurden die Grenzen im Basismodell bei dem Wert des 1%-Perzentils begrenzt. Es wurden zwei alternative Szenarien gerechnet (keine Begrenzung, Begrenzung beim 2,5%-Perzentil), die Ergebnisse zeigten sich hierbei sehr robust. Es kam nur vereinzelt zu Verschiebungen (6 Krankenhäuser) zwischen der Gruppe mit einem Stern und mit zwei Sternen. Würde man die Grenzen von CMS Hospital Compare hier anwenden, würde es je nach Qualitätskategorie einige Krankenhäuser betreffen, bei denen der z-Wert des Zwischenergebnisses sehr stark verändert wird.

## 5 Zusammenfassung und Ausblick

Das Ziel der hier vorliegenden Untersuchung bestand in der Entwicklung und beispielhaften Anwendung einer Methodik für die Qualitätsmessung und -darstellung mittels eines sog. Composite Measures (Machbarkeitsuntersuchung). Im Gegensatz zur Darstellung vieler einzelner Qualitätsinformationen werden bei einem Composite Measure zwar auch viele Informationen berücksichtigt, diese jedoch zu einer einzigen Kennzahl zusammengefasst und aggregiert dargestellt. Composite Measures kommen aktuell im internationalen Kontext vermehrt vor [6–8], insbesondere auch auf in der Literatur bekannten Report Cards (z. B. Hospital Compare, Nursing Home Compare, NHS Choices, US News Best Hospital, Home Health star rating system). Der nützliche Aspekt des Composite Measures soll dadurch entstehen, dass durch die reduzierte Menge an bereitgestellten Informationen die Komplexität der Entscheidungsfindung reduziert wird, was sich positiv auf das Wahlverhalten von Konsumenten auswirken kann [7]. Des Weiteren können Composite Measures dazu geeignet sein, künftige Qualitätsergebnisse von Leistungserbringern vorherzusagen [10].

Bei der Entwicklung und Ermittlung von Composite Measures sind zahlreiche Entscheidungen zu treffen, die eine grundlegende Auswirkung auf die finalen Ergebnisse haben. Die hier vorliegende Machbarkeitsanalyse stellt eine mögliche Vorgehensweise für die Ermittlung von Composite Measures dar; der Vergleich mit internationalen Instrumenten hat allerdings auch alternative Ansätze aufgezeigt. Die Methodik sollte hierbei sowohl auf andere Leistungsbereiche übertragen werden können als auch weitere Qualitätsdaten einbinden können. Gezeigt wurden die Ausführungen anhand der beiden Fallbeispiele (1) Hüftendoprothesenversorgung und (2) Mammachirurgie. Verwendet wurden (insbesondere) die öffentlich zugänglichen Daten der externen stationären Qualitätssicherung (im Folgenden Gesetzliche Qualitätssicherung, Gesetzliche QS) sowie bereits auf der Weisen Liste gezeigte Informationen. Insgesamt wurden 9 Schritte durchgeführt, um einen spezifischen Composite Measure für jedes Krankenhaus zu ermitteln und darüber hinaus spezielle Empfehlungen oder Warnungen abzuleiten. Es soll abschließend noch einmal darauf hingewiesen werden, dass die angewandte Vorgehensweise nur eine von vielen möglichen Ansätzen darstellt und an vielen Stellen modifiziert werden könnte. Insbesondere die Vergleiche mit internationalen Composite Measures könnten Anregungen für Weiterentwicklungen geben.

## 6 Literaturverzeichnis

- [1] M. Emmert, N. Meszmer, A. Simon, U. Sander, Internetportale für die Krankenhauswahl in Deutschland: Eine leistungsbereichsspezifische Betrachtung, *Gesundheitswesen* (2015). <https://doi.org/10.1055/s-0035-1549968>.
- [2] M. Emmert, U. Sander, A.S. Esslinger, M. Maryschok, O. Schoeffski, Public reporting in Germany: the content of physician rating websites, *Methods Inf Med* 51 (2012) 112–120. <https://doi.org/10.3414/ME11-01-0045>.
- [3] K. Kast, M. Emmert, C.B. Maier, Public Reporting über stationäre Pflegeeinrichtungen in Deutschland: Eine Bestandsaufnahme und Bewertung der Qualitätsinformationen, *Gesundheitswesen* (2020). <https://doi.org/10.1055/a-1160-5720>.
- [4] M. Emmert, N. Meszmer, U. Sander, Public Reporting zu Krankenhäusern in den USA und in Großbritannien, in: J. Klauber, F. Dormann (Eds.), *Qualitätsmonitor 2017*, first. Auflage, neue Ausgabe, MWV Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, 2016, 65-84.
- [5] F. Rothenfluh, P.J. Schulz, Content, Quality, and Assessment Tools of Physician-Rating Websites in 12 Countries: Quantitative Analysis, *J Med Internet Res* 20 (2018) e212. <https://doi.org/10.2196/jmir.9105>.
- [6] M. Schwartz, A.K. Rosen, J.F. Burgess, Can Composite Measures Provide a Different Perspective on Provider Performance Than Individual Measures?, *Med Care* 55 (2017) e99-e103. <https://doi.org/10.1097/MLR.0000000000000407>.
- [7] R.M. Werner, R.T. Konetzka, D. Polsky, Changes in Consumer Demand Following Public Reporting of Summary Quality Ratings: An Evaluation in Nursing Homes, *Health Serv Res* 51 Suppl 2 (2016) 1291–1309. <https://doi.org/10.1111/1475-6773.12459>.
- [8] N. Bhandari, D.P. Scanlon, Y. Shi, R.A. Smith, Why Do So Few Consumers Use Health Care Quality Report Cards? A Framework for Understanding the Limited Consumer Impact of Comparative Quality Information, *Med Care Res Rev* 76 (2019) 515–537. <https://doi.org/10.1177/1077558718774945>.
- [9] M. Schwartz, J.D. Restuccia, A.K. Rosen, Composite Measures of Health Care Provider Performance: A Description of Approaches, *Milbank Quarterly* 93 (2015) 788–825. <https://doi.org/10.1111/1468-0009.12165>.
- [10] L.M. Chen, D.O. Staiger, J.D. Birkmeyer, A.M. Ryan, W. Zhang, J.B. Dimick, Composite quality measures for common inpatient medical conditions, *Med Care* 51 (2013) 832–837. <https://doi.org/10.1097/MLR.0b013e31829fa92a>.
- [11] Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation, Overall Hospital Quality Star Rating on Hospital Compare Methodology Report (v3.0), 2017.
- [12] U. Nimptsch, T. Mansky, Hospital volume and mortality for 25 types of inpatient treatment in German hospitals: observational study using complete national data from 2009 to 2014, *BMJ Open* 7 (2017) e016184. <https://doi.org/10.1136/bmjopen-2017-016184>.
- [13] C. Bolczek, U. Nimptsch, M. Möckel, T. Mansky, Versorgungsstrukturen und Mengen-Ergebnis-Beziehung beim akuten Herzinfarkt – Verlaufsbetrachtung der deutschlandweiten Krankenhausabrechnungsdaten von 2005 bis 2015, *Gesundheitswesen* (2019). <https://doi.org/10.1055/a-0829-6580>.
- [14] A. Donabedian, The quality of care. How can it be assessed?, *JAMA* 260 (1988) 1743–1748.

- [15] A. Donabedian, Evaluating the quality of medical care, *Milbank Mem Fund Q* 44 (1966) Suppl:166-206.
- [16] S.M. Campbell, M.O. Roland, S.A. Buetow, Defining quality of care, *Soc Sci Med* 51 (2000) 1611–1625.
- [17] A. Donabedian, Evaluating the quality of medical care. 1966, *Milbank Q* 83 (2005) 691–729. <https://doi.org/10.1111/j.1468-0009.2005.00397.x>.
- [18] M. Emmert, T. Adelhardt, U. Sander, V. Wambach, J. Lindenthal, A cross-sectional study assessing the association between online ratings and structural and quality of care measures: results from two German physician rating websites, *BMC Health Services Research* 15 (2015) 414. <https://doi.org/10.1186/s12913-015-1051-5>.
- [19] B. Kolb, M. Emmert, U. Sander, C. Patzelt, O. Schöffski, Do German public reporting websites provide information that office-based physicians consider before referring patients to hospital? A four-step analysis, *Z Evid Fortbild Qual Gesundheitswes* (2018). <https://doi.org/10.1016/j.zefq.2018.07.010>.
- [20] X. Lin, H. Tao, M. Cai, A. Liao, Z. Cheng, H. Lin, A Systematic Review and Meta-Analysis of the Relationship Between Hospital Volume and the Outcomes of Percutaneous Coronary Intervention, *Medicine (Baltimore)* 95 (2016) e2687. <https://doi.org/10.1097/MD.0000000000002687>.
- [21] A.T. Malik, U.Y. Panni, B.A. Masri, S. Noordin, The impact of surgeon volume and hospital volume on post-operative mortality and morbidity after hip fractures: A systematic review, *Int J Surg* 54 (2018) 316–327. <https://doi.org/10.1016/j.ijsu.2017.10.072>.
- [22] E.J.A. Wiegers, C.A. Sewalt, E. Venema, N.W.L. Schep, J.A.N. Verhaar, H.F. Lingsma, D. Den Hartog, The volume-outcome relationship for hip fractures: a systematic review and meta-analysis of 2,023,469 patients, *Acta Orthop.* 90 (2019) 26–32. <https://doi.org/10.1080/17453674.2018.1545383>.
- [23] P.M. Dodek, Volume-outcome relationships in critical care. Understanding the mechanism, *Am J Respir Crit Care Med* 190 (2014) 601–603. <https://doi.org/10.1164/rccm.201401-0132ED>.
- [24] Gemeinsamer Bundesausschuss, Mindestmengenregelungen, 2020. <https://www.g-ba.de/themen/qualitaetssicherung/vorgaben-zur-qualitaetssicherung/vorgaben-mindestmengenregelungen/>.
- [25] Institut für Qualitätssicherung und Transparenz im Gesundheitswesen, Bericht zum Strukturierten Dialog 2019: Erfassungsjahr 2018, 2020. <https://iqtig.org/qs-berichte/sd/>.
- [26] M. Emmert, O. Schöffski, Berichterstattung über die Versorgungsqualität von Krankenhäusern im Raum Nürnberg: Zielsetzung, Vorgehensweise und kritische Betrachtung, Nürnberg, 2019.
- [27] Centers for Medicare & Medicaid Services, Design for Nursing Home Compare Five-Star Quality Rating System: Technical Users' Guide: Updated September 2019 with a description of changes to the Five-Star Quality Rating System that will be made in October 2019., 2019.
- [28] M.G. Olmsted, R. Powell, J. Murphy, D. Bell, M. Morley, M. Stanley, Methodology - U.S. News & World Report 2019-20 Best Hospitals: Specialty Rankings (2019).
- [29] S. Chimonas, E. Fortier, D.G. Li, A. Lipitz-Snyderman, Facts and Fears in Public Reporting: Patients' Information Needs and Priorities When Selecting a Hospital for Cancer Care, *Med Decis Making* 39 (2019) 632–641. <https://doi.org/10.1177/0272989X19855050>.
- [30] O.C. Damman, P. Spreeuwenberg, J. Rademakers, M. Hendriks, Creating compact comparative health care information: what are the key quality attributes to present for cataract and total hip or knee replacement surgery?, *Med Decis Making* 32 (2012) 287–300. <https://doi.org/10.1177/0272989X11415115>.

- [31] S. Fischer, S. Pelka, R. Riedl, Understanding patients decision-making strategies in hospital choice: Literature review and a call for experimental research, *Cogent Psychology* 2 (2015). <https://doi.org/10.1080/23311908.2015.1116758>.
- [32] M. Emmert, M. Schlesinger, Hospital Quality Reporting in the United States: Does Report Card Design and Incorporation of Patient Narrative Comments Affect Hospital Choice?, *Health Serv Res* (2016) Jun 20. <https://doi.org/10.1111/1475-6773.12519>.
- [33] M. Emmert, K. Kast, U. Sander, Characteristics and decision making of hospital report card consumers: Lessons from an onsite-based cross-sectional study, *Health Policy* (2019). <https://doi.org/10.1016/j.healthpol.2019.07.013>.
- [34] I.B. de Groot, W. Otten, J. Dijs-Elsinga, H.J. Smeets, J. Kievit, P.J. Marang-van de Mheen, Choosing between hospitals: the influence of the experiences of other patients, *Med Decis Making* 32 (2012) 764–778. <https://doi.org/10.1177/0272989X12443416>.
- [35] D.A. Marshall, K. Deal, B. Conner-Spady, E. Bohm, G. Hawker, L. Loucks, K.V. MacDonald, T. Noseworthy, How do patients trade-off surgeon choice and waiting times for total joint replacement: a discrete choice experiment, *Osteoarthr. Cartil.* 26 (2018) 522–530. <https://doi.org/10.1016/j.joca.2018.01.008>.
- [36] P.J. Marang-van de Mheen, J. Dijs-Elsinga, W. Otten, M. Versluijs, H.J. Smeets, R. Vree, W.J. van der Made, J. Kievit, The relative importance of quality of care information when choosing a hospital for surgical treatment: a hospital choice experiment, *Med. Decis. Making* 31 (2011) 816–827. <https://doi.org/10.1177/0272989X11386799>.
- [37] L.M. O'Hara, I. Caturegli, N.N. O'Hara, R.V. O'Toole, D.F. Dalury, A.D. Harris, T.T. Manson, What publicly available quality metrics do hip and knee arthroplasty patients care about most when selecting a hospital in Maryland: a discrete choice experiment, *BMJ Open* 9 (2019) e028202. <https://doi.org/10.1136/bmjopen-2018-028202>.
- [38] J.M. Reuter, C.A. Hutyra, C.S. Politzer, C.C. Calixte, D.J. Scott, D.E. Attarian, R.C. Mather, Characterizing Patient Preferences Surrounding Total Knee Arthroplasty, *JB JS Open Access* 3 (2018) e0017. <https://doi.org/10.2106/JBJS.OA.18.00017>.
- [39] M. Varkevisser, S.A. van der Geest, Why do patients bypass the nearest hospital? An empirical analysis for orthopaedic care and neurosurgery in the Netherlands, *Eur J Health Econ* 8 (2007) 287–295. <https://doi.org/10.1007/s10198-006-0035-0>.
- [40] A. Victoor, D.M.J. Delnoij, R.D. Friele, J.J.D.J.M. Rademakers, Determinants of patient choice of healthcare providers: a scoping review, *BMC Health Serv Res* 12 (2012) 272. <https://doi.org/10.1186/1472-6963-12-272>.
- [41] A.T. Yahanda, K.J. Lafaro, G. Spolverato, T.M. Pawlik, A Systematic Review of the Factors that Patients Use to Choose their Surgeon, *World J Surg* 40 (2016) 45–55. <https://doi.org/10.1007/s00268-015-3246-7>.
- [42] N.C. Zwijnenberg, M. Hendriks, E. Bloemendal, O.C. Damman, J.D. de Jong, D.M. Delnoij, J.J. Rademakers, Patients' Need for Tailored Comparative Health Care Information: A Qualitative Study on Choosing a Hospital, *J Med Internet Res* 18 (2016) e297. <https://doi.org/10.2196/jmir.4436>.
- [43] N. Zwijnenberg, O. Damman, P. Spreeuwenberg, M. Hendriks, J. Rademakers, Different patient subgroup, different ranking? Which quality indicators do patients find important when choosing a hospital for hip- or knee arthroplasty?, *BMC Health Serv Res* 11 (2011) 299. <https://doi.org/10.1186/1472-6963-11-299>.

#### Adresse | Kontakt

Bertelsmann Stiftung  
Carl-Bertelsmann-Straße 256  
33311 Gütersloh  
Telefon +49 5241 81-0

Uwe Schwenk  
Director Programm Gesundheit  
Telefon +49 5241 81-81418  
[uwe.schwenk@bertelsmann-stiftung.de](mailto:uwe.schwenk@bertelsmann-stiftung.de)

Weisse Liste gemeinnützige GmbH  
Werderscher Markt 6  
10117 Berlin

Johannes Strotbek, Hannah Wehling  
Senior Project Manager  
Telefon +49 30 275788-320, -326  
[johannes.strotbek@weisse-liste.de](mailto:johannes.strotbek@weisse-liste.de)  
[hannah.wehling@weisse-liste.de](mailto:hannah.wehling@weisse-liste.de)

[www.bertelsmann-stiftung.de](http://www.bertelsmann-stiftung.de)