



# Public Reporting im Gesundheitswesen

Eine systematische Ausarbeitung von Erfolgsfaktoren und  
Barrieren in Bezug auf das Design von Report Cards

# Public Reporting im Gesundheitswesen

Eine systematische Ausarbeitung von Erfolgsfaktoren  
und Barrieren in Bezug auf das  
Design von Report Cards

## **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)  
Dr. Benjamin Kolb  
Juli 2023

Titelbild: Adobe Stock #576980630

# Inhalt

<b>1</b>	<b>Hintergrund und Zielsetzung</b>	<b>4</b>
<b>2</b>	<b>Methodik der Untersuchung</b>	<b>4</b>
<b>3</b>	<b>Ergebnisse</b>	<b>6</b>
3.1	Systematische Literaturrecherche	6
3.2	Beschreibung der Primärstudien	7
3.2.1	Ambulanter Sektor bzw. Arztebene-bezogenes Public Reporting	7
3.2.2	Stationärer Sektor	15
3.2.3	Pflegesektor	35
3.3	Gesamtergebnis	41
3.3.1	Ambulanter Sektor bzw. Arztebene-bezogenes Public Reporting	41
3.3.2	Stationärer Sektor	43
3.3.3	Pflegesektor	44
3.4	Gegenüberstellung der Ergebnisse	45
<b>4</b>	<b>Zusammenfassung und Diskussion der Ergebnisse</b>	<b>47</b>
<b>5</b>	<b>Verzeichnisse</b>	<b>50</b>
5.1	Literaturverzeichnis	50
5.2	Abbildungsverzeichnis	53
5.3	Tabellenverzeichnis	53

## 1 Hintergrund und Zielsetzung

Public Reporting umfasst die Darstellung von (Qualitäts)Informationen über medizinische Leistungserbringer. Das umfasst sowohl niedergelassene Ärzte und Krankenhäuser aber auch Rehabilitationseinrichtungen und Pflegeheime. Public Reporting wird dabei nicht einheitlich betrieben, sondern es bestehen erhebliche Unterschiede im Umsetzen von Public Reporting, sowohl in Deutschland als auch international [1–8]. Zahlreiche Studien haben dabei gezeigt, dass es bei der Darstellung von Qualitätsinformationen über medizinische Leistungserbringer zu Verständnisschwierigkeiten kommen kann [9, 10]. Insbesondere das Design von Report Cards nimmt eine zentrale Bedeutung für die Verständlichkeit ein [5, 11, 12]. So hat beispielsweise eine Studie aus den USA zur Verständlichkeit der Report Card Hospital Compare gezeigt, dass nur 51% der Befragten das Krankenhaus mit den besten Qualitätsergebnissen gewählt hat [13]. Eine weitere Studie hatte bereits im Jahr 2014 gezeigt, dass die Verständlichkeit deutscher Public Reporting Instrumente bei ca. 61% lag [14].

Die Weisse Liste strebt eine möglichst verständliche Darstellung von Qualitätsinformationen für die Ausgestaltung der Arzt-, Krankenhaus- und Pflegeheimsuche an. Im Rahmen der folgenden Ausführungen erfolgt daher eine evidenzgestützte Ableitung von Handlungsempfehlungen zur Qualitätsdarstellung durch die systematische Identifizierung von Erfolgsfaktoren und Barrieren für das Public Reporting. Der Fokus soll auf dem Design der Darstellung liegen. Zielsetzung soll eine übersichtliche Darstellung von Empfehlungen sein, wie Public Reporting aus der Nutzerperspektive erfolgreich und verständlich umgesetzt werden kann. Es soll also herausgearbeitet werden, welche Darstellungsvarianten tendenziell zu höheren Verständniswerten führen. Hierzu wird unter Einhaltung wissenschaftlicher Standards eine systematische Literaturrecherche in den beiden Datenbanken PubMed (Public Medline) sowie der Cochrane Library durchgeführt. Eingeschlossen werden wissenschaftliche Studien der vergangenen 10 Jahre.

## 2 Methodik der Untersuchung

Im Rahmen der folgenden Ausführungen erfolgt eine evidenzgestützte Ableitung von Handlungsempfehlungen zur Qualitätsdarstellung durch die systematische Identifizierung von Erfolgsfaktoren und Barrieren für das Public Reporting. Es wird daher die Fragestellung adressiert: „Welche Public Reporting Gestaltungselemente zur Darstellung von Qualitätsdaten in Bezug auf Ärzte, Krankenhäuser und Pflegeeinrichtungen werden in der Literatur als Best Practice identifiziert?“ Der Überblick zu den in der Literatur empfohlenen Präsentationsmerkmalen wird mittels einer systematischen Literaturrecherche erarbeitet. Der Durchführung liegen die Vorgaben des PRISMA-Statements [15, 16] sowie dem Cochrane Handbook [17] zugrunde.

Zur Erarbeitung einer geeigneten Suchsyntax wurde im ersten Schritt nach bereits publizierten Übersichtsarbeiten zu Best Practice Evidenz zu Public Reporting von Qualitätsdaten durchgeführt. Grundlage hierfür waren die Suchstrategien vergangener, bekannter systematischer Übersichtsarbeiten zu einer ähnlichen Thematik von Sander et al. (2015) [10], Kurtzman et al. (2016) [9] sowie Faber et al. (2009) [18]. Die Syntax wurde jeweils in den Datenbanken Medline (via PubMed) und The Cochrane Library eingegeben und ausgeführt. Aus den genannten Quellen sowie weiteren ergänzenden Begriffen im Zusammenhang mit Public Reporting von Qualitätsdaten in der Gesundheitsversorgung wurde die im Folgenden dargestellte Suchsyntax entwickelt (siehe Tabelle 1).

Block 1	Block 2
<p>"public reporting"[Title/Abstract] OR "public release"[Title/Abstract] OR "public report"[Title/Abstract] OR "public reports"[Title/Abstract] OR "publicly reported"[Title/Abstract] OR "report card"[Title/Abstract] OR "report cards"[Title/Abstract] OR publicly report*[Title/Abstract] OR "reporting system"[Title/Abstract] OR "reporting systems"[Title/Abstract] OR "reporting instrument"[Title/Abstract] OR "reporting instruments"[Title/Abstract] OR "quality report"[Title/Abstract] OR "quality reports"[Title/Abstract] OR "quality reporting"[Title/Abstract] OR "consumer report"[Title/Abstract] OR "consumer reports"[Title/Abstract] OR "performance report"[Title/Abstract] OR "performance reports"[Title/Abstract] OR "performance information"[Title/Abstract] OR "patient rating"[Title/Abstract] OR "patient ratings"[Title/Abstract] OR "online rating"[Title/Abstract] OR "online ratings"[Title/Abstract] OR "web-based ratings"[Title/Abstract] OR "physician rating"[Title/Abstract] OR "physician ratings"[Title/Abstract] OR "hospital rating"[Title/Abstract] OR "hospital ratings"[Title/Abstract] OR "doctor rating"[Title/Abstract] OR "doctor ratings"[Title/Abstract] OR "nurse rating"[Title/Abstract] OR "nurse ratings"[Title/Abstract] OR "review site"[Title/Abstract] OR "review sites"[Title/Abstract] OR "review website"[Title/Abstract] OR "review websites"[Title/Abstract] OR "comparative report"[Title/Abstract] OR "comparative reports"[Title/Abstract] OR "public disclosure"[Title/Abstract] OR "information dissemination"[Title/Abstract]</p>	<p>presentation[Title/Abstract] OR format[Title/Abstract] OR numeracy[Title/Abstract] OR design[Title/Abstract] OR ("information literacy"[Title/Abstract] OR ("information"[Title/Abstract] AND "literacy"[Title/Abstract])) OR "choice behavior"[Title/Abstract] OR "decision making"[Title/Abstract] OR consumer behavior[MeSH] OR information services[MeSH] OR quality indicators, health care[MeSH] OR delivery of health care[MeSH] OR decision making[MeSH] OR health literacy[MeSH]</p>
AND	

Tabelle 1: Finale Suchsyntax für die Datenbanksuche auf PubMed und The Cochrane Library

Zusätzlich zur Datenbankenrecherche erfolgte eine Handsuche auf Google Scholar sowie in Referenzlisten der identifizierten Artikel. Die Suchergebnisse der unterschiedlichen Datenquellen wurden anschließend in das Literaturverwaltungsprogramm Citavi importiert und verwaltet. Hierbei wurden zunächst Duplikate entfernt. Alle verbleibenden Suchtreffer wurden anschließend von zwei Forschern (ME, BK) unabhängig voneinander entsprechend der in Tabelle 2 genannten Ein- und Ausschlusskriterien überprüft.

Die Relevanz der Treffer wurde zunächst auf Basis von Titel und Abstract und anschließend auf Basis des Volltextes beurteilt. Unklarheiten bzw. unterschiedliche Einschätzungen wurden im Konsensverfahren geklärt.

	Einschluss	Ausschluss
Zeitraum	2009 – 2019	Vor 2009
Sprache	Deutsch, Englisch	Alle Sprachen außer Deutsch und Englisch
Art der Publikation	Artikel aus Zeitschriften mit Peer-Review-Verfahren und Experteneinschätzungen	Kommentare, Guidelines, Interviews, Briefe, Vorlesungen, Nachrichten, Konferenzartikel (sofern diese keinem Peer-Review-Verfahren unterlagen), Buchbesprechungen, Editorials etc.

Inhalt	Gestaltungselemente zur Darstellung von Qualitätsinformationen mit Bezug zu Ärzten, Krankenhäusern und Pflege auf Public Reporting Portalen im Internet	Gestaltungselemente zur Darstellung anderer Daten und Informationen (z.B. in Bezug auf Arzneimittel [19], Gesundheitsinformationen [20], Health Plans [21–23])
Analyseverfahren	Designelemente zur Darstellung von Qualitätsinformationen mussten im Analyseverfahren enthalten sein (z.B. als Variable im Regressionsmodell)	Designelemente wurden nicht explizit evaluiert. Studien, die beispielsweise nur den Einfluss soziodemografischer Variablen auf die Verständlichkeit untersuchen, hierbei aber keine Variablen zum Design aufnehmen [14]. Auch wurden Studien ausgeschlossen, die keine Ergebnisse bzw. Varianten nachvollziehbar erläuterten bzw. darstellten [24].
Studiendesign	Experimentelle Studien zur Untersuchung der Verständlichkeit von Report Cards	Keine experimentellen Studien (z.B. Expertenmeinungen, Präferenzbefragungen)

*Tabelle 2: Ein- und Ausschlusskriterien*

Die eingeschlossenen Studien wurden jeweils einzeln inhaltlich analysiert und die Ergebnisse diskutiert. Hierbei wurden zu jeder Studie jeweils der Hintergrund sowie das methodische Vorgehen dargelegt. Der Fokus lag auf der Darstellung der Ergebnisse zur Verständlichkeit sowie den Handlungsempfehlungen, die sich aus der jeweiligen Studie ergeben. Im Anschluss daran wurden die Studien zusammenfassend dargestellt und Handlungsempfehlungen für die Qualitätsdarstellung auf der Weissen Liste abgeleitet.

## 3 Ergebnisse

### 3.1 Systematische Literaturrecherche

Auf den beiden Datenbanken konnten insgesamt 3.750 Treffer erzielt werden (vgl. Abbildung 1). Nach Entfernung der Duplikate und der Durchsicht von Titel und Abstract konnten 202 potentiell relevante Publikationen identifiziert werden. Nach einer Volltextanalyse wurden insgesamt 23 Studien als relevant für diese Literaturrecherche eingestuft, davon 20 Primärstudien und 3 Sekundärstudien. Die 20 Primärstudien beziehen sich sowohl auf die Darstellung individueller Ärzte (N=6), Krankenhäuser (N=11) und Pflegeheime (N=3); diese werden nun im folgenden Kapitel einzeln beschrieben und zusammengefasst. Die Ergebnisse der 3 Sekundärstudien werden zur Diskussion der hier abgeleiteten Empfehlungen herangezogen.

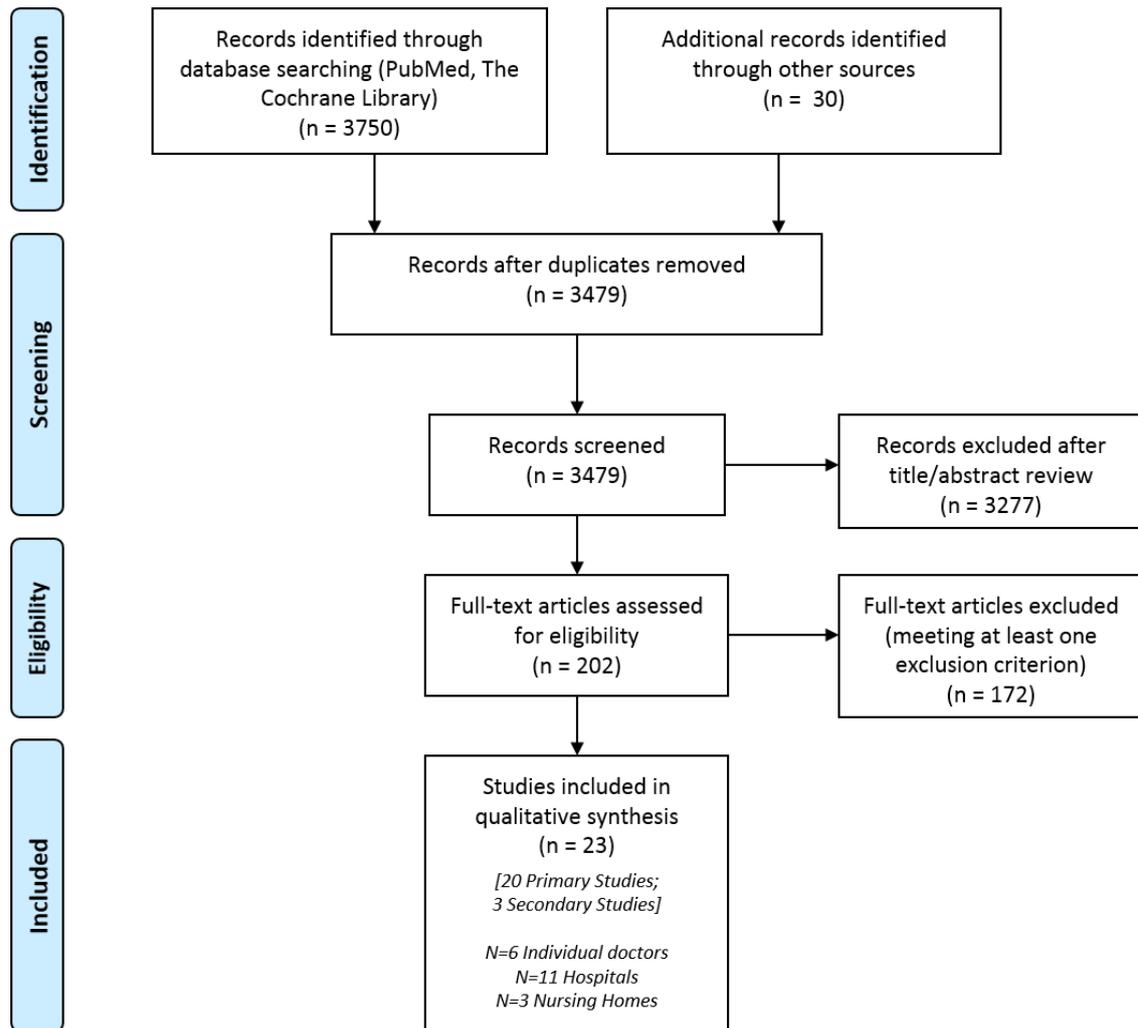


Abbildung 1: Darstellung der Literaturrecherche mittels PRISMA Flow Diagram [25]

## 3.2 Beschreibung der Primärstudien

### 3.2.1 Ambulanter Sektor bzw. Arzzebene-bezogenes Public Reporting

**Cerully J et al. (2018) *Improving Patients' Choice of Clinician by Including Roll-up Measures in Public Healthcare Quality Reports: an Online Experiment* [26]**

In der Studie von Cerully et al. wurde untersucht, wie sich die Darstellung von Qualitätsindikatoren mit unterschiedlichem Detaillierungsgrad auf die Arztauswahl von Patienten auswirkt. Dafür wurde ein Experiment durchgeführt, in dem Probanden eines von drei entwickelten Datensets vorgelegt wurde. Die Datensets enthielten jeweils entweder aggregierte Summenscores (Roll-up measures), detaillierte Ergebnisse zu einzelnen Indikatoren (Drill-down measures) oder beides gleichzeitig zu jeweils 12 Ärzten (siehe Abbildung 2). Ebenfalls wurde getestet, ob sich die Ergebnisse von erkrankten Patienten im Vergleich zu gesunden Patienten unterschieden. Insgesamt 550 Personen nahmen an der durchgeführten Onlinebefragung teil.

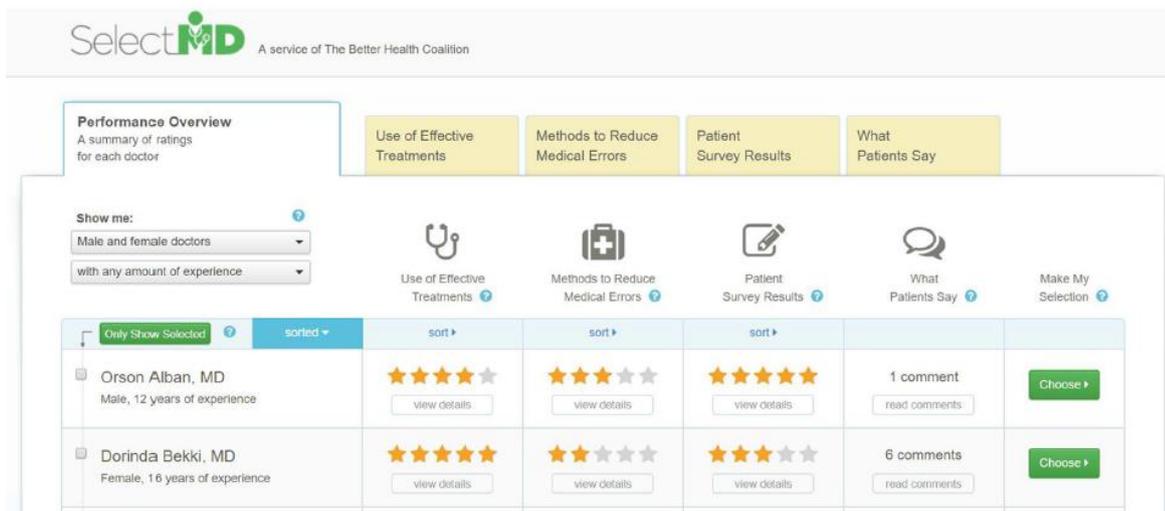


Abbildung 2: Sample screenshot from the SelectMD experimental website

Als ein erstes Ergebnis der Studie konnte festgestellt werden, dass Probanden, die detaillierte Ergebnisse zu mehreren Indikatoren mit oder ohne zusätzlichen Summenscore sahen (MW der Anzahl an Aktivitäten auf der Report Card<sup>1</sup>: 19,2 bzw. 14,9), wesentlich aktiver waren als Probanden, die lediglich die Summenscores (Roll-up measures) sahen (MW: 10,5;  $p < .05$ ) (s. Tabelle 3). Dieses Ergebnis war unabhängig vom Gesundheitszustand; lediglich bei kränkeren Probanden wurden Unterschiede zwischen der detaillierten Ergebnisdarstellung und Summenscores gezeigt. Als Ergebnis bezüglich der Verständlichkeit konnte festgestellt werden, dass diejenigen Probanden, die ausschließlich die detaillierten Ergebnisse sahen, im Vergleich zu den anderen Probanden signifikant häufiger einen suboptimalen Arzt wählten (36,3% vs. 25,6% bzw. 23,4%;  $p < .05$ ). Das gleiche Ergebnis wurde bezüglich der Übereinstimmung mit zuvor genannten Präferenzen für die Arztwahl und tatsächlicher Arztwahl festgestellt. Insgesamt wählten 51,2% der Probanden, die ausschließlich die detaillierten Ergebnisse sahen, einen nicht ihren eigentlichen Präferenzen entsprechenden Arzt; unter den anderen Probanden waren dies 47,5% bzw. 45,6% ( $p < .05$ ). Auf der Basis der Ergebnisse empfehlen die Autoren die Verwendung von Summenscores (Summary Measures, Roll-up measures), unabhängig davon, ob diese mit detaillierteren Ergebnissen zu einzelnen Indikatoren ergänzt werden oder nicht („These findings suggest that it is advantageous to include rollup scores in reports on healthcare quality with or without accompanying drill-down information“).

<sup>1</sup> [...] the number of concrete actions taken (e.g., clicking on items, hovering over a pop-up element)

Categories of quality	Measure (scale)	Mean (SD) by experimental arm		
		Drill-down plus roll-up (n = 176)	Roll-up (n = 184)	Drill-down (n = 190)
Amount of effort	Time deliberating (number of seconds)	437.6 (482.9)	373.9 (321.7)	450.4 (724.6)
	Active exploring of website (number of actions)	19.2 <sup>a</sup> (28.5)	10.5 <sup>ab</sup> (9.5)	14.9 <sup>b</sup> (19.0)
Quality of decision	Proportion of times clinician choice did not match with stated preferences (percentage)	47.5 <sup>a</sup> (14.7)	45.6 <sup>b</sup> (16.0)	51.2 <sup>ab</sup> (14.5)
	Choice of a suboptimal clinician (percentage)	25.6 <sup>a</sup>	23.4 <sup>b</sup>	36.3 <sup>ab</sup>

*Means/percentages within the same row sharing the same superscript are significantly different ( $p < 0.05$ ) from one another*

*Tabelle 3: Effects of Quality Data Presentation on Decision Quality*

### **Donelan K et al. (2011) Consumer Comprehension of Surgeon Performance Data for Coronary Bypass Procedures [27]**

In der Studie von Donelan et al. wurden 337 Personen im Alter von mindestens 18 Jahren befragt. Für die Befragung wurden vier verschiedene Präsentationen zu Qualitätsdaten von fünf unterschiedlichen Herzchirurgen entwickelt. Dabei gab es bei jeder Präsentation einen absolut besten Chirurgen, den die Personen erkennen auswählen sollten. Es wurde bei den Präsentationen insbesondere die Darstellung von Tabellen mit ausschließlich Text, Tabellen mit Text und Symbolen (die als Interpretationshilfe die Ergebnisse hervorheben) sowie die Darstellung mit einem Balkendiagramm untersucht. Die Qualitätsmerkmale waren die Fallzahl, die beobachtete Todesrate, die erwartete Todesrate, die risikoadjustierte Todesrate, das 95%-Konfidenzintervall für die risikoadjustierte Todesrate, die Rate der Wiedereinweisung sowie die durchschnittliche Länge des Krankenhausaufenthalts (siehe Abbildung 3).

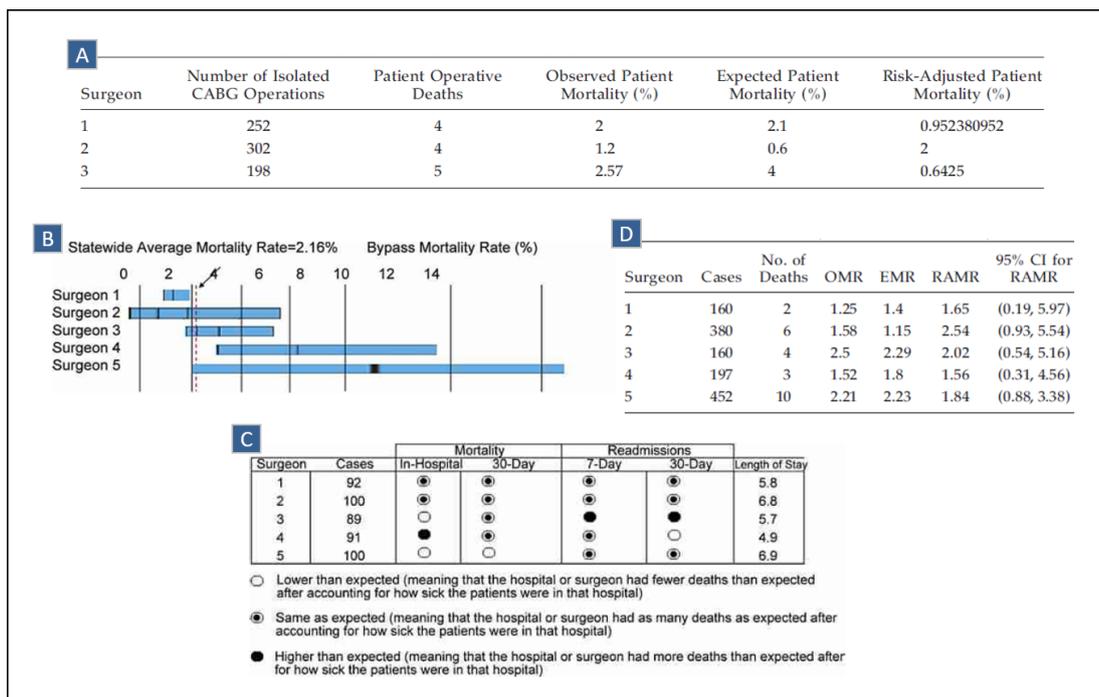


Abbildung 3: Presentation formats a-d

Als Ergebnis zeigte sich, dass die Verbindung von Text und Symbolen (66%, Typ C) sowie die Darstellung mit Balkendiagrammen (53%, Typ B) am häufigsten zur Auswahl des Chirurgen mit der niedrigsten risikoadjustierten Mortalitätsrate führte (siehe Tabelle 4). Am schlechtesten schnitten die beiden Darstellungen mit ausschließlich zahlenbasierten Informationen in Tabellenform ab (22% für Typ D bzw. 16% für Typ A). Nur 6,4% der Probanden wählten bei allen vier Präsentationen den Chirurgen mit der niedrigsten risikoadjustierten Todesrate, 16,8% bei zumindest drei Präsentationen. Die Autoren empfehlen entsprechend der von den Probanden geäußerten unterschiedlichen Präferenzen bezüglich verschiedener Präsentationsformen, dass Report Cards die Präsentation der Daten in unterschiedlicher Form anbieten sollten. Grafische Darstellungen sind reinen text- bzw. zahlenbasierten Tabellen vorzuziehen. Aus Sicht der Autoren bedarf die Darstellung der risikoadjustierten Todesrate weiterer Forschung. Für die Probanden scheint dies schwer verständlich und deshalb nicht relevant zu sein.

	Surgeon 1 (%)	Surgeon 2 (%)	Surgeon 3 (%)	Surgeon 4 (%)	Surgeon 5 (%)	Need More Information (%)
Display A	14	44	16 <sup>a</sup>	N/A	N/A	26
Display B	53 <sup>a</sup>	15	1	1	6	25
Display C	7	3	3	4	66 <sup>a</sup>	17
Display D	32	10	2	22 <sup>a</sup>	13	21

Surgeon preference for each display is shown ("which surgeon would you choose given the information shown"). <sup>a</sup>Indicates the surgeon with the lowest risk-adjusted mortality in the display.

N/A = not available.

Tabelle 4: Surgery Results: Surgeon Preference for Display Type

**Hibbard et al. (2012) An Experiment Shows That A Well-Designed Report On Costs And Quality Can Help Consumers Choose High-Value Health Care [28]**

Die Studie von Hibbard und Kolleginnen greift das Zusammenspiel der Präsentation von Kosten- und Qualitätsinformationen über Leistungserbringer auf. Es wird darauf hingewiesen, dass die Zielsetzung des Experiments die

Wahl sog. high-value Leistungserbringer sein solle. Diese zeichnen sich durch eine gute Qualität bei gleichzeitig niedrigen Kosten aus. Es wurde daher untersucht, inwieweit das Präsentationsformat von Qualitäts- und Kosteninformationen zu Ärzten und auch Kliniken die Wahrscheinlichkeit erhöht, dass Patienten eine sog. high-value Auswahl treffen. Die Probanden wurden randomisiert in drei verschiedene Gruppen eingeteilt. Die Teilnehmer einer jeden Gruppe sahen 6 unterschiedliche Präsentationen zu Kosteninformationen und Ressourcenverbräuchen von Ärzten und Krankenhäusern. Die Darstellung der Informationen wurde hierbei variiert („Each group was presented with the same descriptive data on health care providers, but the data were displayed, labeled, or framed differently“). Eine der Aufgaben im Rahmen des Experiments bestand in der Wahl des jeweils besten Leistungserbringers; d.h. welchen Leistungsanbieter würden sie als qualitativ besten identifizieren. Die beste Wahl bestand hierbei in dem Leistungserbringer mit der höchsten Qualität und den gleichzeitig niedrigsten Kosten. In den Präsentationen wurden die Kosten mit Sternen, Dollarzeichen oder Dollarangaben dargestellt bei gleichzeitiger Abbildung von keinen, schwachen oder starken Qualitätssignalen (siehe Abbildung 4):

- Kein Qualitätssignal: keine Informationen über Qualität, lediglich über den Zugang der Versorgung
- Schwaches Qualitätssignal: schwer zu verstehende und verarbeitende Qualitätsinformationen
- Starkes Qualitätssignal: einfach zu verstehende und verarbeitende Qualitätsinformationen

Die Kosteninformationen wurden dabei in Form von Kreisen (hier allerdings als „Stars“ bezeichnet), absoluten Angaben in Dollar sowie Dollarzeichen dargestellt (siehe Darstellung linke Seite). Insgesamt wurden im Rahmen der Studie 1.421 Teilnehmer befragt.

EXHIBIT 1				
Examples Of Cost Data With No Quality Signal Presented To Consumers				
Doctor	Access data			Cost data
	Saturday hours	Driving distance (miles)	Same-day office visits	
<b>GROUP 1*</b>				
Dr. White	9:00-noon	6	No	●●●
Dr. Ramsey	9:00-3:00	5	Yes	●●●●
Dr. Abbot	None	10	Yes	●●●●●
Dr. James	9:00-3:00	5	Yes	●●●●
Dr. Albright	None	8	No	●●●●●
Dr. Casey	9:00-noon	7	Yes	●●●●
<b>GROUP 2*</b>				
Dr. White	9:00-noon	6	No	\$150
Dr. Ramsey	9:00-3:00	5	Yes	\$95
Dr. Abbot	None	10	Yes	\$180
Dr. James	9:00-3:00	5	Yes	\$195
Dr. Albright	None	8	No	\$110
Dr. Casey	9:00-noon	7	Yes	\$140
<b>GROUP 3*</b>				
Dr. White	9:00-noon	6	No	⚡
Dr. Ramsey	9:00-3:00	5	Yes	⚡
Dr. Abbot	None	10	Yes	⚡⚡
Dr. James	9:00-3:00	5	Yes	⚡⚡
Dr. Albright	None	8	No	⚡
Dr. Casey	9:00-noon	7	Yes	⚡

SOURCE Authors' analysis. NOTES The three groups received the same information, but the cost data were presented differently, as shown. The groups received access data but no quality data. \*Cost data presented as 1-3 stars (represented here by small circles), with the label "careful with your health care dollars." \*Cost data presented as dollar amounts, with the label "average cost for office visit." \*Cost data presented as 1-3 dollar signs, with the label "average cost for office visit."

EXHIBIT 2				
Examples Of Cost Data Presented To Respondents With A Weak Quality Signal				
Doctor	Quality data (weak signal)			Cost data
	Diabetes patients tested for blood sugar (%)	Uses electronic health record	Patients say office staff is helpful (%)	
Dr. Pieters	75	Yes	80	●●●
Dr. Rabin	75	Yes	80	●●●

SOURCE Authors' analysis. NOTES Groups 2 and 3 received the same information, but the cost data were presented differently, as shown in Exhibit 1. Each group received information about six doctors, as explained in the text. The information included the following key: 3 stars is very careful (lower costs); 2 stars is somewhat careful (average costs); 1 star is less careful (higher costs). In this exhibit, small circles denote stars.

EXHIBIT 3				
Examples Of Cost Data Presented To Respondents With A Strong Quality Signal				
Doctor	Quality data (strong signal)			Cost data
	Uses treatments proven to get results	Has safeguards to protect patients from medical errors	Is responsive to patients' needs and preferences	
Dr. Friedman	Better	Better	Better	●
Dr. Hunter	Better	Better	Better	●●●

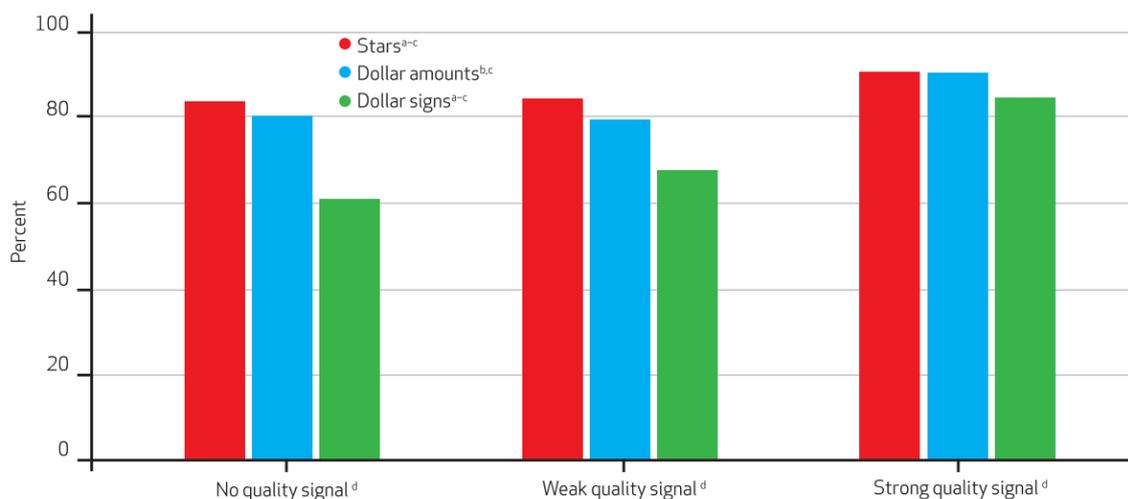
SOURCE Authors' analysis. NOTES Groups 2 and 3 received the same information, but the cost data were presented differently, as shown in Exhibit 1. Each group received information about six doctors, as explained in the text. The information included the following key: 3 stars is very careful (lower costs); 2 stars is somewhat careful (average costs); 1 star is less careful (higher costs). In this exhibit, small circles denote stars.

Abbildung 4: Examples Of Cost Data Presented To Respondents With No, A Weak Or Strong Quality Signal

Als Ergebnis zeigte sich, dass bei der Darstellung von Kostendaten insbesondere die Darstellung in Kreis- bzw. Sternenform sowie absoluten Dollarangaben als effektiv erwiesen haben, unabhängig davon, ob keine, schwache oder starke Qualitätssignale zu sehen waren (s. Abbildung 5). Bei starken Qualitätssignalen waren die Unterschiede zwischen den Darstellungsformen allerdings geringer. Die Darstellung in Dollarzeichen erwies sich als die am wenigsten effektive Form der Darstellung. Darüber hinaus steigt das Vertrauen in die Auswahl, je stärker das Qualitätssignal ist. Die Studienergebnisse verdeutlichen insgesamt die Herausforderung der gleichzeitigen Darstellung von Qualitäts- und Kostendaten. So zeigte sich beispielsweise, dass Kosteninformationen nur unwahrscheinlich herangezogen und interpretiert werden, wenn keine starken Qualitätssignale von der Darstellung

ausgehen. Des Weiteren zeigte sich auch, dass erläuternde bzw. interpretierende Labels die Interpretation von Darstellungen im Vergleich zur ausschließlichen Darstellung mittels vergleichender Zahlen positive Effekte auf die Verständlichkeit haben. Die Autoren empfehlen daher, Qualitätssignale möglichst stark darzustellen und Kosten in Form von Kreisen bzw. Sternen oder als absoluten Betrag darzustellen.

**Respondents Selecting High-Value Providers, By Type Of Cost Data And Quality Signal**



**SOURCE** Authors' analysis. **NOTES** Group 1 of the respondents saw cost data presented as 1–3 stars, with the label "careful with your health care dollars." Group 2 saw the data presented as dollar amounts, with the label "average cost for office visit." Group 3 saw the data presented as 1–3 dollar signs, with the label "average cost for office visit." <sup>a</sup> $p < 0.05$  for within-subject differences between no quality signal and weak quality signal. <sup>b</sup> $p < 0.05$  for within-subject differences between no quality signal and strong quality signal. <sup>c</sup> $p < 0.05$  for within-subject differences between weak and strong quality signals. <sup>d</sup> $p < 0.05$  for between-group differences.

Abbildung 5: Respondents Selecting High-Value Providers, By Type Of Cost Data And Quality Signal

### **Kanouse et al. (2016) How Patient Comments Affect Consumers' Use of Physician Performance Measures [29]**

In dieser Studie wurde untersucht, wie sich die Integration von Patientenkommentaren bei der Berichterstattung zur Qualität ambulanter Leistungserbringer zusammen mit Qualitätsindikatoren (HEDIS Measures) sowie standardisierten Patientenbefragungsergebnissen (CAHPS) auf die Qualität der Auswahlentscheidung auswirkt. Dafür wurden insgesamt 848 Personen befragt, welchen jeweils randomisiert eine von sechs Versionen einer Arztbewertungsplattform zugewiesen wurde. Diese unterschieden sich ausschließlich hinsichtlich der Art und der Menge dargestellter Informationen (CAHPS/HEDIS/Comments) sowie in der Anzahl der dargestellten fiktiven ambulant tätigen Ärzte (12 vs. 24 Ärzte).

Als Ergebnis zeigte sich, dass Probanden, die lediglich die standardisierten Indikatorenergebnisse für 12 Ärzte sahen, mit 61,2% (CAHPS) bzw. 61,6% (CAHPS+HEDIS) am häufigsten den besten Arzt auswählten (siehe Tabelle 5). Wurden die Patientenkommentare zusätzlich integriert, verringerte sich dieser Wert signifikant auf 49,3% (CAHPS) bzw. 43,7% (CAHPS+HEDIS) ab. Bei einer weiteren Erhöhung der Anzahl angezeigter Ärzte von 12 auf 24 konnte eine weitere Reduzierung der richtigen Auswahlentscheidungen festgestellt werden (34,6% bzw. 37,0%). Auch diese Reduzierung war signifikant. Der gleiche Trend, nur entgegengesetzt, konnte beim Anteil der Auswahl dominierter Ärzte festgestellt werden. Allerdings kann auch angeführt werden, dass sich durch die Integration der Patientenkommentare sowohl die auf den Report Cards verbrachte Zeit als auch die Anzahl der durchgeführten Handlungen signifikant steigerte. Die Autoren schlussfolgern aus den Ergebnissen, dass eine geringere Informationskomplexität durch weniger Ergebnisdaten und weniger angezeigten Leistungserbringern zu

einer höheren Entscheidungsqualität führt. Darüber hinaus ist zu überlegen, welche Qualitätsinformationen gleichzeitig angezeigt werden. Durch die Aufnahme der Patientenkommentare in die Darstellungen wurden standardisierte Qualitätsergebnisse von den Probanden weniger berücksichtigt, was mit einer Verringerung der Entscheidungsqualität einherging.

Experimental Measures	Experimental Arm					
	Without Comments (12 Doctors)		With Comments (12 Doctors)		With Comments (24 Doctors)	
	Arm 1 (CAHPS)	Arm 2 (CAHPS + HEDIS)	Arm 3 (CAHPS)	Arm 4 (CAHPS + HEDIS)	Arm 5 (CAHPS)	Arm 6 (CAHPS + HEDIS)
				Engagement		
Time spent on website [s (ln s)]	311.4 (5.45)	345.8 (5.52)	421.5 (5.72)**	489.6 (5.77)*	514.7 (5.87)**	517.7 (5.72)**
No. actions on website	3.6	5.2	7.8*	10.7*	9.0*	9.1*
Time spent probing standardized measures [s (ln s)]	42.5 (4.14)	60.6 (4.36)	22.6 (3.56)**	45.5 (4.21)	30.6 (3.83)	50.8 (4.10)
Actions probing standardized measures	0.7	1.4	0.6	1.1	0.7	1.2
Percentage of time on website probing CAHPS	11.8	9.0	5.4**	5.8	9.4	5.5
Percentage of time on website probing HEDIS	—	10.1	—	3.3**	—	5.8**
				Perceived usefulness		
Usefulness of website <sup>†</sup>	78.3	74.4	80.9	81.0	83.2	77.4
Satisfaction with choice of clinicians <sup>†</sup>	67.4	69.6	72.4	67.6	77.4	71.9
Usefulness of CAHPS <sup>‡</sup>	89.9	85.6	84.2	86.6	87.1	84.9
Usefulness of HEDIS <sup>‡</sup>	—	89.6	—	88.0	—	84.2
Trustworthiness of CAHPS <sup>‡</sup>	82.2	83.2	80.3	82.4	83.0	84.9
Trustworthiness of HEDIS <sup>‡</sup>	—	84.8	—	83.8	—	86.3
				Quality of choice (%)		
Selected best clinician <sup>§</sup>	61.2	61.6	49.3*	43.7*	34.6*	37.0*
Selected lower performing (dominated) clinician	—	17.6	—	37.3*	—	50.0*

<sup>†</sup>Entries are the percentages of participants who report that they definitely or probably would use.

<sup>‡</sup>Entries are the percentages of participants who respond very or somewhat satisfied/trustworthy/useful.

<sup>§</sup>Best = highest average score on standardized performance metrics.

CAHPS indicates Consumer Assessment of Healthcare Providers and Systems; HEDIS, Healthcare Effectiveness Data and Information Set.

\*Statistically significant difference from the parallel arm (1 or 2) without patient comments ( $P < 0.05$ ).

\*\*Statistically significant difference from the parallel arm (1 or 2) without patient comments ( $P < 0.01$ ).

\*\*\*Statistically significant difference from the parallel arm (1 or 2) without patient comments ( $P < 0.001$ ).

Tabelle 5: Engagement, Perceived Usefulness of Comparative Performance Information, and Quality of Choices When Presented With and Without Patient Comments

### Kenny P et al (2017) Choosing a Doctor: Does Presentation Format Affect the Way Consumers Use Health Care Performance Information? [30]

Die Studie von Kenny und Kollegen untersucht den Einfluss von Präsentationsformaten auf die qualitätsbasierte Auswahl eines Allgemeinarztes. Dafür wurden 1.208 Personen eines Onlinepanels befragt. Jedem Probanden wurden randomisiert zwei Präsentationen von Qualitätsdaten zu zwei fiktiven Hausarztpraxen vorgelegt. Die vier variierenden Designmerkmale waren (1) Häufigkeiten und Prozentangaben mit einem Symbol, (2) Sternebewertung, (3) Sternebewertung mit Wort-Symbol („Besser“, „Durchschnitt“, „Schlechter“) sowie (4) nur Prozentangaben (siehe Abbildung 6).

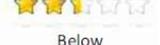
Level	Style 1	Style 2	Style 3	Style 4
1	 5 in 10 (50%)		 Below National average	50%
2	 7 in 10 (70%)		 Average	70%
3	 9 in 10 (90%)		 Above National average	90%

Abbildung 6: Presentation formats used for practice quality attributes

Insgesamt wählten 10,4% der Befragten bei zumindest einem der beiden Auswahlentscheidungen (siehe oben) den schlechteren Arzt aus. Bei Formaten mit ausschließlich Prozentangaben wurde in 8,5% der Fälle die schlechtere Alternative gewählt. Mit 4,4-5,3% fiel dieser Wert bei den Formaten 1-3 niedriger aus. Sahen Probanden nur Prozentangaben, erhöhte sich die Wahrscheinlichkeit für eine schlechte Entscheidung im Vergleich zu den Angaben von Häufigkeiten mit einem Symbol um den Faktor 1,7 ( $p < .05$ ), im Vergleich zu Sternebewertungen um 1,9 ( $p < .05$ ) und im Vergleich zu Sternebewertungen mit Wort-Symbolen um 2,2

( $p < .05$ ). Das erstgenannte Format (Style 1) wurde von 64% der Befragten als bestes Format angegeben und war damit unter allen Formaten das beliebteste, gefolgt von Style 4 (50%), Style 2 (45%) und Style 3 (41%). Die Autoren schlussfolgern in Bezug auf die Darstellung von Qualitätsinformationen, dass diese graphisch mit Symbolen oder Sternbewertungen dargestellt werden sollten. Numerische Angaben (z.B. Prozentwerte) können mit Symbolen graphisch unterstützt werden. In Bezug auf die beste Auswahl über alle Beobachtungen hinweg hat sich das Format Häufigkeiten und Prozentangaben mit einem Symbol am Besten bewährt (Tabelle 6). Als zentrale Implikation schlussfolgern die Autoren schließlich, dass die „Presentation of performance quality information for consumers will be most useful if presented as frequency and percentage with icons or star ratings.“ Numerische Ergebnisse sollten demnach mit einer grafischen Erläuterung verbunden werden.

	Model 1 Choosing worst alternative			Model 2 Difficult to understand				
	Odds ratio	95% CI	<i>p</i>	Odds ratio	95% CI	<i>p</i>		
Style 2	0.891	0.519	1.530	0.677	0.931	0.714	1.214	0.597
Style 3	0.763	0.463	1.256	0.287	0.976	0.745	1.278	0.860
Style 4	1.678	1.038	2.713	0.035	0.980	0.720	1.336	0.900
Choice time <sup>b</sup> <3 min	4.282	2.756	6.655	<0.001	2.328	1.701	3.188	<0.001
High numeracy <sup>c</sup>	0.573	0.377	0.870	0.009	0.905	0.685	1.196	0.484
Second style seen	1.113	0.784	1.581	0.550	0.988	0.848	1.150	0.874
Age								
30–44	0.620	0.404	0.949	0.028	0.661	0.463	0.943	0.022
45–59	0.297	0.161	0.548	<0.001	0.495	0.336	0.730	<0.001
60–74	0.256	0.123	0.534	<0.001	0.470	0.304	0.728	0.001
75 or more	0.441	0.139	1.406	0.166	0.502	0.223	1.129	0.096
Male	1.602	1.040	2.469	0.032	1.107	0.835	1.468	0.481
Constant	0.055	0.032	0.095	<0.001	0.253	0.176	0.365	<0.001
Hosmer–Lemeshow $\chi^2_{(8)}$	6.00				9.90			
<i>p</i>	0.65				0.27			

Robust standard errors were used to account for clustering by individual (2416 observations for 1208 respondent)

Model 1: Wald chi-squared test for 3 style coefficients  $p = 0.01$ ; Wald chi-squared test for 4 age coefficients  $p < 0.001$

Model 2: Wald chi-squared test for 3 style coefficients  $p = 0.95$ ; Wald chi-squared test for 4 age coefficients  $p = 0.002$

CI confidence interval

<sup>a</sup> Rated 4 or 5, where 1 = extremely easy and 5 = extremely difficult

<sup>b</sup> Time spent on all 18 choice questions

<sup>c</sup> Above sample median numeracy score of 4.66 on a scale of 1–6

*Tabelle 6: Logit models for evaluating presentation styles: choosing the worst alternative (model 1) and finding the choice information difficult to understand (model 2)*

### **Martino SC et al. (2017) Choosing Doctors Wisely: Can Assisted Choice Enhance Patients' Selection of Clinicians? [31]**

Im Rahmen der von Martino und Kollegen durchgeführten Studie wurden 688 Teilnehmer des US-GfK-Panels hinsichtlich des Wahlverhaltens von Probanden nach individuellen Ärzten untersucht. Die Report Cards wurden dabei folgendermaßen gestaltet: (1) eine Gesamtübersicht von Ärzten mittels quantitativer Qualitätskennzahlen sowie eine (2) Gesamtübersicht von Ärzten, die zusätzlich qualitative Patientenkommentare beinhaltet (siehe

Abbildung 7). Die dritte und vierte Darstellung bezog sich auf den Einfluss durch die Integration von Entscheidungshilfen („assisted-choice arms“). (3) Eine zweite, detaillierte Website enthielt eine Entscheidungshilfe mittels Labeling von Patientenkommentaren, d.h. das Thema des Patientenkommentars wurde kategorisch dargestellt (Tagged Comments). (4) Im vierten Studienarm wurde zudem ein erfahrener Experte im Umgang mit den Report Cards zur Seite gestellt, der telefonisch mit dem Probanden verbunden war (Tagged Comments plus trained Navigator). Die Darstellungen der vier Studienarme enthielten jeweils 12 Ärzte, zu denen die Teilnehmer randomisiert zugewiesen wurden. Zur Untersuchung der Verständlichkeit der Informationen bzw. der Report Card

wurden mehrere Kennzahlen verwendet, unter anderem die Arztwahl in Abhängigkeit der individuell angegebenen Präferenzen; d.h. die Übereinstimmung der Arztwahl mit den vom Proband genannten Präferenzen (sog. Preference congruent choices).

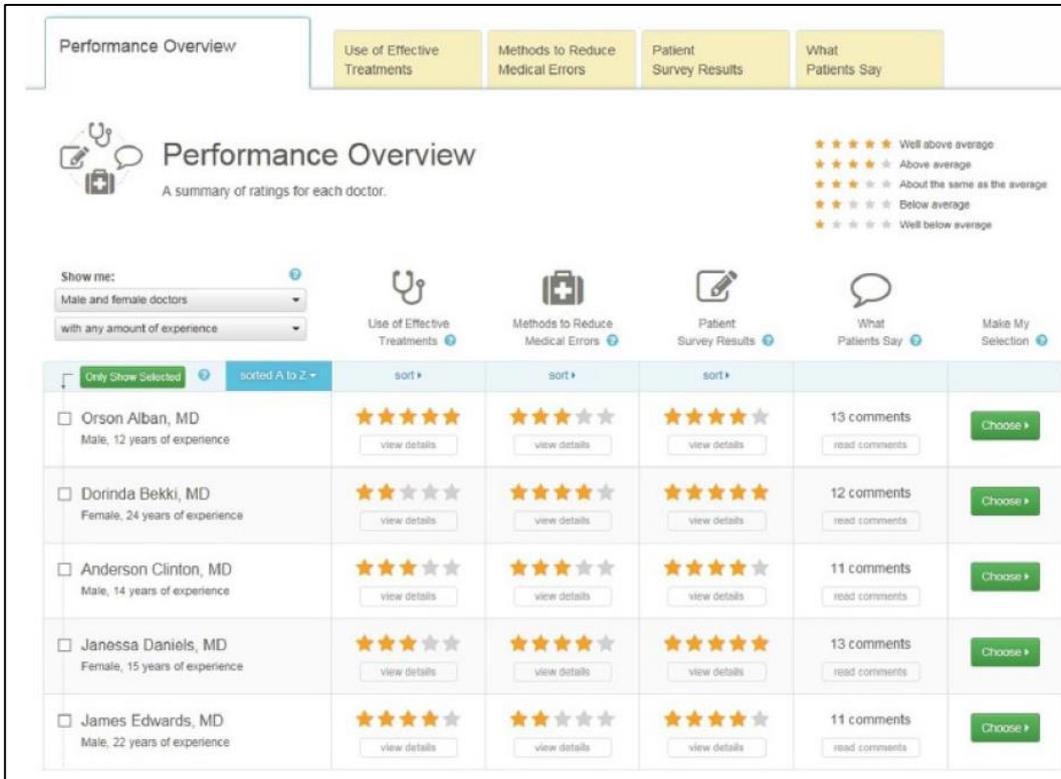


Abbildung 7: Overview page, SelectMD, Conventional Comments Arm. Figure A3.

Es zeigten sich insgesamt relativ niedrige Verständlichkeitswerte für die Arztwahl, welche zwischen 22% und 37% rangierten (No comments: 33.7%, Conventional comments: 21.6%, Tagged comments: 30.2%, Tagged comments + navigator: 36.7%;  $p < 0.05$ ). Auch wenn die Einführung bzw. Darstellung der Patientenkommentare das Engagement mit den Qualitätsinformationen erhöhte, so zeigte sich eine Abnahme der Entscheidungsqualität. Sollten Patientenkommentare dargestellt werden, so führten die beiden getesteten Entscheidungshilfen (Tagged Comments, Tagged Comments plus trained Navigator) zu verbesserten Entscheidungen.

### 3.2.2 Stationärer Sektor

**Emmert M et al. (2019) Characteristics and decision making of hospital report card consumers: lessons from an onsite-based cross-sectional study [6]**

Im Rahmen der Untersuchung wurden im Jahr 2017 insgesamt 635 Besucher der Krankenhaussuche der Report Card Weisse Liste zu einer Befragung eingeladen, die an der Studie vollständig teilnahmen. Zielsetzung der Untersuchung war mehr über die Nutzergruppe der Besucher der Weissen Liste zu erfahren, der Fokus der Studie lag auf der Verständlichkeit in Abhängigkeit der dargestellten Qualitätsinformationen und der Komplexität der Darstellung. Verglichen wurde die Verständlichkeit der aktuellen Darstellung auf der WL mit zwei leicht modifizierten Versionen der WL; einerseits wurde basierend auf Nutzerpräferenzen zusätzlich ein Composite Measure und

andererseits wurden die Krankenhäuser in Abhängigkeit der gewählten Anzahl und Reihenfolge der Krankenhäuser dargestellt. Die Komplexität wurde über die Anzahl der dargestellten Krankenhäuser variiert (5, 10, 15 Krankenhäuser) (siehe Abbildung 8).

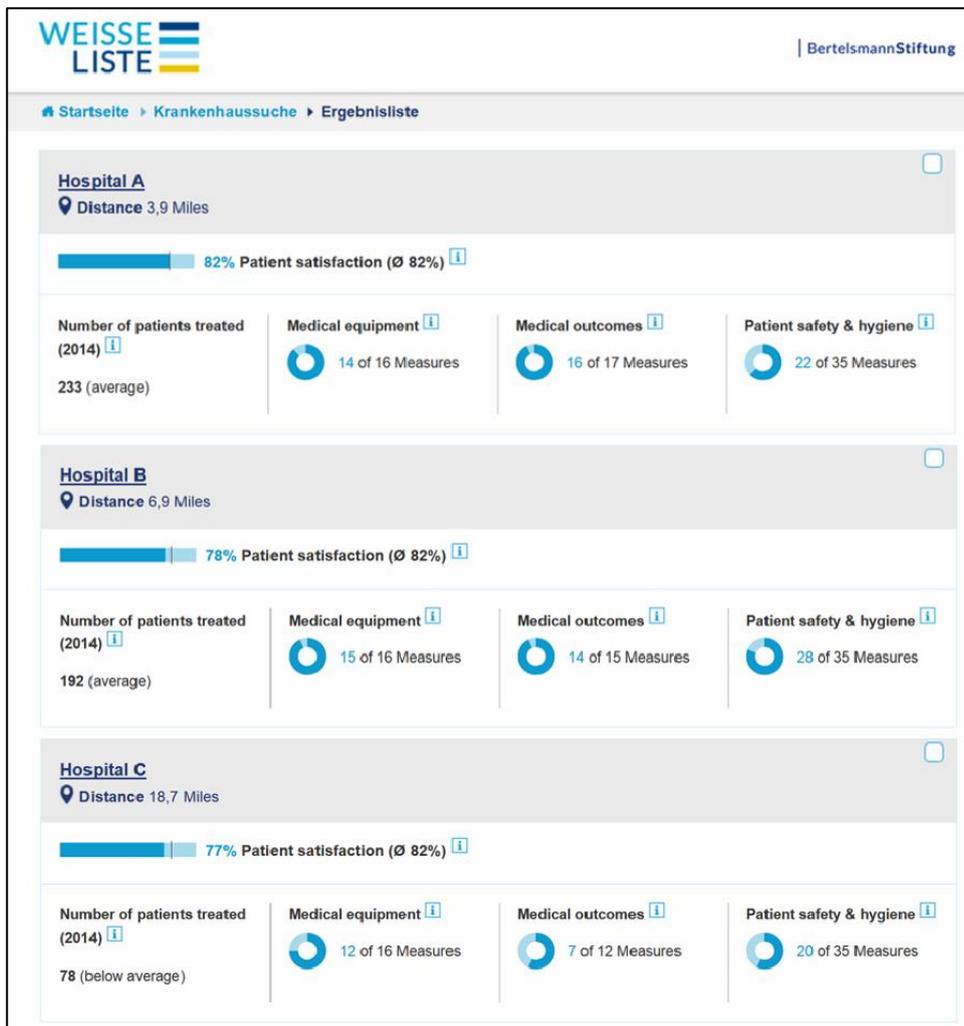


Abbildung 8: Experimental Report Card Design on Weisse Liste

Als Gesamtergebnis zeigte sich, dass die Weisse Liste insgesamt hohe Verständlichkeitswerte aufweist; insgesamt 81,5% aller Befragten wählten das jeweils beste Krankenhaus.

Weiterhin konnte festgestellt werden, dass die Verständlichkeit mit zunehmender Komplexität, d.h. zunehmender Anzahl von Krankenhäusern, für alle drei Darstellungsvarianten abgenommen hat (Basisversion: 87,5% bei 5 KH, 78,4% bei 10 KH, 68,3% bei 15 KH;  $p < .001$ ). Zwischen den drei Darstellungsvarianten konnten hingegen keine signifikanten Unterschiede beobachtet werden. Anschließend wurde in einem multivariaten Regressionsmodell der Einfluss von insgesamt 11 unabhängigen Variablen auf die Verständlichkeit, d.h. die Wahl des besten Krankenhauses, untersucht. Wie in Tabelle 7 dargestellt, zeigte sich als Ergebnis, dass die Komplexität der Darstellung als einzige Variable einen signifikanten Einfluss auf die Verständlichkeit aufwies. Die unterschiedlichen Designvarianten zeigten hingegen keinen signifikanten Einfluss. Als Empfehlung für das Design von Report Cards schlussfolgerten die Autoren, dass insbesondere die Komplexität der Darstellung, hier gemessen an der Anzahl der dargestellten Krankenhäuser, einen entscheidenden Einfluss auf die korrekte Interpretation von Qualitätsinformationen hat. Daher soll insbesondere an weniger komplexen Darstellungsvarianten gearbeitet werden. Diese könnten neben einer geringeren Anzahl dargestellter Leistungserbringer beispielsweise auch durch sog.

Qualitätsfilter erreicht werden. In Abhängigkeit bestimmter vorab definierter Qualitätsziele werden nur Krankenhäuser dargestellt, welche diese Ziele auch erreichen.

Choice of the best performing hospital	
Variablen	Model 4 <sup>&amp;</sup> OR (95% CI)
Complexity of decision making	**
Low complexity decision making (N=451)	1.00 (ref)
Middle complexity decision making (N=407)	0.57 (0.37;0.87)*
High complexity decision making (N=412)	0.25 (0.16;0.39)**
HRC design types	
Report card design: basis version (N=635)	1.00 (ref)
Customized report card design I: composite measure (N=295)	1.11 (0.71;1.73)
Customized report card design II: number of hospitals and ordering (N=340)	0.87 (0.58;1.31)
Statistical model details	$\chi^2(22)=75.686,$ $R^2 = .123; p < .001$
<sup>&amp;</sup> Model 4: Adjusted for demographics, health related demographics, perceived differences in the quality of care and maximal travel distance	

*Tabelle 7: Logistic regression models predicting likelihood of selection of the best performing hospital based on the complexity of decision making and different design types (N=1270)*

**Emmert M, Schlesinger M (2017) Hospital Quality Reporting in the United States: Does Report Card Design and Incorporation of Patient Narrative Comments Affect Hospital Choice? [13]**

In der im Jahr 2015 in den USA durchgeführten Studie wurden im Rahmen einer Onlinebefragung 1.350 Personen zum Verständnis von Krankenhausqualitätsinformationen befragt. Hierbei wurde insbesondere der Einfluss des Designs (Darstellung von Qualitätsinformationen in Text, Sternenformat, numerisch) von Report Cards auf das Verständnis adressiert. Als zugrundeliegende Report Card wurde Hospital Compare herangezogen, die über verschiedene Dimensionen von Krankenhausinformationen berichtet (z.B. klinische Qualität, Struktur, Hygiene, Kosten; siehe Abbildung 9).

	HOSPITAL A	HOSPITAL B	HOSPITAL C	HOSPITAL D	HOSPITAL E	U.S. NATIONAL RATE
<b>General information</b>						
Hospital type	Acute Care Hospitals	Acute Care Hospitals	Critical Access Hospitals	Acute Care Hospitals	Acute Care Hospitals	
Provides emergency services	No	Yes	Yes	Yes	Yes	
<b>Survey of patients' experiences</b>						
Patients who gave their hospital a rating of 9 or 10 on a scale from 0 (lowest) to 10 (highest)	67%	67%	66%	78%	71%	71%
Patients who reported YES, they would definitely recommend the hospital	69%	74%	71%	80%	80%	71%
<b>Readmissions, complications, &amp; deaths</b>						
Rate of unplanned readmission after hip/knee surgery	5.2%	6.5%	5.4%	4.2%	Not Available <sup>1</sup>	5.4%
Rate of complications for hip/knee replacement patients	3.3%	5.6%	3.5%	2.1%	Not Available <sup>1</sup>	3.4%
<b>Costs of Care</b>						
Provider included in the plan network	Yes	Yes	Yes	Yes	Yes	
Out-of-pocket costs	No Different than U.S. National Rate					
<b>Healthcare-associated infections</b>						
Surgical site infections from hip replacement surgery (SSI: Hip)	0.867	Not Available <sup>1</sup>	Not Available <sup>1</sup>	0.442	Not Available <sup>1</sup>	
Clostridium difficile (C. diff.) Laboratory-identified Events (Intestinal infections)	0.931	1.782	1.035	0.309	Not Available <sup>1</sup>	

Abbildung 9: Screenshot of Hospital Report Card

Als Ergebnis zeigten sich unterschiedliche Ergebnisse für die Verständlichkeit in Abhängigkeit vom Design der Darstellung auf der Report Card. Wie in Abbildung 10 dargestellt, konnten insgesamt 51% der Befragten das jeweils beste Krankenhaus identifizieren, das beste Design wurde für die Darstellung in numerischer Form (Prozentangaben) ermittelt (56% vs. 48% vs. 47%;  $p < .001$ ). Die Ergebnisse hinsichtlich der Wahl des schlechtesten Krankenhauses unterschieden sich zwischen den drei Designs nicht signifikant ( $p > .05$ ).

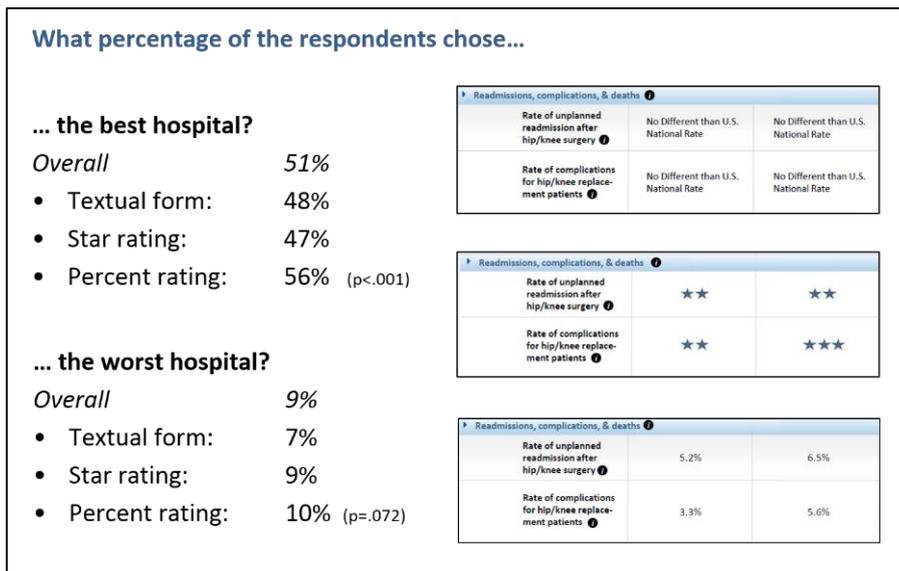


Abbildung 10: Selection of the Quantitatively Dominant Hospital and Quantitatively Dominated Hospital

Die durchgeführte Regressionsanalyse zeigte einen konstant signifikanten Einfluss des Designs der Darstellung auf die Verständlichkeit der Qualitätsinformationen (siehe Tabelle 8); so auch im umfassendsten Modell mit 12 unabhängigen Variablen. Hierbei zeigte sich, dass eine numerische Darstellung in Prozentangaben die Wahrscheinlichkeit der Wahl der richtigen Klinik um das 1,42fache erhöht (OR 1.42; 95% KI 1.17 - 1.73; p<.001) im Vergleich zu einer reinen textbasierten Darstellung. Zwischen der Darstellung in Textform oder Sternen wurden keine signifikanten Unterschiede ermittelt. Als Schlussfolgerung kann gezogen werden, dass das Design von Report Cards für die Verständlichkeit von Bedeutung ist. Es konnten hier signifikante Unterschiede zur Verständlichkeit zwischen verschiedenen Präsentationsdesigns nachgewiesen werden. Basierend auf den Ergebnissen sollten klinische Qualitätsergebnisse in der Form numerischer Darstellungen in Prozentangaben erfolgen. Die Darstellung in Textform oder Sternen führt zu ähnlichen Verständlichkeitswerten.

Choice of the quantitatively dominant hospital	
Variablen	Model 4& OR (95% CI)
<b>Design types</b>	
Textual information (N=902)	1.00 (ref)
Star ratings (N=867)	0.95 (0.78;1.16)
Numerical information (N=931)	1.42 (1.17;1.73)**
<b>Statistical model details</b>	
	$\chi^2(31)=282.621,$
	$R^2 = .135; p=.000$

\* Model 4: Adjusted for demographics, health related demographics, high vs. low maximizer and perceived differences in the quality of care

Tabelle 8: Logistic regression models predicting likelihood of selection of the quantitatively dominant hospital and quantitatively dominated hospital based on different design types (N= 2700)

**Geraedts et al. (2012) Communicating quality of care information to physicians: A study of eight presentation formats [32]**

Gegenstand dieser Untersuchung war herauszufinden, welche Präsentationsformate niedergelassene Ärzte für Krankenhausqualitätsdaten präferieren und wie sie diese verstehen. Dafür wurden 300 zufällig ausgewählte niedergelassene Ärzte in einem computerunterstützten Telefoninterview befragt. Den Ärzten wurden 8 unterschiedliche Präsentationsformate, die sich im Detaillierungsgrad sowie der Verwendung verschiedener Symbole unterschieden, in verschiedener Reihenfolge vorgelegt. Die Ärzte bewerteten daraufhin die Übersichtlichkeit, die Verständlichkeit, den Inhalt der Informationen, die Akzeptanz und ihre Präferenz hinsichtlich des Präsentationsformats (siehe Abbildung 11).



Abbildung 11: Format 1-8

Es zeigte sich, dass Präsentationen mit einfacher Sternebewertung als am übersichtlichsten (2,01) und verständlichsten (2,16) bewertet wurden, die Gesamtbewertung (4,21) und die Bewertung des Informationsinhalts (4,73) am schlechtesten ausfielen. Hinsichtlich der getesteten Verständlichkeit durch ein fiktives Auswahllexperiment erwies sich die einfache Sternebewertung als verständlichstes Präsentationsformat (94,3%; Tabelle 9). Die Tabellen mit Zahlenwerten, sortiert nach einem errechneten Gesamtergebnis und als Rangliste dargestellt, schnitt ebenfalls sehr gut ab (91,3%). Eine separate Darstellung der Qualitätsergebnisse in Balkenform und für jedes Krankenhaus individuell ergab die niedrigsten Verständnisswerte (55,3%). Betrachtet man allerdings ausschließlich die Ergebniswerte für Darstellungen mit einem Präsentationsformat (Nummer 4-7), so erscheint eine Darstellung mittels Ampelsymbolen verständlicher als mittels Textform, Sternenform oder Schulnotensystematik. Die in einem Gesamtscore dargestellte Variante mittels Sternenform (Nummer 8) erzielt die höchsten Verständlichkeitswerte, ist aber auch deutlich weniger komplex in der Interpretation, da hier nur ein Gesamtscore zu interpretieren ist. Die Autoren empfehlen auf der Basis ihrer Ergebnisse die Verwendung von detaillierten Indikatorenergebniswerten, die in Verbindung mit Ranglisten oder Ampelsymbolen dargestellt werden. Rein aggregierte Ergebnisse unter Verwendung von ausschließlich einfachen Symbolen wie Sternen oder Ampelfarben werden von den Ärzten zwar nicht gut akzeptiert; erzielen allerdings die höchsten Verständnisswerte.

	Ranked Numeric Table	Bar Charts	Numeric Table with Traffic Lights	School Grades Table	Table with Stars	Comparative Table	Traffic Lights Table	Simple Star Rating
Clarity <sup>a,c</sup>	2.32	3.68	2.58	3.60	3.90	4.42	2.88	2.01
Comprehensibility <sup>a,c</sup>	2.06	2.91	2.22	2.96	3.25	3.82	2.65	2.16
Information content <sup>a,c</sup>	1.84	2.54	1.86	3.14	3.40	3.98	3.19	4.73
Overall assessment <sup>a,c</sup>	2.14	3.31	2.29	3.44	3.77	4.27	3.16	4.21
Acceptance: yes (%) <sup>c</sup>	52.7	44.3	73.7	24.3	15.7	10.3	35.0	10.0
Acceptance: no/don't know (%)	47.3	55.7	26.3	75.7	84.3	89.6	65.0	90.0
Comprehension: yes <sup>b</sup> (%) <sup>c</sup>	91.3	55.3	76.7	72.7	70.7	64.0	80.7	94.3
Comprehension: no/don't know (%)	8.7	44.7	23.3	27.3	29.3	36.0	19.3	5.7

<sup>a</sup> Mean values of German school grades 1 to 6 (1 = very good; 6 = insufficient).

<sup>b</sup> Percentage of physicians answering both comprehension questions accurately.

<sup>c</sup>  $p < 0.001$  (Friedman test: differences between presentation formats).

<sup>d</sup>  $p < 0.001$  (Cochran's Q test: differences between presentation formats).

Tabelle 9: Physicians' assessment, acceptance and comprehension of eight presentation formats

### Greene J und Sacks RM (2018) Presenting Cost and Efficiency Measures That Support Consumers to Make High-Value Health Care Choices [33]

Die Studie untersuchte Ansätze zur Präsentation von Kosten- und Qualitätsdaten, die Konsumenten bei der Auswahl von qualitativ sehr guten Leistungserbringern bzw. Leistungserbringern mit dem besten Preis-Leistungs-Verhältnis unterstützen. Hierzu wurden 420 Mitarbeiter eines großen Halbleiterherstellers im Rahmen von zwei Experimenten befragt; bei beiden Experimenten gab es immer eine dominierende Auswahlalternative. (1) Die Probanden sahen (neben einigen Qualitätsinformationen) jeweils randomisiert eine von insgesamt fünf verschiedenen Darstellungen in Bezug auf die Kosten der Behandlung von Pneumonie für vier Krankenhäuser (siehe Abbildung 12). Für die Designs wurden Wort-Symbole, Zahlentabellen, Kostenwerte etc. dargestellt. Aus jeweils vier dargestellten Krankenhäuser sollten die Probanden das beste Krankenhaus auswählen. Gemessen wurde jeweils die Zeit, bis eine Auswahl getroffen wurde sowie der Anteil der Probanden, der das beste Krankenhaus auswählte.

	County Hospital	Evergreen Hospital	University Hospital	Woodland Hospital
<b>Effective Care for Pneumonia</b> Getting the right care at the right time				
Blood test is done before an antibiotic is given	Average	Better	Better	Better
Antibiotic is given within six hours of arrival to the hospital	Better	Better	Better	Better
Correct antibiotic drug is given	Better	Average	Better	Better
<b>Pneumonia Outcomes of Care</b>				
Rate of death within 30 days of leaving the hospital	Average	Better	Better	Better
<b>Cost of Pneumonia Care</b>				
Median cost of care	\$12,985	\$10,032	\$8,320	\$13,281

Note. The cost of pneumonia care presented in the display above (with actual dollar amounts), was one of five versions that were randomized to participants. The other four versions are presented below.

Alternate versions of cost					
v2	Spending per hospital patient (displayed in ratio)	1.32	1.02	0.85	1.35
v3	Median cost of care	High Cost	Average	Affordable	High Cost
v4	Median cost of care	High Cost	Average	Low Cost	High Cost
v5	Median cost of care	\$12,985	\$10,032	\$8,320	\$13,281
	High Quality, Affordable Hospital	Below	Average	Better	Average

Abbildung 12: An Example of the Display from the Cost Presentation Experiment, with the Alternate Cost Displays below

Als Ergebnis zeigte sich, dass sich bei der Präsentation der Kostendaten das visualisierte Wort-Symbol „Erschwinglich“ (93,8%) und der durchschnittliche in einer absoluten Zahl dargestellte Kostenwert in Dollar (93,1%) als am effektivsten in Bezug auf die Auswahl des besten Krankenhauses herausstellte (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**). Die kürzeste Zeit bis zur Wahlentscheidung brauchten die Probanden, wenn ein Indikator für das Krankenhaus mit dem besten Preis-Leistungsverhältnis abgebildet war. Die Wahlentscheidungen waren dann allerdings in Bezug auf die Auswahl des besten Leistungserbringers nicht so optimal wie im zuvor beschriebenen Fall (84,5%).

	Median Cost (\$) (n = 80)	CMS Spending Ratio (n = 75)	Word Icon: Affordable (n = 87)	Word Icon: Low Cost (n = 86)	Indicator for High-Quality, Affordable Hospital (n = 84)	p-Value
Choice of high-quality/low-cost hospital (University Hospital) (%; 95% CI)	93.8 (88.4–99.1)	77.3 (67.8–86.9)	93.1 (87.7–98.5)	84.9 (77.3–92.5)	84.5 (76.7–92.3)	.01
Time to make choice in seconds (mean, 95% CI)	30.5 (24.0–36.9)	35.8 (30.4–41.1)	30.8 (25.9–35.7)	30.1 (23.7–36.5)	22.1 (18.2–26.0)	.01

Note. p-Values are based upon a chi-square test for the choice of high-quality/low-cost hospital, and ANOVA for time to make choice.

Tabelle 10: Cost Presentation Experiment: Choice of Hospital by Cost Presentation Approach

Im Rahmen des zweiten Experiments wurde neben der leistungsbereichsbezogenen Gesamtpformance die Gesamt-Wiedereinweisungsrate für vier Krankenhäuser auf ebenfalls fünf unterschiedliche Art und Weisen dargestellt (siehe Abbildung 13).

	Lakeview Hospital	Parkdale Hospital	District Hospital	Regional Hospital
<b>Hospital Quality</b> Receiving the right care at the right time				
Overall Surgical Care	99%	99%	90%	99%
Overall Heart Failure	100%	100%	95%	100%
Overall Heart Attack	92%	95%	87%	93%
Overall Pneumonia	90%	85%	85%	89%
Overall Readmission Rates	35%	30%	25%	20%

Note. The readmission rates presented in the display above (with percentages), was one of five versions that were randomized to participants. The other four versions are presented below.

Alternate Versions of Readmissions					
v2	Overall Readmission Rates Lower score is better	35%	30%	25%	20%
v3	Rate of Patients <b>Not</b> Returning to the Hospital	65%	70%	75%	80%
v4	Overall Readmission Rates Discharged patients returning to the hospital	35%	30%	25%	20%
v5	Overall Readmission Rates Discharged patients returning to the hospital	Below	Average	Average	Better

Abbildung 13: An Example of the Display from the Hospital Readmission Presentation Experiment, with the Alternate Readmission Displays below

Auch bei der Darstellung von Wiedereinweisungsraten schnitten im Hinblick auf die Auswahl des besten Leistungserbringers die visualisierten Wort-Symbole („Better“ hinterlegt mit einem gelben Pfeil) sehr gut ab. Hier wählten insgesamt 91,2% der Probanden das beste Krankenhaus. Gleichzeitig war die verstrichene Zeit bis zur Auswahlentscheidung am kürzesten (26,2 Sekunden). Alle anderen Formate schnitten deutlich schlechter ab. Darüber hinaus gab es zwischen diesen Darstellungen mit z.B. Erklärungen der Wiedereinweisungsrate oder der Angabe einer Interpretationshilfe („Niedrigerer Wert ist besser“) keine nennenswerten Unterschiede im Verständnis gemessen am Wahlverhalten der Probanden (s. Tabelle 11).

	Readmission Rate (%) (n = 81)	Readmission Rate with "Lower Score Is Better" (n = 75)	Rate of Patients Not Returning to the Hospital (n = 87)	Readmission Rate with Explanation (n = 86)	Readmission Rate Using Word Icons (n = 91)	p-Value
Choice of highest quality hospital (regional) (%; 95% CI)	58.0 (47.2–68.9)	57.3 (46.0–68.6)	65.5 (55.4–75.6)	59.3 (48.8–69.8)	91.2 (85.3–97.1)	<.00
Time to make choice, in seconds (mean, 95% CI)	47.6 (41.1–54.2)	39.9 (35.2–44.6)	44.4 (38.9–49.9)	56.4 (48.2–64.6)	26.2 (22.2–30.2)	<.00

Note. p-Values are based upon a chi-square test for the choice of high-quality/low-cost hospital, and ANOVA for time to make choice.

Tabelle 11: Readmissions Presentation Experiment: Choice of Hospital by Readmission Presentation Approach

Als Gesamtergebnis der Studie empfehlen die Autoren daher sowohl für die Darstellung von Kostendaten als auch für die Darstellung von Wiedereinweisungsraten die Verwendung von visualisierten bzw. gelabelten Word-Symbolen bzw. Begriffen. Im Fall von Kostendaten zeigten die visualisierten Wort-Symbole (z.B. „Erschwinglich“) die beste Verständlichkeit, bei der Wiedereinweisungsrate die visualisierten Wort-Symbole „Besser“, „Durchschnitt“ und „Schlechter“.

### Hibbard et al. (2010) What Is Quality Anyway? Performance Reports That Clearly Communicate to Consumers the Meaning of Quality of Care [34]

In dieser Studie von Hibbard geht es weniger um das Verständnis der Befragten zu einzelnen Qualitätsindikatoren als vielmehr um die Frage, ob und wenn ja, welche Darstellung eines Frameworks zur Erläuterung von Qualitätsinformationen einen Einfluss auf die Verständlichkeit hat. Im Grunde wird untersucht, wie die „Spaltenüberschriften“ von Qualitätsinformationen gestaltet sein sollten, um das Verständnis der Informationen zu erhöhen; dies wird in der Studie als Qualitätsframework bezeichnet. Dafür wurden nach vorbereitenden Fokusgruppen und Interviews insgesamt 439 Personen im Alter zwischen 18 und 64 Jahren befragt. Jede Person bekam randomisiert eine von drei verschiedenen Darstellungen zur Arzt- und Krankenhausqualität vorgelegt. Die Designs unterschieden sich hinsichtlich der verwendeten Sprache (einfache vs. technische Sprache) und ob die Informationen entsprechend des Qualitätskonzeptes strukturiert in Kategorien („Wirksamkeit der Behandlung“, „Patientensicherheit“, „Patientenzentrierte Behandlung“) unterteilt dargestellt wurden oder nicht (siehe Abbildung 14: The header from each version of the quality doctor report). Die Befragten mussten anschließend einige Fragen dazu beantworten; z.B. in welchem Krankenhaus es am wahrscheinlichsten ist, eine wirksame Behandlung zu erhalten.

**A**

Doctor	<u>Effective</u> Uses treatments proven to get results		<u>Safe</u> Uses safeguards to protect patients from medical errors		<u>Patient Focused</u> Responsive to patients' needs and preferences	
	Uses best practices for treating chronic illness	Uses best practices for screening for diseases	Has procedures to prevent medication errors	Has a system to prevent lab results from getting lost	Communicates well with patients	Spends enough time with patients
Dr. Allard	average	average	<b>better</b>	<b>better</b>	average	average
Dr. French	<b>below</b>	average	<b>better</b>	average	<b>below</b>	<b>below</b>
Dr. Cain	average	average	<b>below</b>	<b>below</b>	<b>better</b>	average
Dr. Emory	<b>better</b>	<b>better</b>	average	average	average	average
Dr. Castle	average	average	average	average	<b>better</b>	<b>better</b>

**B**

Doctors	Uses best practices for treating chronic illness	Uses best practices for screening for diseases	Has procedures to prevent medication errors	Has a system to prevent lab results from getting lost	Communicates well with patients	Spends enough time with patients

**C**

Doctors	Percent of patients with diabetes who had HbA1c test	Percent of women receiving breast cancer screening	Provider uses electronic prescribing to prevent medication errors	Provider uses a computer system to prevent lab results from getting lost	Patients report that provider explains information, listens and shows respect to patients	Patients report that provider spends enough time with patient

Abbildung 14: The header from each version of the quality doctor report

Das beste Konzeptverständnis für Arzt- und Krankenhausqualität konnte erzielt werden, wenn die Ergebnisse in kategorisierter Form entsprechend dem verwendeten Qualitätskonzept verwendet wurden (d.h. unter Verwendung der drei Hauptkategorien; siehe A in Abbildung 14) sowie eine einfache Sprache verwendet wurde (86,7% bzw. 72,9%) (s. Tabelle 12). Bezüglich der Einschätzung der Probanden, wie hilfreich sie die Darstellung fanden, wurden beide Darstellungen mit 3,8 gleich gut bewertet. Die Bewertung für die Darstellung mit technischer Sprache wurde hingegen mit 3,3 hingegen schlechter bewertet.

Indices	Experimental Groups		
	Condition 1: Plain Language With Framework (n = 147)	Condition 2: Plain Language Without Framework (n = 160)	Condition 3: Technical Language Without Framework (n = 132)
Comprehension of quality report (% correct)			
Doctor quality concepts	86.7	76.0	67.7 <sup>a,b,c</sup>
Hospital quality concepts	72.9	59.3	39.3 <sup>a,b,c</sup>
Understanding definitions of quality concepts (% correct)	74.2	69.0	62.4 <sup>b</sup>
Usefulness of quality report (mean) <sup>d</sup>			
Doctor chart	3.8	3.9	3.5 <sup>b,c</sup>
Hospital chart	3.8	3.8	3.3 <sup>b,c</sup>

a.  $p < .05$  between plain language with framework versus plain language without framework.

b.  $p < .05$  between plain language with framework versus technical language without framework.

c.  $p < .05$  between plain language without framework versus technical language without framework.

d. Scale: 1 = not at all useful, 5 = very useful.

Tabelle 12: Comprehension Indices and Usefulness by Experimental Group

Als Gesamtfazit empfehlen die Autoren deshalb die Verwendung von einfacher Sprache sowie die strukturierte Darstellung von Qualitätsinformationen in Form eines vorher definierten Qualitätskonzepts. Dadurch wird die Verständlichkeit der Informationen erhöht und diese zudem als hilfreicher eingeschätzt. Dieser Befund ist konstant über die verschiedenen Altersgruppen und Geschlechter hinweg.

**Mazor M et al. (2009) Communicating Hospital Infection Data to the Public: A Study of Consumer Responses and Preferences [35]**

In der US-Studie von Mazor und Kollegen wurde die Verständlichkeit von Qualitätsinformationen zu "Healthcare-associated Infections" (HAI) untersucht. Hierbei wurden Infektionsraten auf Ebene von Krankenhäusern von acht unterschiedlichen Report Cards präsentiert, welche sich im Design und der Gestaltung unterschieden. Bezüglich des Designs wurde eine textbasierte Darstellung („29 out of 1000 people got an infection“; siehe Abbildung 15) sowie eine Darstellungsform mittels Balken integriert; neben der zusätzlichen Darstellung von Konfidenzintervallen in einigen Studienarmen. Befragt wurden insgesamt 201 zufällig ausgewählte Bürger aus Worcester (MA) postalisch auf dem Briefweg.

<b>Safe Practice Score</b>	
How well did a hospital <b>follow procedures that can reduce infections?</b>	
The higher the score, the more closely the hospital is following the recommended practices.	
The best possible score is 100.	
<b>Best</b>	
Green Valley Hospital	had a safe practice score of 93
Clear Falls Hospital	had a safe practice score of 88
Orchard Park Hospital	had a safe practice score of 87
<b>Massachusetts Average</b>	<b>had a safe practice score of 87</b>
Coral City Hospital	had a safe practice score of 79
<b>Worst</b>	

<b>Infection Rate*</b>	
How many people <b>got an infection</b> while in the hospital?	
<b>Best</b>	
Orchard Park Hospital	29 out of 1,000 patients got an infection
Green Valley Hospital	36 out of 1,000 patients got an infection
<b>Massachusetts Average</b>	<b>41 out of 1,000 patients got an infection</b>
Coral City Hospital	42 out of 1,000 patients got an infection
Clear Falls Hospital	51 out of 1,000 patients got an infection
<b>Worst</b>	

Abbildung 15: First Page from a Word report with inconsistent indicators

Als Ergebnis zeigten sich (entgegen der Erwartungen der Autoren der Studie) im Rahmen dieser Untersuchung keine Unterschiede bezüglich der Interpretationsqualität zwischen Text- und Balkenform sowie zwischen der Darstellung mit und ohne Konfidenzintervall.

### **Sander U et al (2016) Verständlichkeit der Texte von Qualitätsvergleichen zu Krankenhausleistungen [36]**

Es wurde untersucht, inwieweit die Verständlichkeit der Texte von Qualitätsvergleichen und die Anzahl der verwendeten Fachbegriffe auf Krankenhausvergleichsportalen mit dem Verständnis der Informationen und den Wahlentscheidungen der Nutzer zusammenhängen. Für die Studie wurden zunächst die 10 wichtigsten deutschen Krankenhausvergleichsportale ermittelt und anschließend die Darstellungen zur risikoadjustierten Mortalitätsrate bei Herzkatheterbehandlungen (PCI) untersucht. Gemessen wurde zunächst die Anzahl der Worte und Fachbegriffe. Anschließend wurde mit zwei computergestützten Verfahren (Flesch und Wiener Sachtextformel) und einem Experten basierten heuristischen Verfahren (Hamburger Verständlichkeitsmodell) die Lesbarkeit und die Verständlichkeit analysiert. In einer Online-Befragung sahen insgesamt 2.027 Probanden randomisiert jeweils 3 der 10 Präsentationen der risikoadjustierten Mortalitätsrate zu fünf fiktiven Krankenhäusern. Die Aufgabe bestand in der Auswahl des jeweils besten Krankenhauses. Darüber hinaus sollten die Probanden in einer Selbsteinschätzung angeben, wie verständlich sie verschiedene Fachbegriffe fanden (z.B. Strukturierter Qualitätsbericht, Referenzbereich) und inwieweit die Fachbegriffe das Verständnis der Darstellung verschlechtern haben. Als Ergebnis zeigte sich, dass ein geringerer Fachwortanteil und bessere Werte für die Dimension Einfachheit signifikant besser verständlich aufgefasst wurden ( $-0,399$ ;  $P < 0,001$  bzw.  $0,161$ ;  $P < 0,001$ ) sowie auch zu einer häufigeren Auswahl der Klinik mit der niedrigsten Todesrate führten ( $-0,147$ ;  $P < 0,001$  bzw.  $0,036$ ;  $P = 0,002$ ; siehe Tabelle 13).

Textanalyse: Anzahl Worte und Fachbegriffe				Textanalyse: Messung von Lesbarkeit und Verständlichkeit				Online-Befragung: Unklare Fachbegriffe haben Verständnis der Informationen verschlechtert				Online Befragung	
Portal	Worte	Anzahl-Fachbegriffe	Anteil-Fachbegriffe	Flesch (d)	WSTF1	Einfachheit	Kürze/Prägnanz	Stimme vollkommen zu	Stimme eher zu	Stimme eher nicht zu	Stimme nicht zu	Auswahl KRH niedrige Mortalität	Einschätzung Verständlichkeit
Berlin	206	10	0,5%	34	12	0,3	0,0	15,6%	35,2%	34,2%	15,1%	77,2%	4,17
BKK	199	18	9,0%	36	12	-0,7	0,3	29,1%	38,7%	23,6%	8,6%	31,4%	3,22
DKV	542	46	8,5%	35	13	-0,7	0,3	44,4%	34,9%	14,4%	6,3%	27,5%	2,38
Hamburg	327	10	3,1%	31	12	0,3	1,0	9,7%	30,8%	37,0%	22,5%	76,9%	4,68
IQM	127	22	17,3%	32	12	-1,7	0,0	22,6%	41,1%	26,7%	9,6%	65,1%	3,76
KRHde	497	39	7,8%	27	13	-1,0	-1,0	31,5%	39,5%	20,9%	8,1%	73,0%	3,70
QK	79	9	11,4%	22	14	-1,7	0,7	34,6%	39,0%	18,5%	7,9%	64,5%	2,99
Rheinland	365	4	1,1%	20	14	0,7	0,7	20,2%	42,0%	28,0%	9,9%	67,9%	3,92
TK	229	11	4,8%	36	11	1,0	1,0	22,2%	37,1%	27,0%	13,7%	58,2%	3,53
WL	209	21	10,0%	17	14	0,5	1,0	16,6%	39,4%	27,5%	16,5%	66,1%	3,53
Durchschnitt	278	19	7,4%	29	13	-0,3	0,4	24,6%	37,8%	25,8%	11,8%	60,8%	3,59

Flesch (d): 100: sehr leicht verständlich, 0: völlig unverständlich; WSTF1: 4 = leichteste Stufe, 15 = schwerste Stufe; Einfachheit: -2 = kompliziert, +2 = einfach; Kürze/Prägnanz: -2 = weitschweifig, +2 = kurz und prägnant; Auswahl KRH niedrige Mortalität: Prozentsatz der Probanden, welche das Krankenhaus mit der niedrigsten Mortalitätsrate auswählten. Einschätzung Verständlichkeit: Einschätzung der Probanden der Verständlichkeit des präsentierten Qualitätsindikators, 1 = sehr verständlich, 7 = sehr unverständlich

**Tabelle 13: Ergebnisse zur Verständlichkeit der Texte von Qualitätsvergleichen**

Die Autoren schlussfolgern aus den Ergebnissen, dass eine bessere Verständlichkeit der Texte durch die Vermeidung von Fachbegriffen bzw. verständlichere Darstellung von Fachbegriffen häufiger zu einer richtigen Interpretation der Informationen auf Seiten der Nutzer führt.

### Sander U et al (2017) Verstehen Laien Informationen über die Krankenhausqualität? Eine empirische Überprüfung am Beispiel der risikoadjustierten Mortalität [37]

Der Fokus dieser Studie bestand in der Untersuchung, ob Laien die Konzepte hinter Qualitätsvergleichen verstehen und ob ein Zusammenhang des Verständnisses mit der Krankenhauswahl besteht. Für die Studie wurden zunächst Fachbegriffe identifiziert, die zur Erklärung der risikoadjustierten Mortalitätsrate bei Herzkatheteruntersuchungen (PCI) auf Report Cards verwendet werden. In einer Online-Befragung wurden anschließend 535 Probanden befragt. Jedem Probanden wurden randomisiert drei von insgesamt sechs fiktiven Report Cards gezeigt. Aus den Darstellungen sollte das beste aus fünf Krankenhäusern (Krankenhaus mit der niedrigsten Risikoadjustierten Mortalitätsrate) ausgewählt werden (siehe Abbildung 16).

#### Variante 1

##### Todesfälle bei Ballonkatheter-Behandlungen

Ein gutes Qualitätsergebnis eines Krankenhauses zeichnet sich durch eine möglichst niedrige risikoadjustierte Todesfallrate aus

Das Schaubild zeigt, wie häufig Patienten im Zusammenhang mit einer Ballonkatheter-Behandlungen im Krankenhaus sterben.

Ausführlichere Informationen sind mit der Maus über Grafiken und Texten abrufbar.



Abbildung 16: Online-Darstellung der RAM für fünf Krankenhäuser (Variante 1)

Für alle fünf untersuchten Fachbegriffe (Tatsächliche Todesrate, Erwartete Todesrate, Risikoadjustierte Todesrate, Referenzbereich, Bundesdurchschnitt) hing ein besseres Verständnis der Fachbegriffe signifikant mit der Auswahl des Krankenhauses mit der niedrigsten risikoadjustierten Todesrate zusammen (s. Tabelle 14). So wählten beispielsweise 86,7% der Probanden, die Item 5 korrekt verstanden haben, auch gleichzeitig das Krankenhaus mit der niedrigsten risikoadjustierten Mortalitätsrate aus ( $p=0,001$ ). Zwischen den verschiedenen Portaldesigns variierte dabei der Anteil der Probanden, der das Krankenhaus mit der niedrigsten Risikoadjustierten Mortalitätsrate wählte, von 38,4% bis 81,0%. Die Anteile der korrekten Antworten zu den Fachbegriffen lagen dabei zwischen 60,2% und 75,0%, wobei die Anteile der korrekten Antworten zu einzelnen Aussagen zu den Konzepten stark schwankten (14,2% bis 89,5%). Am schwierigsten zu verstehen waren die Begriffe Referenzbereich und Risikoadjustierte Todesrate (s. Tabelle 14).

Fachbegriff	Items zur Prüfung des Konzeptverstehens	Antwort	Probanden (N=353)	KH mit niedrigster RAM gewählt (N=285)	Nicht KH mit niedrigster RAM gewählt (N=68)	Phi-Koeffizient (Zusammenhangsstärke)	p	
a) Die Risiko-adjustierte Todesrate	1) wird auf der Grundlage der unterschiedlichen Risiken von Patienten wie Alter, Geschlecht, Vor- und Begleiterkrankungen, etc. berechnet. (r)	korrekt	193 (54,7%)	165 (57,9%)	28 (41,2%)	0,096	0,121	
		inkorrekt	66 (18,7%)	51 (17,9%)	15 (22,1%)			
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
	2) ist bei Krankenhäusern mit vielen Risiko-Patienten in der Regel genauso hoch wie bei Häusern mit wenigen Risiko-Patienten. (r)	korrekt	50 (14,2%)	42 (14,7%)	8 (11,8%)	0,008	0,899	
		inkorrekt	209 (59,2%)	174 (61,1%)	35 (51,5%)			
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
	3) ist für alle Krankenhäuser in Deutschland gleich hoch. (f)	korrekt	235 (66,6%)	196 (68,8%)	39 (57,4%)	-0,001	0,993	
		inkorrekt	24 (6,8%)	20 (7,0%)	4 (5,9%)			
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
	b) Die Tatsächliche Todesrate	4) stellt die Gesamtzahl der im Krankenhaus verstorbener Patienten dar. (f)	korrekt	216 (61,2%)	185 (64,9%)	31 (45,6%)	-0,136*	0,029*
inkorrekt			43 (12,2%)	31 (10,9%)	12 (17,6%)			
keine Antwort			94 (26,6%)	69 (24,2%)	25 (36,8%)			
5) zeigt wie viele der behandelten Patienten verstorben sind. (r)		korrekt	694 (67,0%)	588 (68,1%)	106 (61,6%)	0,158	0,168	
		inkorrekt	342 (33,0%)	276 (31,9%)	66 (38,4%)			
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
c) Die Erwartete Todesrate		6) sollte nicht höher als die erwartete sein, um eine gute Behandlungsqualität darzustellen. (r)	korrekt	288 (81,6%)	247 (86,7%)	41 (60,3%)	0,183*	0,001*
			inkorrekt	47 (13,3%)	31 (10,9%)	16 (23,5%)		
			keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)		
		7) beinhaltet die erwartete Sterblichkeit der Patienten aufgrund ihrer Alters- und Geschlechtsverteilung. (f)	korrekt	124 (35,1%)	105 (36,8%)	19 (27,9%)	0,035	0,527
	inkorrekt		211 (59,8%)	173 (60,7%)	38 (55,9%)			
	keine Antwort		94 (26,6%)	69 (24,2%)	25 (36,8%)			
	8) gibt den Wert für gute Qualität an. (f)	korrekt	303 (85,8%)	255 (89,5%)	48 (70,6%)	-0,096	0,079	
		inkorrekt	32 (9,1%)	23 (8,1%)	9 (13,2%)			
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
	d) Der Referenzbereich	9) zeigt an, wieviel Sterbefälle bei den behandelten Patienten einer Klinik zu erwarten sind. (r)	korrekt	1005 (75,0%)	846 (76,1%)	159 (69,7%)	0,164	0,060
inkorrekt			335 (25,0%)	266 (23,9%)	69 (30,3%)			
keine Antwort			94 (26,6%)	69 (24,2%)	25 (36,8%)			
10) hängt u. a. vom Alter, Geschlecht sowie Vor- und Begleiterkrankungen der Patienten des jeweiligen Krankenhauses ab. (r)		korrekt	249 (70,5%)	220 (77,2%)	29 (42,6%)	0,247*	0,000*	
		inkorrekt	78 (22,1%)	52 (18,2%)	26 (38,2%)			
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
11) zeigt an wieviel Patienten verstorben sind. (f)		korrekt	190 (53,8%)	157 (55,1%)	33 (48,5%)	0,017	0,755	
		inkorrekt	272 (77,1%)	233 (81,8%)	39 (57,4%)	-0,148*	0,008*	
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
e) Die Erwartete Todesrate		12) ist für alle KH in Deutschland gleich hoch. (f)	korrekt	55 (15,6%)	39 (13,7%)	16 (23,5%)	-0,100	0,070
	inkorrekt		304 (86,1%)	256 (89,8%)	48 (70,6%)			
	keine Antwort		94 (26,6%)	69 (24,2%)	25 (36,8%)			
	13) gibt den Bereich eines Qualitätsindikators an, in dem sich der gemessene Wert befinden sollte um erreichbar gute Qualität darzustellen. (r)	korrekt	26 (7,4%)	13 (4,6%)	7 (10,3%)	-0,220*	0,000*	
		inkorrekt	962 (73,5%)	824 (75,7%)	138 (62,7%)	0,242*	0,001*	
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
	f) Die Tatsächliche Todesrate	14) ist für einen Qualitätsindikator für alle deutschen Krankenhäuser gleich. (r)	korrekt	161 (45,6%)	142 (49,8%)	19 (27,9%)	0,182*	0,003*
			inkorrekt	108 (30,6%)	80 (28,1%)	28 (41,2%)		
			keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)		
		15) zeigt an, wieviel der behandelten Patienten verstorben sind. (f)	korrekt	79 (22,4%)	64 (22,5%)	15 (22,1%)	-0,026	0,673
inkorrekt			190 (53,8%)	158 (55,4%)	32 (47,1%)			
keine Antwort			94 (26,6%)	69 (24,2%)	25 (36,8%)			
16) ist für jedes Krankenhaus unterschiedlich. (f)		korrekt	236 (66,9%)	200 (70,2%)	36 (52,9%)	-0,156*	0,010*	
		inkorrekt	33 (9,3%)	22 (7,7%)	11 (16,2%)			
		keine Antwort	94 (26,6%)	69 (24,2%)	25 (36,8%)			
Summe		korrekt	84 (23,8%)	63 (22,1%)	21 (30,9%)	-0,081	0,127	
	inkorrekt	648 (60,2%)	555 (62,5%)	93 (49,5%)	0,233*	0,006*		

Tabelle 14: Zusammenhang zwischen den Ergebnissen des Instruments zur Bestimmung des Konzeptverstehens und der Auswahl eines Krankenhauses mit der niedrigsten RAM

Auf der Basis dieser Ergebnisse geben die Autoren zwei wesentliche Empfehlungen. Zum einen sollten Entwickler von Report Cards darauf achten, bei der Darstellung von Qualitätsdaten und erklärenden Texten zu einzelnen Indikatoren möglichst wenig Fachbegriffe zu verwenden. Zum anderen sollten auf oberster Ebene eher einfache oder selbsterklärende Begriffe wie Erwartete Mortalitätsrate oder Tatsächliche Mortalitätsrate zur Darstellung der Ergebnisse verwendet werden als ein schwerer zu verstehendes Konzept wie die Risikoadjustierte Mortalitätsrate. Die für die fünf Fachbegriffe im Rahmen der Untersuchung verwendeten erklärenden Aussagen können ebenfalls als Basis für die Entwicklung neuer Darstellungen auf Krankenhausvergleichsportalen verwendet werden.

### Zwijnenberg NC et al. (2012) Understanding and using comparative healthcare information; the effect of the amount of information and consumer characteristics and skills [38]

In der niederländischen Studie von Zwijnenberg und Kollegen wurde unter anderem untersucht, welchen Einfluss die Menge an Qualitätsinformationen auf das Verständnis bei der Krankenhauswahl für eine Hüft- bzw. Knie-TEP Operation haben. Hierzu haben die Autoren insgesamt 414 Personen befragt und in einem randomisierten Design vier unterschiedliche Gruppen gebildet, denen 3, 7, 11 oder 15 Qualitätsinformationen zu drei Krankenhäusern gezeigt wurden (siehe die Subgruppendarstellung in Abbildung 17). Die Verständlichkeit wurde anhand von zwei Kennzahlen untersucht: (1) die Wahl der Klinik (beste Klinik vs. nicht beste Klinik) sowie (2) die Anzahl richtiger Antworten auf 5 Verständnisfragen bezüglich der Qualitätsinformationen der gezeigten drei Krankenhäuser.

	Hospital A	Hospital B	Hospital C	Subgroup
Conduct of doctors	★★★	★★☆	★★☆	1, 2, 3, 4
Conduct of nurses	★★☆	★☆☆	★★☆	2, 3, 4
Pain control	★★☆	★★☆	★★☆	3, 4
Information about new medication	★★★	★★☆	★★☆	4
Information provision before surgery	Yes	No	Yes	2, 3, 4
Procedures to prevent adverse effects of thrombosis	Yes	Yes	No	1, 2, 3, 4
Registration of complications related to THA/TKA	Yes	No	Yes	3, 4
National registration of orthopaedic implants	Yes	Yes	Yes	3, 4
Transfusion of homologous blood	Yes	Yes	No	4
Specialist areas of orthopaedist	No	No	No	2, 3, 4
Number of performed total knee- or hip replacements among adults in a year	314	98	244	1, 2, 3, 4
Number of performed total knee- or hip replacements among children in a year	0	1	0	4
Number of orthopaedists in the hospital	8	2	4	2, 3, 4
Information provision approach	Written information material and briefings	Written information material	Written information material	3, 4
Group-hospital admission	Yes	No	No	4

★★★ better than average.

★★☆ average.

★☆☆ worse than average.

The information provided was derived from patient experience indicators measured by the Consumer Quality Index (CQI: the Dutch standard for measuring patient experiences in healthcare), and indicators about hospital services and clinical performance indicators were derived from hospital registrations.

Abbildung 17: Aspects of comparative healthcare information shown to each subgroup

Als Ergebnis für die Kennzahl (1) zeigte sich, dass zwischen 97% (Gruppe 1 bis 3) und 99% (Gruppe 4) der Teilnehmer das jeweils beste Krankenhaus ausgewählt haben. Signifikante Unterschiede konnten hierbei nicht festgestellt werden, ein angedachtes Regressionsmodell konnte aufgrund der Verteilung der Antworten der Probanden nicht durchgeführt werden (keine ausreichende Anzahl an „falschen“ Antworten bezüglich der Klinikwahl; siehe Tabelle 15). Bezüglich der Kennzahl (2) zeigten sich bei der deskriptiven Auswertung ebenfalls sehr positive Ergebnisse; die Verständnismerte, gemessen am Mittelwert der richtig beantworteten Fragen, rangierten zwischen 3,8 (Gruppe 2) und 4,1 (Gruppe 4). Auch hier konnten keine signifikanten Unterschiede festgestellt werden. In einem anschließend durchgeführten Regressionsmodell wurde gezeigt, dass die Menge an gezeigten Informationen einen geringen, aber signifikanten Einfluss auf das Verständnis hat ( $\beta = -0.12$ ) (Tabelle 15). Probanden, denen sieben Qualitätsinformationen gezeigt wurden, hatten mehr Schwierigkeiten beim Verständnis gegenüber Probanden, denen drei Qualitätsinformationen gezeigt wurden.<sup>2</sup>

<sup>2</sup> Ein weiteres Ergebnis zeigte allerdings auch, dass der Nutzen der Informationen für die Klinikwahl bei der Bereitstellung von 7 Qualitätsinformationen höher als der von 3 Qualitätsinformationen eingestuft wurde.

	Comprehension of information			Ease of hospital choice			Usefulness of information		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
<b>Predictors</b>									
Male (female = reference)	-.01	.04	.03	-.02	.02	.03	.01	.04	.05
Age	-.32**	-.28**	-.29**	-.16*	-.15*	-.15*	-.11	-.13*	-.12*
Education level	.29**	.18**	.18**	.16*	.09	.09	.10*	.06	.06
Literacy		.22**	.22**		.15*	.15*		.07	.08
Numeracy		.13*	.14*		.07	.08		.09	.09
Active choice behaviour		.04	.05		.13	.13		.24**	.23**
<b>Amount of information</b>									
3 aspects (reference)			-			-			-
7 aspects			-.12*			.11			.16*
11 aspects			-.08			-.05			.10
15 aspects			-.04			.01			.04
R <sup>2</sup>	0.19**	0.26**	0.27**	0.05**	0.09**	0.11**	0.02	0.09**	0.11**
Change in R <sup>2</sup>		0.07**	0.01		0.04*	0.02*		0.07**	0.02

\* p < 0.05 \*\* p < 0.001.  
For the outcome measure 'correct hospital choice' no regression models were estimated, due to the skew distribution of this outcome measure.

Tabelle 15: Hierarchical regression models with regression coefficients (Beta) predicting the outcome measures [N = 349]

**Peters E et al. (2007) Less Is More in Presenting Quality Information to Consumers [39]**

In der Studie von Peters und Kollegen wurden drei einzelne Untersuchungen durchgeführt, um insbesondere die Bedeutung der Darstellung von Qualitätsinformationen über Krankenhäuser für die Verständlichkeit zu untersuchen. Befragt wurden insgesamt 303 Personen im erwerbsfähigen Alter. In der ersten Studie wurde der Einfluss der Anordnung von Krankenhäusern sowie der Menge der dargestellten Informationen auf die Verständlichkeit untersucht; bezüglich der Informationen wurde zudem unterschieden in „Quality and Nonquality Hospital Information“ (siehe beispielsweise Abbildung 18).

Indicators	Hospital X	Hospital Y	Hospital Z
Your out-of-pocket costs	\$	\$\$\$	\$\$
No. of general care beds	550	231	180
Rated quality of hospital food (higher is better)	4.1	1.1	2.0
% of time guidelines for heart attack care are followed	82%	92%	87%
% of time guidelines for pneumonia care are followed	60%	89%	78%
No. of visiting hours per day	11	6	8
No. of registered nurses per 100 patients	18	38	29
Patient references available	Limited	Limited	Limited
Has computer system to prevent medication errors	No	Yes	Limited

Abbildung 18: Studie 1: Unordered—Both Quality and Nonquality Hospital Information Is Presented Unordered

Als Ergebnis zeigten sich die höchsten Verständniswerte, wenn einerseits eine Darstellung nur ausgewählter Informationen (d.h. Kosten und Quality Hospital Information) erfolgt und diese zudem entsprechend der Reihenfolge Kosten – Quality Information angeordnet waren (57% bzw. 67%) (siehe Abbildung 19). Eine entsprechend der Reihenfolge Kosten – Quality Information – Nonquality Hospital Information gruppierte, angeordnete Darstellung erzielte niedrigere Verständniswerte (53% bzw. 43%), die aber höher waren als die ungruppierte Darstellung aller Informationen der drei Kategorien (44% bzw. 38%).

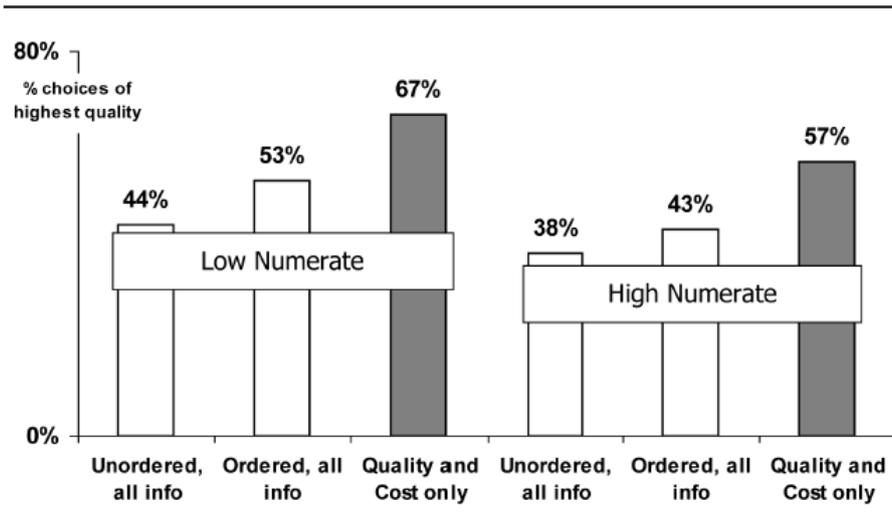


Abbildung 19: Study 1. Choice of the Highest-Quality Hospital by Presentation Format and Numeracy

In der zweiten Teilstudie wurde untersucht, ob die Darstellung von Symbolen (schwarz weiße Kreise bzw. in Ampelfarben gefüllte Kreisdarstellung, siehe Abbildung 20) zur leichteren Interpretierbarkeit von Zahlen dazu führen, dass Nutzer diese stärker gewichten im Vergleich zu weniger bedeutsamen Informationen. Ebenso wurde untersucht, ob die grafisch unterstützten Darstellungen von nur indikationsspezifischen Kennzahlen im Gegensatz zur grafisch unterstützten Darstellung aller Kennzahlen zu mehr Bedeutung der jeweils wichtigeren Kennzahlen bzw. bessern Verständniswerten (Wahl des Krankenhauses mit der geringsten Sterberate) führen. Hierzu wurden fünf Designs entwickelt und evaluiert, jeweils 15 Krankenhäuser wurden präsentiert.

	Cost	Overall Patient Satisfaction	Death Rate for Heart-Failure Patients
Hospital E	\$\$\$	68	● 1.4%
Hospital H	\$\$\$	76	◐ 4.1%
Hospital B	\$\$\$	81	◑ 4.8%
Hospital J	\$\$\$	87	◒ 5.3%
Hospital K	\$\$\$	90	○ 6.1%
Hospital D	\$\$	71	● 3.9%
Hospital F	\$\$	78	◐ 4.4%
Hospital I	\$\$	85	◑ 4.9%
Hospital N	\$\$	90	○ 6.1%
Hospital G	\$\$	94	○ 6.6%
Hospital M	\$	88	◑ 5.2%
Hospital A	\$	91	○ 6.4%
Hospital C	\$	95	○ 7.3%
Hospital O	\$	96	○ 8.9%
Hospital L	\$	98	○ 9.6%

Note: A key indicated that ● = better than average, ◐ = average, and ○ = worse than average. For the conditions using traffic-light symbols, the three symbols above were systematically replaced with circles filled in with green, yellow, and red, respectively. The corresponding key looked like a traffic light.

Abbildung 20: Study 2: Death Rate Only Is Made Easier to Evaluate with Black and White Symbols

Die höchsten Verständniswerte zeigten sich, wenn ausschließlich die Mortalitätsrate grafisch unterstützt dargestellt wurde (74%) im Gegensatz zur vollständig nicht grafisch-unterstützten Darstellung (60%) sowie wenn sowohl Patientenzufriedenheit als auch die Mortalitätsrate grafisch unterstützt wurden (60%;  $p < .05$ ) (siehe Abbildung 21). Die Verwendung schwarz weißer Kreise bzw. in Ampelfarben gefüllter Kreise hatte keinen signifikanten Einfluss auf die Verständlichkeit.

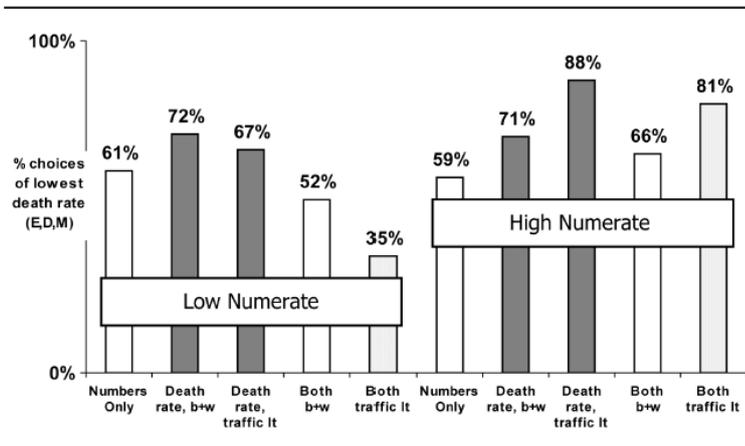


Abbildung 21: Study 2. Choice of the Lowest-Death-Rate Hospitals by Presentation Format and Numeracy

Im Rahmen der dritten Teilstudie wurde die Richtungsdarstellung numerischer Informationen (höhere Zahlen stehen für bessere Ergebnisse) sowie die grafisch-unterstützte Darstellung numerischer Ergebnisse zu wiederum 15 Krankenhäusern untersucht. Vier Studienarme waren enthalten (1: higher is better, no symbols; 2: higher is better, symbols – siehe Abbildung 22; 3: lower is better, no symbols; 4: lower is better, symbols).

	No. of Registered Nurses per 100 Patients	Your Out-of-Pocket Cost
Hospital A	38 ⊕	\$\$\$
Hospital B	36 ⊕	\$\$\$\$
Hospital C	32 ⊕	\$\$\$\$
Hospital D	38 ⊕	\$\$\$\$
Hospital E	24 ○	\$
Hospital F	37 ⊕	\$\$\$\$
Hospital G	19 ⊖	\$\$\$\$
Hospital H	20 ⊖	\$
Hospital I	18 ⊖	\$
Hospital J	21 ○	\$
Hospital K	24 ○	\$
Hospital L	23 ○	\$\$\$\$
Hospital M	26 ○	\$
Hospital N	24 ○	\$\$\$
Hospital O	17 ⊖	\$

Note: ⊕ = more registered nurses per 100 patients; ○ = average number of registered nurses per 100 patients; ⊖ = fewer registered nurses per 100 patients.

Abbildung 22: Study 3: Patient-to-Nurse Ratio Is Presented as “Higher Is Better” with Easier-to-Evaluate Symbols

Als Ergebnis (siehe Abbildung 23) zeigte sich einerseits, dass die Verwendung höherer Zahlen für positive Ergebnisse zu höheren Verständniswerten führt im Vergleich zur Verwendung niedrigerer Zahlen für positive Ergebnisse (73% vs. 60%,  $p < .001$ ). Die zusätzliche Verwendung von Symbolen hingegen hatte keinen Einfluss auf die Verständlichkeit.

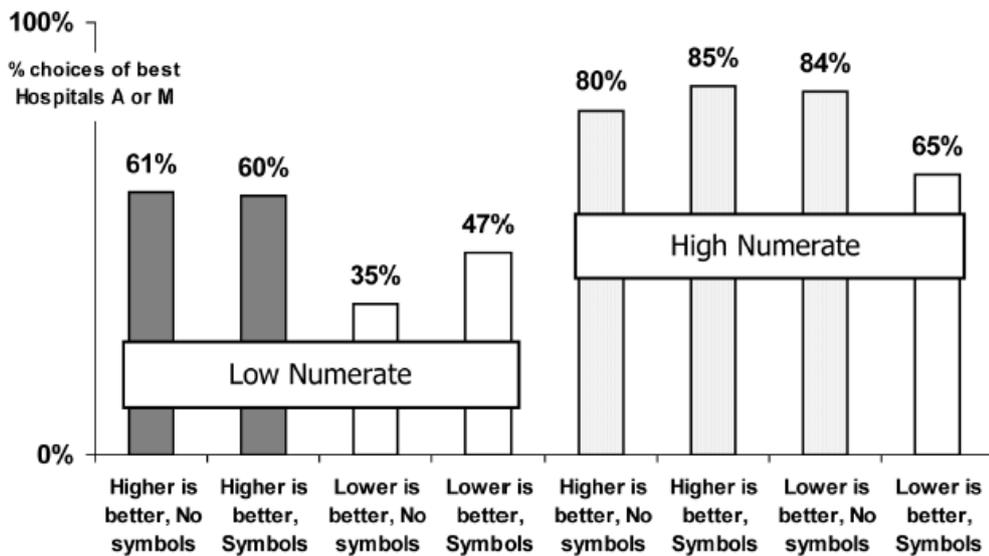


Abbildung 23: Study 3. Choice of the Best Hospitals by Presentation Format and Numeracy

Als Gesamtergebnis der drei Teilstudien kann das Fazit „less is more“ gezogen werden. Die Teilstudien zeigten, dass für eine möglichst hohe Verständlichkeit die folgenden Empfehlungen abgeleitet werden können: (1) Fokus auf die relevanten Informationen und deren Anordnung, (2) grafische Hervorhebung von numerisch dargestellten, indikationsspezifischen und primär relevanten Informationen, (3) Verwendung höherer Zahlen für die Darstellung positiver Ergebnisse (die zusätzliche Verwendung von Symbolen hatte hierbei keinen Einfluss auf die Verständlichkeit mehr gezeigt). Somit sollten entweder nur jeweils relevante Qualitätsinformationen gezeigt werden bzw. die primär relevanten Qualitätsinformationen dargestellt und grafisch-unterstützt aufbereitet werden.

**Sander U et al. (2015) Information presentation features and comprehensibility of hospital report cards: design analysis and online survey among users [10]**

In der Studie von Sander und Kollegen sollten evidenzbasierte Empfehlungen zur Darstellung von Krankenhausqualitätsinformationen (hier exemplarisch die „risk-adjusted mortality rate (RAMR) for coronary angiography“) abgeleitet werden. Hierzu wurden die Darstellungen auf den zehn am häufigsten in Deutschland genutzten Krankenhaus Report Cards im Hinblick auf die Verständlichkeit der Darstellung betrachtet und im Rahmen einer Online-Befragung mit 2.027 Personen evaluiert. Jedem Probanden wurden randomisiert drei der zehn Report Card Designs zugewiesen, die Aufgabe bestand jeweils in der Identifikation des besten Krankenhauses (niedrigste RAMR); gezeigt wurden jeweils fünf Krankenhäuser.

Table 5. Choice of the hospital with the lowest risk-adjusted mortality rate (RAMR).

Information presentation feature	Feature included			Feature not included			Choice of lowest RAMR		Comprehensibility	
	Respondents, n	Selected hospital with the lowest RAMR, n (%)	Comprehensibility, <sup>a</sup> mean (SD)	Respondents, n	Selected hospital with the lowest RAMR, n (%)	Comprehensibility, <sup>a</sup> mean (SD)	$\chi^2$ (df)	P	t (df)	P
Table without symbols	1245	575 (46.18)	3.07 (1.85)	4836	3119 (64.50)	3.77 (1.92)	139.2 (1)	<.001	-11.657 (6079)	<.001
Table with symbols	1787	928 (51.93)	3.58 (1.84)	4294	2766 (64.42)	3.65 (1.96)	82.5 (1)	<.001	-1.300 (6979)	.19
Bar chart without symbols	608	392 (64.47)	2.99 (1.84)	5473	3302 (60.33)	3.70 (1.92)	3.9 (1)	.047	-8.626 (6079)	<.001
Bar chart with symbols	2441	1799 (73.70)	4.11 (1.92)	3640	1895 (52.06)	3.31 (1.86)	286.9 (1)	<.001	16.289 (6079)	<.001
Bar chart with traffic light symbols	1814	1341 (73.93)	4.25 (1.93)	4267	2353 (55.14)	3.36 (1.86)	188.3 (1)	<.001	16.774 (6079)	<.001
Bar chart with thumb symbols	627	458 (73.05)	3.70 (1.83)	5454	3236 (59.33)	3.62 (1.94)	44.4 (1)	<.001	1.030 (6079)	.30
Providers ranked by performance	1221	883 (72.32)	4.29 (1.91)	4860	2811 (57.84)	3.46 (1.89)	85.8 (1)	<.001	13.620 (6079)	<.001
Explicit statement about whether higher or lower values indicate better performance	3017	2079 (68.91)	3.85 (1.97)	3064	1615 (52.71)	3.41 (1.86)	167.3 (1)	<.001	9.112 (6079)	<.001
No statement about scale direction, but range for good quality presented	1220	884 (72.46)	4.04 (1.91)	4861	2810 (57.81)	3.52 (1.92)	87.8 (1)	<.001	8.440 (6079)	<.001
*Incomplete data (N/A labels)	2445	1212 (49.57)	3.32 (1.88)	3636	2483 (68.30)	3.84 (1.93)	214.2 (1)	<.001	-10.436 (6079)	<.001

Tabelle 16: Choice of the hospital with the lowest risk-adjusted mortality rate (RAMR)

Als Ergebnis zeigte sich über alle zehn Report Cards ein Durchschnittswert von 61% korrekter Identifizierung des besten Krankenhauses, 7% hingegen haben jeweils das schlechteste Krankenhaus gewählt. Lediglich jeder dritte Proband (32%) wählte auf allen drei Report Cards das jeweils beste Krankenhaus; 14% hingegen konnten auf keiner Report Card das jeweils beste Krankenhaus identifizieren. Des Weiteren wurde evaluiert, ob das Wahlverhalten auf Report Cards, welche evidenzbasierte Designfeatures beachten, signifikant besser ausgefallen sind im Vergleich zu Report Cards ohne die Beachtung der zehn evidenzbasierten Designfeatures (siehe Tabelle 16). Hier zeigten sich für alle zehn evidenzbasierten Designfeatures signifikant bessere Verständlichkeitswerte; beispielsweise sollten Tabellen mit Symbolen (64% vs. 52%;  $p < .001$ ) oder Leistungserbringer nach Qualitätsergebnissen gerankt dargestellt werden (72% vs. 58%;  $p < .001$ ). Basierend auf den Ergebnissen und der internationalen Evidenz wurden die folgenden fünf Empfehlungen für das Public Reporting in Deutschland abgeleitet: (1) Vermeidung von Tabellen ohne erläuternde Symbole, (2) Darstellung von Balken- bzw. Säulendiagrammen mit erläuternden Symbolen, (3) ausdrücklicher Hinweis, ob niedrige oder hohe Werte eine gute Qualität darstellen bzw. Aufzeigen einer guten Qualitätsergebnisrange, (4) Vermeidung unvollständiger Daten sowie (5) Anordnung von Krankenhäusern nach Performance.

### 3.2.3 Pflegesektor

#### ***Damman OC et al. (2012): Consumers' interpretation and use of comparative information on the quality of health care: the effect of presentation approaches [40]***

Gegenstand der Untersuchung war zu zeigen, welche Präsentationsmerkmale die richtige Interpretation und die intendierte Nutzung von Qualitätsvergleichsdaten zur Gesundheitsversorgung bewirken und welche Rolle dabei bestimmte Nutzercharakteristika spielen. Dabei waren insbesondere folgende Präsentationsmerkmale von Interesse: Balkendiagramme, Sternbewertungen, Sortierung der Daten, Art der Sternbewertung und die Anzahl der Sterne sowie eine Gesamtbewertung der Qualitätsergebnisse (Summenscore). Ebenfalls wurde die Kombination verschiedener Merkmale, wie beispielsweise die Präsentation eines Summenscores in Form einer Sternbewertung, in Bezug auf die Verständlichkeit untersucht. Auf der Seite der Nutzercharakteristika wurden das Alter, das Geschlecht und der Bildungsstand berücksichtigt. Für die Untersuchung wurde eine Conjoint-Analyse zu Effekten von fünf Präsentationsmerkmalen in folgender Form durchgeführt: (1) Kombination von Balkendiagrammen und Sternbewertungen vs. nur Sternbewertungen, (2) alphabetische Sortierung der Leistungserbringer vs. Rangliste entsprechend der Qualitätsergebnisse, (3) Sternbewertung entsprechend der absoluten Werte der Qualitätsergebnisse vs. Sternbewertung entsprechend der relativen Werte der Qualitätsergebnisse, (4) Bewertungen mit drei Sternen vs. Bewertungen mit fünf Sternen sowie (5) Darstellung mit einer Gesamtbewertung der Leistungserbringer mit Summenscore vs. Darstellung ohne Gesamtbewertung der Leistungserbringer. Die Kombination aller Merkmale mit unterschiedlichen Ausprägungen führte zu insgesamt 32 verschiedenen experimentellen Präsentationsformaten. Davon wurden insgesamt 15 Präsentationsformate ausgewählt, Abbildung 24 zeigt ein beispielhaftes Design der Studie. In einer Online-Befragung wurden den Probanden randomisiert jeweils vier Präsentationsformate angezeigt. 438 Probanden beendeten die Online-Befragung vollständig. (1) Die korrekte Interpretation der Informationen (Identifikation des besten und/oder schlechtesten Leistungsanbieters) und (2) die tatsächliche (fiktive) Wahlentscheidung auf der Basis der gegebenen Informationen standen im Zentrum.



Abbildung 24: Example of experimental format: a combination of bar chart and star ratings, a rank ordering of providers, stars based on relative performance, three stars, and no inclusion of an overall rating

Die Ergebnisse der Studie sind übersichtlich in Tabelle 17 dargestellt. Erkennbar ist, dass keine der unterschiedlichen Varianten einen Einfluss auf die Auswahl des besten Leistungserbringers hatte.

Hinsichtlich der Auswahl des schlechtesten Leistungserbringers zeigten sich signifikant bessere Verständlichkeitswerte für die Kombination aus Balkendiagrammen und Sternebewertungen im Vergleich zu einer Darstellung nur mit Sternebewertung ( $\beta=0,43$ ;  $p<0,05$ ) sowie die Darstellungen ohne Gesamtbewertung im Vergleich zu Darstellungen mit einer Gesamtbewertung in der Form von Summenscores ( $\beta=1,81$ ;  $p<0,05$ ). Bezüglich der tatsächlichen Auswahlentscheidung innerhalb des Experiments zeigte sich, dass sich die Sortierung der Anbieter nach Alphabet im Vergleich zur Rangfolge nach Performance positiv auf die Auswahlentscheidung auswirkt ( $\beta=0,51$ ;  $p<0,05$ ), genauso wie die Verwendung von Sternebewertungen mit drei Sternen im Vergleich zu Bewertungen mit fünf Sternen ( $\beta=0,77$ ;  $p<0,05$ ).

Designelemente	Interpretation (best provider)	Interpretation (worst provider)	Effective use (choice)
$\beta$ Combination bar chart and stars <sup>1</sup>	0.08 (0.21)	<b>0.43 (0.15)*</b>	0.03 (0.19)
$\beta$ Ordering by alphabet <sup>1</sup>	0.23 (0.21)	0.15 (0.13)	<b>0.51 (0.19)*</b>
$\beta$ Absolute stars <sup>1</sup>	-0.02 (0.27)	-0.16 (0.18)	0.26 (0.23)
$\beta$ Three stars <sup>1</sup>	0.13 (0.34)	0.23 (0.23)	<b>0.77 (0.31)*</b>
$\beta$ No inclusion overall rating <sup>1</sup>	-0.32 (0.21)	<b>1.81 (0.14)*</b>	-0.26 (0.18)
1Reference group of display = stars only; reference group ordering = rank ordering by performance; reference group type of stars = relative stars; reference group number of stars = five stars; reference group overall rating = inclusion overall rating; *P < 0.05.			

Tabelle 17: Ergebnisse zur Verständlichkeit in Abhängigkeit bestimmter Designelemente

Empfehlungen entsprechend der Ergebnisse dieser Studie: Die korrekte Interpretation der Informationen wurde insbesondere beeinflusst durch (1) die Kombination von Balkendiagrammen und Sternebewertungen sowie (2) das Weglassen von Gesamtbewertungen, z.B. in Form von Summenscores. Die tatsächliche Nutzung der Daten hinsichtlich der Auswahlentscheidung im Sinne der intendierten Nutzung wurde maßgeblich durch die (3) Sortierung der Leistungserbringer nach dem Alphabet und (4) der Verwendung von Sternebewertungen mit drei statt mit fünf Sternen beeinflusst.

**Damman OC et al. (2016) Making comparative performance information more comprehensible: an experimental evaluation of the impact of formats on consumer understanding [41]**

Gegenstand der Studie von Damman et al. war zu untersuchen, wie sich verschiedene Präsentationsformate auf das Verständnis und die Verwendung von vergleichenden Qualitätsinformationen zu Pflegeeinrichtungen auswirken. Hierfür wurden insgesamt 902 Personen mit einem Mindestalter von 18 Jahren mittels Fragebogen befragt. Untersucht wurden die Effekte der folgenden Features auf die Verständlichkeit: (1) displaying an overall performance score (yes/no); (2) displaying a small number of quality indicators (5 vs 9); and (3) displaying different types of evaluative symbols (star ratings, coloured dots and word icons vs numbers and bar graphs). Ebenso wurde der Effekt einer reduzierten Anzahl von gezeigten Pflegeheimen evaluiert (5 vs. 20 Pflegeheime).

Jeder Befragte sah zunächst eine Ergebnisseite mit 20 Pflegeheimen, danach eine reduzierte Ergebnisseite mit 5 Pflegeheimen. Aus einer Vorgängerstudie wurden entsprechend der Patientenpräferenzen insgesamt 9 Qualitätsindikatoren eingeschlossen (z.B. Unabhängigkeit, Reinigungsservice, Essen oder Kompetenz des Personals). Unter den gezeigten Pflegeheimen war jeweils ein Pflegeheim, dass bezüglich der Qualitätsergebnisse allen anderen überlegen war (Top-Performer) und zwei weitere, die jeweils häufiger über dem Durchschnitt lagen als der Rest der gezeigten Heime (Top 3). Das Design war angelehnt an die staatliche Seite „kiesBeter“ (siehe Abbildung 25). Der Fokus lag auf der richtigen Auswahl der Top 3-Performer.



Abbildung 25: Experimental Formats

Als Ergebnis zeigte sich, dass in der Version mit 20 Pflegeheimen 69,5% der Befragten die Top 3-Performer identifizieren konnten, in der reduzierten Version mit 5 Anbietern hingegen 80,2% ( $p < .001$ ). Im Gegensatz dazu zeigte die Reduktion von Qualitätsindikatoren keine signifikanten Effekte. Gleiches gilt für die Identifikation des besten Pflegeheims, des schlechtesten Pflegeheims und die tatsächliche Auswahlentscheidung, die bei der reduzierten Version mit 5 Pflegeheimen jeweils signifikant (jeweils  $p < .001$ ) besser ausfielen. Die Abbildung eines Summescores für die Gesamtbewertung eines Pflegeheims hatte ebenfalls einen signifikanten Einfluss auf die richtige Identifikation der Top 3-Performer im Vergleich zur Version ohne Gesamtscore sowohl bei der Darstellung von 20 Pflegeheimen (79,8% vs. 59,5%;  $p < 0,01$ ) als auch 5 Pflegeheimen. Auch die Verwendung evaluativer Symbole (Zahlen, Balken, Sternebewertungen, Farbige Punkte, Wort-Symbole) wirkte sich jeweils signifikant positiv auf die Auswahl der Top-3 Performer, des besten und des schlechtesten Pflegeheims sowie die tatsächliche Wahlentscheidung aus (siehe Abbildung 26). Die Verwendung evaluativer Symbole (coloured dots, word icons) hatte ebenfalls einen signifikanten Einfluss auf die Verständlichkeit. Bei der Präsentation mit farbig hinterlegten Punkten (siehe rechte Seite in Abbildung 25) identifizierten 84,3% die Top 3-Performer, im Vergleich zu 76,6% bei Wort-Symbolen, 70,6% bei Sternebewertungen, 62,0% bei Zahlen sowie 54,2% bei balkenbasierten Darstellungen.

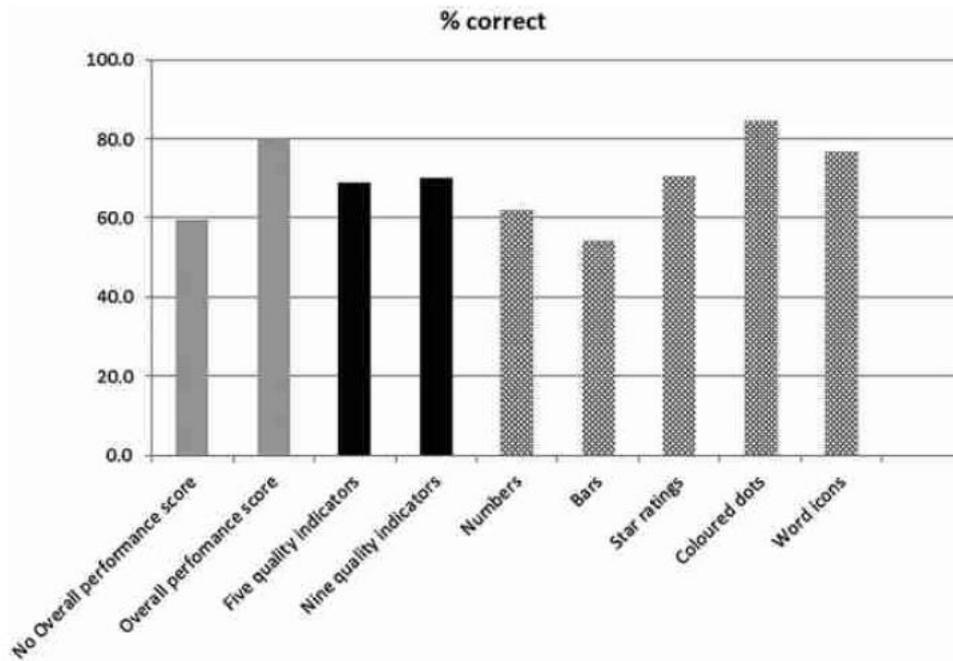


Abbildung 26: Percentages of respondents correctly selecting the top three best-performing nursing homes from a set of 20 nursing homes, for the different between-subjects manipulations of presentation approaches

Die Verständlichkeitswerte für die korrekte Identifizierung der Top-3 Performer bei einer Darstellung von nur 5 Pflegeheimen zeigt die folgende Abbildung 27. Erkennbar sind bessere Ergebniswerte, wenn eine Gesamtscore präsentiert wird, eine reduzierte Anzahl an Qualitätsindikatoren gezeigt wird sowie für die Designarten Wort-Symbole, Sternebewertungen sowie farbig hinterlegten Punkte.

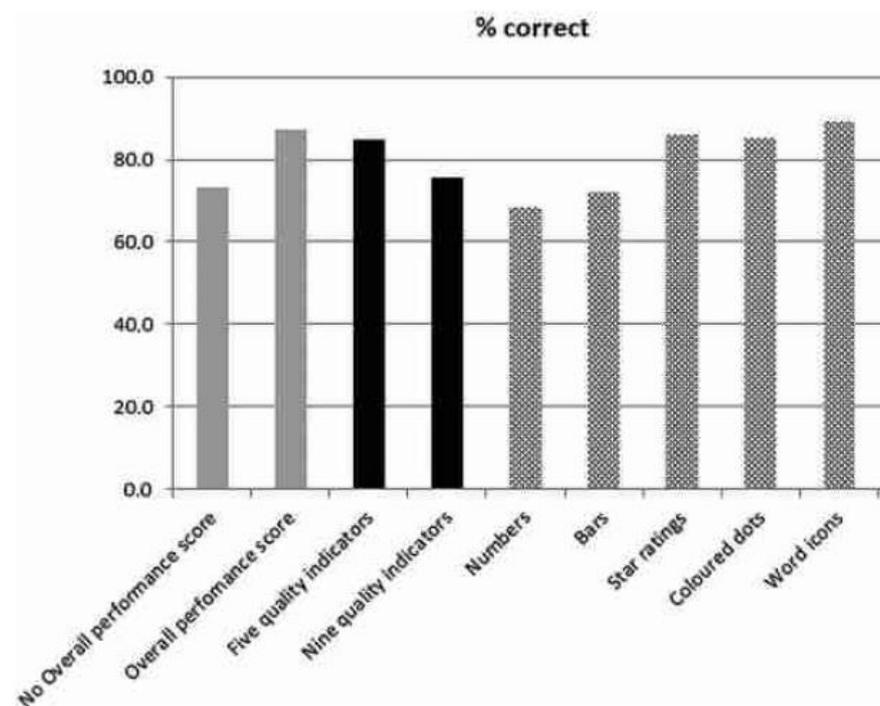


Abbildung 27: Percentages of respondents correctly selecting the top three of best-performing nursing homes from a set of five nursing homes, for the different between-subjects manipulations of presentation approaches

Insgesamt kann daher als Ergebnis gezeigt werden, dass insbesondere die Präsentation eines Gesamt-Performancescores sowie die Verwendung evaluativer Symbole (coloured dots, word icons) eine positive Wirkung auf die Verständlichkeit aufzeigen (siehe Tabelle 18). Wurde kein Gesamt-Performancescore angezeigt, zeigten lediglich die Wort-Symbole sowie die farbigen Punkte positive Wirkung auf die richtige Identifikation der Top 3-Performer (0,063;  $p < 0,01$ ). Gleiches gilt für die Identifikation des besten Pflegeheims (0,015;  $p < 0,05$ ), sowohl in der Version mit 20 Pflegeheimen als auch in der Version mit 5 Pflegeheimen. Die Autoren empfehlen daher als Implikation für die praktische Umsetzung von Public Reporting die Verwendung von Summenscores, evaluativer Symbole (Wort-Symbole, farbige Punkten) sowie die Reduktion der Darstellungen auf wenige Anbieter.

	Selection of top three nursing homes Effect size†	Selection of best nursing home Effect size†	Selection of worst nursing home Effect size†	Choice of nursing home Effect size†
Main effects of manipulations in presentation formats				
Display of overall performance score	0.039**	0.005*	0.038**	0.001
Small number of quality indicators	0.001	0.004	0.002	0.000
Type of evaluative symbols	0.034**	0.016*	0.015*	0.021*
Interaction effects of manipulations in presentation formats				
Display of overall performance score×small number of quality indicators	0.003	0.003	0.000	0.001
Display of overall performance score×type of evaluative symbols	0.063**	0.015*	0.010	0.007
Small number of quality indicators×type of evaluative symbols	0.018*	0.010	0.006	0.003
Display of overall performance score×small number of quality indicators×type of evaluative symbols	0.012	0.013*	0.009	0.018*
Main effects of vulnerability-related consumer characteristics				
Educational level	0.002	0.003	0.000	0.007
Health numeracy	0.066**	0.034**	0.068**	0.033**
Subjective health literacy	0.000	0.000	0.005	0.000
Patient activation	0.008	0.008	0.006	0.008

\*Significant effect with significance level of 0.05.  
\*\*Significant effect with a significance level of 0.01.  
†Partial eta squared, which is a measure of effect size for use in analyses of variance. The common thresholds for the magnitude of effect are 0.01=small; 0.06=medium; 0.14=large.

*Tabelle 18: Main findings from the analyses of variance testing between-subjects manipulations of presentation formats on the different outcome variables (realistic version of 20 providers)*

### **Werner RM et al. (2016) Changes in Consumer Demand Following Public Reporting of Summary Quality Ratings: An Evaluation in Nursing Homes [42]**

Die Analyse von Werner et al. untersucht den Zusammenhang zwischen der Verwendung eines Summenscores in Form von Sternbewertungen und dem Marktanteil von Pflegeheimen in Abhängigkeit der Qualität dieser Heime, gemessen an der Bewertung von einem (schlechtester Wert) bis fünf Sternen (besten Wert). Damit soll gezeigt werden, inwieweit Patienten sich von solchen Summenscores bei der Auswahl von Pflegeheimen in der Realität leiten lassen. Dafür wurden Daten zu 2.316.649 Einweisungen von 16.147 Pflegeheimen in den USA mit einer Medicare- bzw. Medicaid-Zertifizierung im Zeitraum zwischen 2005 und 2010 ausgewertet.

Gleichzeitig müssen diese eine Sternbewertung auf dem Bewertungsportal der Centers for Medicare & Medicaid Services (CMS) ausweisen. In einer Regressionsanalyse wurde der Einfluss der Einführung des Gesamt-Performancescores durch die Darstellung mittels Sternbewertung im Jahr 2008 auf die Auswahl eines Pflegeheims untersucht.

	General Information	Inspections and Complaints	Staffing	Quality Measures	Penalties
	Nursing Home A		Nursing Home B		Nursing Home C
Overall Rating	★★★★★ Average		★★★★★ Much Below Average		★★★★★ Below Average
Health Inspection Rating	★★★★★ Average		★★★★★ Below Average		★★★★★ Much Below Average
Staffing Rating	★★★★★ Much Below Average		★★★★★ Much Below Average		★★★★★ Above Average
Quality Measure Rating	★★★★★ Much Above Average		★★★★★ Above Average		★★★★★ Average

Abbildung 28: Example of the Star Ratings Available for Nursing Homes on the Nursing Home Compare Website

Als Ergebnis der Studie zeigte sich, dass die Chance in ein Pflegeheim mit höchster Qualität (5 Sterne) eingewiesen zu werden, vor Einführung der Sterneratings niedriger war (-0,403;  $p < 0,01$ ) als nach der Einführung von Sterneratings (0,078;  $p < 0,01$ ) (s.

	All Admissions	PAC Admissions	LTC Admissions
2 stars	-0.051*** (0.003)	-0.055*** (0.003)	-0.031*** (0.009)
3 stars	-0.083*** (0.003)	-0.089*** (0.003)	-0.077*** (0.009)
4 stars	-0.182*** (0.003)	-0.192*** (0.003)	-0.160*** (0.010)
5 stars	-0.403*** (0.004)	-0.426*** (0.004)	-0.296*** (0.014)
post2008*2 stars	0.023*** (0.005)	0.024*** (0.005)	0.030* (0.018)
post2008*3 stars	0.017*** (0.005)	0.019*** (0.005)	0.012 (0.019)
post2008*4 stars	0.017*** (0.005)	0.021*** (0.005)	-0.052** (0.019)
post2008*5 stars	0.078*** (0.007)	0.081*** (0.007)	0.077*** (0.026)
N	180,148,037	164,741,202	15,406,835

Notes: Coefficients from choice models such as these are typically interpreted as changes in utility. Robust standard errors in parentheses. All regressions control for the driving distance between the nursing home resident's home zip code and the nursing home, the nursing home's ownership (for profit, not for profit, or government), total number of beds, occupancy rates, whether it is hospital based, percentage of Medicare patient days, and percentage of Medicaid patient days. Differences between post2008\*5 stars and all other interactions are statistically significantly different from zero ( $p < .001$ ).  
\* $p < .10$ , \*\* $p < .05$ , \*\*\* $p < .01$ .

Tabelle 19). Eine Betrachtung der Marktanteile ergab diesbezüglich, dass 1-Stern-Pflegeheime 8% Marktanteil gegenüber der Zeit vor Einführung des Summenscores auf der Pflegeheimvergleichsseite des CMS verloren, wohingegen 5-Sterne-Pflegeheime 6% Marktanteile gewonnen haben.

	<i>All Admissions</i>	<i>PAC Admissions</i>	<i>LTC Admissions</i>
2 stars	-0.051*** (0.003)	-0.055*** (0.003)	-0.031*** (0.009)
3 stars	-0.083*** (0.003)	-0.089*** (0.003)	-0.077*** (0.009)
4 stars	-0.182*** (0.003)	-0.192*** (0.003)	-0.160*** (0.010)
5 stars	-0.403*** (0.004)	-0.426*** (0.004)	-0.296*** (0.014)
post2008*2 stars	0.023*** (0.005)	0.024*** (0.005)	0.030* (0.018)
post2008*3 stars	0.017*** (0.005)	0.019*** (0.005)	0.012 (0.019)
post2008*4 stars	0.017*** (0.005)	0.021*** (0.005)	-0.052** (0.019)
post2008*5 stars	0.078*** (0.007)	0.081*** (0.007)	0.077*** (0.026)
N	180,148,037	164,741,202	15,406,835

*Notes:* Coefficients from choice models such as these are typically interpreted as changes in utility. Robust standard errors in parentheses. All regressions control for the driving distance between the nursing home resident's home zip code and the nursing home, the nursing home's ownership (for profit, not for profit, or government), total number of beds, occupancy rates, whether it is hospital based, percentage of Medicare patient days, and percentage of Medicaid patient days. Differences between post2008\*5 stars and all other interactions are statistically significantly different from zero ( $p < .001$ ).  
\* $p < .10$ , \*\* $p < .05$ , \*\*\* $p < .01$ .

*Tabelle 19: The Log-Odds of Being Admitted to a Nursing Home with 2, 3, 4, and 5 Stars before (Coefficients on the Uninteracted Terms) and after the Star-Based Ratings Were Released in 2008 (Coefficients on the Interaction Terms), All Compared to Being Admitted to a 1-Star Facility*

Die Autoren schlussfolgern deshalb, dass die Verwendung von Pflegeheim-bezogenen Report Cards durch die Darstellung eines Summenscores in Form von Sternebewertungen erhöht wird und gleichzeitig deren Verständnis verbessert wird. Eine wichtige Limitation bei der Studienbetrachtung liegt allerdings darin, dass im Experiment keine direkte Verbindung zwischen der Auswahl eines Pflegeheims und der tatsächlichen Verwendung des Summenscores auf Seiten der Patienten hergestellt werden kann. Die Auswahlentscheidung könnte auch aufgrund weiterer, in der Studie nicht betrachteter Faktoren getroffen worden sein.

### 3.3 Gesamtergebnis

#### 3.3.1 Ambulanter Sektor bzw. Arzzebene-bezogenes Public Reporting

Wie oben beschrieben, wurden insgesamt sechs Studien identifiziert, die sich mit der Verständlichkeit der Darstellung von Qualitätsinformationen über individuelle Ärzte auseinandersetzen. Die folgende Tabelle 20 gibt einen zusammenfassenden Überblick über die unterschiedlichen Empfehlungen aus den einzelnen Studien. Teilweise sind diese Empfehlungen einheitlich und konsistent über die Literatur hinweg; so beispielsweise zur Verringerung der Komplexität durch die Anzeige nur ausgewählter Ergebnisdaten oder Leistungserbringer [29] oder der Verwendung von Composite/Summary Measures [26]. Bezüglich der Darstellung von Kosteninformationen kann abgeleitet werden, dass diese in Form von Kreisen bzw. Sternen bzw. als absoluter Betrag dargestellt werden sollten [28]. Da Patientenkommentare einen negativen Einfluss auf die Entscheidungsqualität aufweisen [29, 31], sollten entsprechende Entscheidungshilfen (Tagged Comments, Tagged Comments plus trained Navigator) integriert werden [31]. Bezüglich des Designs der Darstellung sind die Empfehlungen teils sehr heterogen.

#### **Empfehlungen für ein verständliches Public Reporting**

- Eine geringere Informationskomplexität durch weniger Ergebnisdaten und weniger angezeigten Leistungserbringern erhöht die Verständlichkeit [29]

- Verwendung von Summenscores (Summary Measures, Roll-up Measures)<sup>3</sup> [26]
- Verbindung von Text und Symbolen sowie die visualisierte Darstellung mit Balkendiagrammen [27]
- Keine Darstellungen mit ausschließlich detaillierten zahlenbasierten Informationen [27]
- Keine reinen text- bzw. zahlenbasierten Tabellen [27]
- Qualitätssignale sind möglichst stark darzustellen [28]
- Erläuternde bzw. interpretierende Labels zeigen positive Verständlichkeitseffekte [28]
- Verwendung des Formats Häufigkeiten und Prozentangaben mit einem Symbol, reine Sternebewertung oder Sternebewertung mit Wort-Symbol [30]
- Numerische Angaben (z.B. Prozentwerte) sollten mit Symbolen graphisch unterstützt werden [30]
- Grafische Darstellung von Qualitätsinformationen mit Symbolen oder Sternebewertungen [30]
- Kosten in Form von Kreisen bzw. Sternen oder als absoluten Betrag darstellen [28]
- Keine ausschließliche Darstellung von Kosten in Dollarzeichen [28]
- Die Darstellung von Patientenkommentaren verringert die Entscheidungsqualität [29, 31]
- Zur Darstellung von Patientenkommentaren sollten Entscheidungshilfen (Tagged Comments, Tagged Comments plus trained Navigator) integriert werden [31]

*Tabelle 20: Empfehlungen für ein verständliches Public Reporting - Individualebene*

---

<sup>3</sup> Unabhängig davon, ob diese dann noch mit detaillierteren Ergebnissen zu einzelnen Indikatoren ergänzt werden oder nicht

### 3.3.2 Stationärer Sektor

Wie oben beschrieben, setzten sich insgesamt elf Studien mit der Verständlichkeit der Darstellung von Qualitätsinformationen über Krankenhäuser auseinander. Die folgende Tabelle 21 gibt einen zusammenfassenden Überblick über die unterschiedlichen Empfehlungen aus den einzelnen Studien. Die hier ableitbaren Empfehlungen beziehen sich insbesondere auf die Reduzierung der Komplexität der Darstellung (d.h. die Anzahl der gezeigten Leistungserbringer) als entscheidende Variable [6], dem Fokus auf relevante Informationen und deren Anordnung [39], der Darstellung eines Qualitätskonzepts als Grundlage für die Darstellung und Nennung von Kategorien [34] sowie der Verwendung einer einfachen und selbsterklärenden Sprache sowie der Vermeidung von Fachwörtern [34, 36, 37]. Des Weiteren sollten keine Krankenhäuser separat bzw. individuell in einzelnen Bereichen gezeigt, sondern komprimiert gegenübergestellt werden [32]. Es wird des Weiteren empfohlen, dass auf Basis der Einzelinformationen ein Gesamtergebnis bzw. ein Composite Measure errechnet wird [32] sowie eine Anordnung der Krankenhäuser als Rangliste bzw. nach Performance erfolgt [10, 32]. Bezüglich des Designs sollten Tabellen ohne erläuternde Symbole vermieden werden [10], vielmehr sollen visualisierte Wort-Symbole<sup>4</sup> Verwendung finden [33]. Die unterschiedlichen Studien empfehlen des Weiteren heterogen die Verwendung von Tabellen mit Zahlenwerten (sofern als Gesamtergebnis und Rangliste ausgestaltet) oder Ampelsymbolen [32], die grafische Hervorhebung von numerisch dargestellten Informationen [39] oder auch die Darstellung von Balken- bzw. Säulendiagramme mit erläuternden Symbolen [10]. Sofern Zahlen dargestellt werden, sollten höhere Zahlen für die Darstellung positiver Ergebnisse verwendet werden [39]. Alternativ sollte ein Hinweis erfolgen, ob niedrige oder hohe Werte eine gute Qualität darstellen bzw. eine gute Qualitätsergebnisrange aufzeigen [10].

<b>Empfehlungen für ein verständliches Public Reporting</b>
<ul style="list-style-type: none"> <li>• Komplexität der Darstellung (Anzahl der gezeigten Leistungserbringer) als entscheidende Variable [6]</li> <li>• Fokus auf relevanten Informationen und deren Anordnung [39]</li> <li>• Darstellung in Form eines Qualitätskonzepts unter Darstellung und Nennung von Kategorien [34]</li> <li>• Keine separate Darstellung von Qualitätsergebnissen für jedes Krankenhaus individuell [32]</li> <li>• Klinische Qualitätsergebnisse in der Form numerischer Darstellungen in Prozentangaben vs. reiner Text bzw. ausschließlicher Sternendarstellung [13]</li> <li>• Einfache Sternebewertung als verständlichstes Präsentationsformat [32]</li> <li>• Ampelsymbole verständlicher als Textform, Sternenform oder Schulnotensystematik [32]</li> <li>• Tabellen mit Zahlenwerten, sortiert nach einem errechneten Gesamtergebnis und als Rangliste [32]</li> <li>• Detaillierte Indikatorenergebniswerte in Verbindung mit Ranglisten oder Ampelsymbolen [32]</li> <li>• Grafische Hervorhebung von numerisch dargestellten, indikationsspezifischen und primär relevanten Informationen [39]</li> <li>• Vermeidung von Tabellen ohne erläuternde Symbole [10]</li> <li>• Darstellung von Balken- bzw. Säulendiagrammen mit erläuternden Symbolen [10]</li> <li>• Darstellung als Gesamtscore mittels Sternenform [32]</li> <li>• Verwendung visualisierter Wort-Symbole<sup>5</sup> für medizinische Qualitätsinformationen [33]</li> <li>• Einfache Sprache, geringer Fachwortanteil oder selbsterklärende Begriffe [34, 36, 37]</li> </ul>

<sup>4</sup> z.B. „Better“ hinterlegt mit einem gelben Pfeil

<sup>5</sup> z.B. „Better“ hinterlegt mit einem gelben Pfeil

- Verwendung einer einfachen Darstellung [36]
- Verwendung höherer Zahlen für die Darstellung positiver Ergebnisse [39] bzw. Hinweis, ob niedrige oder hohe Werte eine gute Qualität darstellen bzw. Aufzeigen einer guten Qualitätsergebnisrange [10]
- Kostendaten als visualisierte Wort-Symbole oder in absolut dargestellten Dollarwerten [33]
- Vermeidung unvollständiger Daten [10]
- Anordnung von Krankenhäusern nach Performance [10]

*Tabelle 21: Empfehlungen für ein verständliches Public Reporting – Krankenhausebene*

### 3.3.3 Pflegesektor

Wie oben beschrieben, bezogen sich insgesamt drei Studien auf die Verständlichkeit der Darstellung von Qualitätsinformationen über Pflegeheime. Die folgende Tabelle 22 gibt einen zusammenfassenden Überblick über die unterschiedlichen Empfehlungen aus den einzelnen Studien. Die hier ableitbaren Empfehlungen beziehen sich auch hier auf die Reduzierung der Komplexität der Darstellung durch die Reduktion der gezeigten Pflegeheime [41]. Des Weiteren wurde auch hier die Verwendung von Summenscores [40–42] empfohlen genauso wie die Verwendung evaluativer Symbole (Wort-Symbole, farbige Punkten) [41]. Eine Studie zeigte ebenso die verständliche Darstellung mittels einer Kombination aus Balkendiagrammen und Sternebewertungen im Vergleich zu einer Darstellung nur mit Sternebewertung [40].

#### ***Empfehlungen für ein verständliches Public Reporting***

- Reduktion der Darstellungen auf wenige Anbieter [41]
- Kombination aus Balkendiagrammen und Sternebewertungen im Vergleich zu einer Darstellung nur mit Sternebewertung [40]
- Verwendung evaluativer Symbole (Wort-Symbole, farbige Punkte) [41]
- Verwendung von Summenscores im Vergleich zu Darstellungen ohne einer Gesamtbewertung [40–42]
- Keine Unterschiede bei den folgenden Gegensatzpaaren: alphabetische Sortierung der Leistungserbringer vs. Ranking, Sternebewertung entsprechend der absoluten Werte vs. Sternebewertung entsprechend der relativen Werte der Qualitätsergebnisse, Bewertungen mit drei Sternen vs. Bewertungen mit fünf Sternen [40]

*Tabelle 22: Empfehlungen für ein verständliches Public Reporting - Pflegeheimebene*

### 3.4 Gegenüberstellung der Ergebnisse

Wie oben beschrieben, wurden drei systematische Übersichtsarbeiten identifiziert, die sich ebenfalls, allerdings in abgewandelter Form, mit der Frage nach der wirksamen Darstellung von Qualitätsinformationen über medizinische Leistungserbringer auseinandersetzen. Diese werden nun im Folgenden beschrieben, wobei jeweils nur auf die zentralen Ergebnisse eingegangen wird.

#### **Faber M et al. (2009) Public Reporting in Health Care: How Do Consumers Use Quality-of-Care Information? A Systematic Review [18]**

Die Verständlichkeit der Informationen und die Wahlentscheidungen wurden besser, wenn weniger Informationen präsentiert oder die wichtigsten Informationen hervorgehoben wurden. Verschiedene Symbole zeigten dabei unterschiedliche Wirkung, wobei nur für die wichtigsten Informationen Symbole (hier: Ampel) verwendet werden sollten. Die Verwendung einfacher Sprache wirkte sich ebenfalls positiv auf die Auswahlentscheidung aus (siehe Tabelle 23).

First Author, Year	Information Types*	Effect
Hibbard, 1996 <sup>30</sup>	B	Explanatory information about condition-specific performance indicators representing quality of HPs: Did not affect comprehension of performance indicators (stage 2) Increased overall salience of performance indicators (stage 3)
Hibbard, 2000 <sup>31</sup>	A	Framing CAHPS information with a risk message increased comprehension, value, and importance given to CAHPS information, while explanatory messages tended to result in the opposite (stages 2 and 3)
Hibbard, 2001 <sup>32</sup>	A	Error rate decreased when HPs were ordered by quality within cost strata or when star ratings were used. No effect of bar charts on error rates (stage 2)
Hibbard, 2002 <sup>20</sup>	A	No effect of evaluative labels added to bar charts on choosing a higher-rated HP (stage 4) Ordering HPs by performance within cost strata resulted in more choices for better-performing HPs (stage 4) More importance was given to trend data on HP performance from previous years and less to current levels of performance when trends were presented with explicit + and - (stage 4)
Peters, 2007 <sup>37</sup>	B	Comprehension of cost and quality information increased when less information was presented or when the most important information was highlighted (stage 2) Choice of a better-quality hospital increased when less information was presented or when the most important information was highlighted (stage 4) Different symbols had different effects on comprehension and choice Increased comprehension and better choices induced by adapted presentation strategies were more pronounced for participants with lower numeracy
Uhrig, 2006 <sup>36</sup>	A, B	Integrated comparative information on costs, benefits, and quality increased the use of quality information, and the quality information was rated as being more important (stage 3) Simplified terminology, increased use of white space, improved formatting, and shorter length were effective for selecting a high-quality HP more often (stage 4)

Stage 1, awareness; stage 2, knowledge; stage 3, attitude; stage 4, behavior.

\*Type A, consumer or patient experiences; type B, clinical performance-based quality measures; type C, expert or peer-assessed quality measures.

CAHPS indicates consumer assessment of health care providers and systems; HP, health plan.

Tabelle 23: Effects of Presentation Formats for Quality-of-Care Information on Consumers

#### **Sander U et al. (2015) Information presentation features and comprehensibility of hospital report cards: design analysis and online survey among users [10]**

Sander und Kollegen leiteten im Rahmen eines systematischen Reviews literaturgestützte Empfehlungen für das Design von Krankenhausqualitätsinformationen auf Report Cards ab. Diese wurden in 12 Kategorien zusammengefasst, so beispielsweise zur Verwendung evaluativer Tabellen mit Symbolen, Tabellen ohne Symbole, Balkendiagrammen oder auch der Anordnung von Leistungserbringern (siehe Abbildung 29). So sollten evaluative Tabellen mit grafischen Elementen wie Sternen dargestellt werden anstelle von alleinigen Säulen- oder

Balkendiagrammen; letztere zeigen in der alleinigen Form häufig die niedrigsten Verständniswerte auf. Tabellen-ähnliche Darstellungen ohne grafische Elemente sollten vermieden werden. Grafische Hervorhebungen sollten erfolgen, allerdings nicht inflationär sondern ausschließlich in Bezug auf die jeweils wichtigen Informationen. Ebenso wurde in vielen Studien gezeigt, dass Leistungserbringer entsprechend ihrer Performance angeordnet werden sollten. Sofern numerische Werte gezeigt werden, sollten höhere Werte gute Ergebnisse repräsentieren.

Category	Recommendations and results
Evaluative table with symbols	Consider using a table design such as the “evaluative table with stars” rather than a bar chart [29] Evaluative tables using words or stars are superior to numerical tables [29] Physicians preferred formats that used traffic light symbols to code the value of indicators (numerical table with traffic lights) [30]
Tables without symbols	Graphic displays were more helpful to users than text-only tables [31]
Bar charts	Bar charts were commonly used (43% of public reporting websites) [32]
Bar charts without symbols	Comprehension was lowest when data were presented in bar charts [30] Standard bar charts were not well-liked by respondents and led to the lowest levels of comprehension [29]
Bar charts with symbols	Symbols and bar charts should be used [31] A combination of bar charts and star ratings facilitated correct interpretation by users [32] Adding stars to bar charts increases comprehension significantly [33]
Symbols	Participants liked to use symbols to identify the best surgeon [31] Physicians preferred formats that used symbols (eg, traffic lights) [30] Star-only formats should be used in preference to numerical values [34] Only important information should be made easier to evaluate using symbols [35]
Evaluative word labels	Adding evaluative labels to bar charts did not increase comprehension [33]
Highlighting	Color-coding important information improves comprehension [36] Highlighting information about quality resulted in greater understanding [37] Presentation formats which highlighted key messages increased comprehension [38]
Order of providers	Physicians prefer presentation formats that combine individual indicator values with evaluative features such as rankings [30] Comprehension of respondents who were low in numeracy was significantly improved by the ordered compared to the unordered condition [35] Providers should be ranked by performance [12] Ranking plans by performance significantly decreased errors in interpreting data [33] Ranking by performance increased the frequency with which users chose higher-performing services [15] Providers should be ranked in descending order of quality, as this was valued by participants and increased their comprehension [36] One of the more powerful display strategies is to rank providers in terms of performance [33] When providers were ordered alphabetically participants were more likely to make effective use of the data (ie, choose the best provider) than when providers were ordered by performance [32]
High values indicate good performance	Performance data should be displayed such that high values always represent high performance [35] Numeric tables and bar charts often led respondents to conclude that the worst performing nursing homes (those with the higher percentages) were the best, notwithstanding the warning label at the top [29]
State explicitly whether high or low values indicate good performance	It should be stated explicitly whether high or low values indicate good performance, regardless of the direction of the scale [8,36]
Incomplete data (“N/A” as a value)	Incomplete data (missing values) have a negative influence on provider assessment and the potential to influence a decision [29]

Abbildung 29: Features of the presentation of information on health care report cards from previous studies

### **Kurtzman ET & Greene J (2016) Effective presentation of health care performance information for consumer decision making: A systematic review [9]**

Kurtzmann und Greene führten einen systematischen Review in sechs Datenbanken durch, um effektive Präsentationsstrategien für das Public Reporting zusammenzufassen. Als Ergebnis der Studie schlussfolgerten die

Autoren, dass die Verständlichkeit erhöht wird, sofern die Komplexität des Designs verringert wird. Eine Vereinfachung kann durch mehrere Wege erzielt werden, so beispielsweise durch die Reduzierung der Auswahloptionen (d.h. Leistungserbringer) bzw. der Attribute je Auswahloption, eine positiv ausgerichtete Darstellung (höhere Werte repräsentieren bessere Ergebnisse), die Vermeidung von Fachsprache, die Nutzung evaluativer Elemente (z.B. Sterne) und die Einbettung der Ergebnisse in einen gewöhnlichen Kontext. Im Detail wurde bezüglich des Designs erstens angeführt, dass numerische Formate incl. Text (z.B. Häufigkeiten dargestellt als „1 in 250“) nicht so verständlich waren wie einfache Formate mittels grafischer Darstellungen oder bekannter Icons. Zweitens wurden grafische Darstellungen als verständlicher als numerische Formate identifiziert; bis auf die Darstellung von Balken- und Säulendiagrammen, die sich häufig als problematisch herausstellte.

Drittens hatten evaluative bzw. interpretierende Elemente (z.B. Kreise, Sterne, Word Icons) positiven Einfluss auf das Verständnis der Informationen. Allerdings wurde auch hier kein Design als allzeit überlegen eingestuft („yet, no single display was found to be superior“). Bezüglich der Darstellung von Kosten führte die Darstellung von \$-Zeichen zu negativen Ergebnissen. Die Darstellung in Ampelfarben half Teilnehmern, die jeweils beste Option zu erkennen (allerdings hier in Bezug auf die Wahl von gesunder Nahrung).

## 4 Zusammenfassung und Diskussion der Ergebnisse

Im Rahmen der hier durchgeführten Untersuchung erfolgte eine evidenzgestützte Ableitung von Handlungsempfehlungen zur Qualitätsdarstellung für das Public Reporting durch die systematische Identifizierung von Erfolgsfaktoren und Barrieren. Hierfür wurde ein systematischer Literaturreview in den Datenbanken von Medline (via PubMed) und The Cochrane Library durchgeführt. Nach einer Volltextanalyse wurden insgesamt 23 Studien als relevant eingestuft, davon 20 Primärstudien und 3 Sekundärstudien. Die 20 Primärstudien beziehen sich sowohl auf die Darstellung individueller Ärzte (N=6), Krankenhäuser (N=11) als auch Pflegeheimen (N=3).

Als Gesamtergebnis für die Darstellung von Qualitätsinformationen über *individuelle Ärzte* konnten unterschiedliche Empfehlungen abgeleitet werden; so beispielsweise zur Verringerung der Komplexität durch die Anzeige nur ausgewählter Ergebnisdaten oder Leistungserbringer [29] oder der Verwendung von Composite/Summary Measures [26]. Bezüglich der Darstellung von Kosteninformationen wurde abgeleitet, dass diese in Form von Kreisen bzw. Sternen bzw. als absoluter Betrag dargestellt werden sollten [28]. Patientenkommentare weisen einen negativen Einfluss auf die Entscheidungsqualität auf [29, 31] und sollten daher durch entsprechende Entscheidungshilfen (Tagged Comments, Tagged Comments plus trained Navigator) begleitet werden [31]. Bezüglich des Designs der Darstellung sind die Empfehlungen teils sehr heterogen.

Für die Darstellung von Qualitätsinformationen über *Krankenhäuser* bezogen sich die abgeleiteten Empfehlungen insbesondere auf die Reduzierung der Komplexität der Darstellung (d.h. die Anzahl der gezeigten Leistungserbringer) [6], dem Fokus auf relevante Informationen und deren Anordnung [39], der Darstellung eines Qualitätskonzepts als Grundlage für die Darstellung und Nennung von Kategorien [34], die Verwendung einer einfachen und selbsterklärenden Sprache sowie die Vermeidung von Fachwörtern [34, 36, 37]. Des Weiteren sollten keine Krankenhäuser separat bzw. individuell in einzelnen Bereichen gezeigt werden, sondern komprimiert gegenübergestellt werden [32]. Es wird des Weiteren empfohlen, dass auf Basis der Einzelinformationen ein Gesamtergebnis bzw. ein Composite Measure errechnet wird [32] sowie eine Anordnung der Krankenhäuser als Rangliste bzw. nach Performance erfolgt [10, 32]. Bezüglich des Designs sollen Tabellen ohne erläuternde Symbole vermieden werden [10], vielmehr sollen visualisierte Wort-Symbole<sup>6</sup> Verwendung finden [33]. Die unterschiedlichen Studien empfehlen des Weiteren heterogen die Verwendung von Tabellen mit Zahlenwerten (sofern

<sup>6</sup> z.B. „Better“ hinterlegt mit einem gelben Pfeil

als Gesamtergebnis und Rangliste ausgestaltet) oder Ampelsymbolen [32], die grafische Hervorhebung von numerisch dargestellten Informationen [39] oder auch die Darstellung von Balken- bzw. Säulendiagramme mit erläuternden Symbolen [10]. Sofern Zahlen dargestellt werden, sollten höhere Zahlen für die Darstellung positiver Ergebnisse verwendet werden [39]. Alternativ sollte ein Hinweis erfolgen, ob niedrige oder hohe Werte eine gute Qualität darstellen bzw. eine gute Qualitätsergebnisrange aufzeigen [10].

Als Empfehlung für die Darstellung von Qualitätsinformationen über *Pflegeheime* konnte herausgearbeitet werden, dass sich auch hier die Empfehlungen auf die Reduzierung der Komplexität der Darstellung durch die Reduktion der gezeigten Pflegeheime beziehen [41]. Des Weiteren wurde auch hier die Verwendung von Summenscores [40–42] empfohlen, genauso wie die Verwendung evaluativer Symbole (Wort-Symbole, farbige Punkten) [41]. Eine Studie zeigte die verständliche Darstellung mittels einer Kombination aus Balkendiagrammen und Sternbewertungen im Vergleich zu einer Darstellung nur mit Sternbewertung [40].

Die genannten Empfehlungen entsprechen bzw. ergänzen die Ergebnisse anderer systematischer Übersichtsarbeiten, die sich einer ähnlichen Fragestellung angenommen haben. So zeigten *Faber und Kollegen* [18], dass die Verständlichkeit der Informationen und der Wahlentscheidungen besser wurde, wenn weniger Informationen präsentiert oder die wichtigsten Informationen hervorgehoben wurden. Verschiedene Symbole zeigten dabei unterschiedliche Wirkung, wobei nur für die wichtigsten Informationen Symbole verwendet werden sollten. Die Verwendung einer einfachen Sprache wirkte sich ebenfalls positiv auf die Auswahlentscheidung aus. *Sander und Kollegen* leiteten in ihrem Review [10] literaturgestützte Empfehlungen für das Design von Krankenhausqualitätsinformationen auf Report Cards ab, die in 12 Kategorien zusammengefasst wurden. Die Autoren empfahlen hier die Darstellung evaluativer Tabellen mit grafischen Elementen wie Sternen anstelle von alleinigen Säulen- oder Balkendiagrammen. Tabellenähnliche Darstellungen ohne grafische Elemente sollten vermieden werden. Grafische Hervorhebungen sollten erfolgen, allerdings ausschließlich in Bezug auf die jeweils wichtigen Informationen. Ebenso wurde empfohlen, dass Leistungserbringer entsprechend ihrer Performance angeordnet werden sollten. Sofern numerische Werte gezeigt werden, sollten höhere Werte gute Ergebnisse repräsentieren. Die dritte Übersichtsarbeit von *Kurtzman und Greene* [9] bestätigt die Erhöhung der Verständlichkeit durch die Verringerung der Komplexität des Designs. Eine Vereinfachung der Darstellung kann beispielsweise durch die Reduzierung der Auswahloptionen (d.h. Leistungserbringer) bzw. der Attribute je Auswahloption erreicht werden. Ebenso sollten höhere Werte bessere Ergebnisse repräsentieren sowie Fachsprache vermieden werden; evaluative Elemente (z.B. Sterne) hingegen haben einen positiven Effekt. Bezüglich des Designs wurde erstens angeführt, dass numerische Formate incl. Text (z.B. Häufigkeiten dargestellt als „1 in 250“) nicht so verständlich waren wie einfache Formate mittels grafischer Darstellungen oder bekannter Icons. Zweitens wurden grafische Darstellungen als verständlicher als numerische Formate identifiziert (bis auf Balken- und Säulendiagramme). Drittens hatten evaluative bzw. interpretierende Elemente einen positiven (z.B. Kreise, Sterne, Word Icons) Einfluss auf das Verständnis der Informationen. Allerdings wurde auch hier kein Design als überlegen eingestuft („yet, no single display was found to be superior“). Bezüglich der Darstellung von Kosten führte die Darstellung von \$-Zeichen zu negativen Ergebnissen.

## **Gesamtbetrachtung**

Nach Betrachtung der Ergebnisse der hier durchgeführten Literaturrecherche und der Ergebnisse anderer Übersichtsarbeiten scheint insbesondere die Reduzierung der Komplexität als eine der Hauptstrategien für ein verständliches Public Reporting elementar zu sein. Dies kann erreicht werden beispielsweise durch die Reduzierung der Anzahl der gezeigten Leistungserbringer bzw. der Anzeige nur ausgewählter Ergebnisdaten [6, 9, 18, 29, 41], der Verwendung von Composite/Summary Measures [26, 32, 40–42], dem Fokus auf relevante Informationen und deren Anordnung [39], der Darstellung eines Qualitätskonzepts als Grundlage für die Darstellung [34], der Anordnung der Leistungserbringer als Rangliste bzw. nach Performance [10, 32], der Verwendung einer einfachen und selbsterklärenden Sprache sowie der Vermeidung von Fachwörtern [9, 18, 34, 36, 37].

Patientenkommentare weisen einen negativen Einfluss auf die Entscheidungsqualität auf [29, 31] und sollten daher durch entsprechende Entscheidungshilfen begleitet werden [31]. Bezüglich der Darstellung von Kosteninformationen wurde abgeleitet, dass diese in Form von Kreisen bzw. Sternen bzw. als absoluter Betrag dargestellt werden sollten [28]. Die Darstellung von Kosten in \$-Zeichen führte zu negativen Ergebnissen [9].

Bezüglich des Designs kann das Fazit von Kurtzman und Greene bestätigt werden, dass sich bislang kein Design als überlegen herausgestellt hat („yet, no single display was found to be superior“) [9]. Hier finden sich teils heterogene Ergebnisse. Jedoch scheint eine weitestgehende Einigkeit zu herrschen, dass rein numerische Formate vermieden werden sollten [9, 10]. Grafische Darstellungen scheinen zudem verständlicher als rein numerische Formate zu sein, wobei Balken- und Säulendiagramme besondere Aufmerksamkeit verlangen. Evaluative bzw. interpretierende Elemente (z.B. Kreise, Sterne, Word Icons) bzw. visualisierte Wort-Symbole<sup>7</sup> [33] haben einen positiven Einfluss auf das Verständnis der Informationen [9, 39, 41], sollten aber begrenzt zur Hervorhebung der jeweils wichtigen Informationen eingesetzt werden [10]. Sofern Zahlen dargestellt werden, sollten höhere Zahlen für die Darstellung positiver Ergebnisse verwendet werden [10, 39]. Alternativ sollte ein Hinweis erfolgen, ob niedrige oder hohe Werte eine gute Qualität darstellen bzw. eine gute Qualitätsergebnisrange aufzeigen [9, 10].

---

<sup>7</sup> z.B. „Better“ hinterlegt mit einem gelben Pfeil

## 5 Verzeichnisse

### 5.1 Literaturverzeichnis

1. Aswani MS, Reagan J, Jin L, Pronovost PJ, Goeschel C. Variation in public reporting of central line-associated bloodstream infections by state. *Am J Med Qual.* 2011;26:387–95. doi:10.1177/1062860611399116.
2. Damman OC, van den Hengel, Ylva Ka, van Loon, A Jeanne M, Rademakers J. An international comparison of web-based reporting about health care quality: content analysis. *J Med Internet Res.* 2010;12:e8. doi:10.2196/jmir.1191.
3. Bridges JFP, Berger Z, Austin M, Nassery N, Sharma R, Chelladurai Y, et al. Public Reporting of Cost Measures in Health: An Environmental Scan of Current Practices and Assessment of Consumer Centeredness. Rockville (MD); 2015 Mar.
4. Christianson JB, Volmar KM, Alexander J, Scanlon DP. A report card on provider report cards: current status of the health care transparency movement. *J Gen Intern Med.* 2010;25:1235–41. doi:10.1007/s11606-010-1438-2.
5. Damberg CL, McNamara P. Postscript: research agenda to guide the next generation of public reports for consumers. *Med Care Res Rev.* 2014;71:97S-107S. doi:10.1177/1077558714535982.
6. Emmert M, Kast K, Sander U. Characteristics and decision making of hospital report card consumers: lessons from an onsite-based cross-sectional study. *Health Policy.* 2019.
7. Emmert M, Wiener M. What factors determine the intention to use hospital report cards? The perspectives of users and non-users. *Patient Educ Couns.* 2017;100:1394–401. doi:10.1016/j.pec.2017.01.021.
8. Emmert M, Meszmer N, Simon A, Sander U. Internet-based Report Cards for Hospital Choice Making in Germany: A Clinical Area-focused Perspective. *Gesundheitswesen.* 2016;78:721–34. doi:10.1055/s-0035-1549968.
9. 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.
10. Sander U, Emmert M, Dickel J, Meszmer N, Kolb B. Information presentation features and comprehensibility of hospital report cards: design analysis and online survey among users. *J Med Internet Res.* 2015;17:e68. doi:10.2196/jmir.3414.
11. Sandmeyer B, Fraser I. New Evidence on What Works in Effective Public Reporting. *Health Serv Res.* 2016;51 Suppl 2:1159–66. doi:10.1111/1475-6773.12502.
12. Bhandari N, Scanlon DP, Shi Y, Smith RA. 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.* 2018:1077558718774945. doi:10.1177/1077558718774945.
13. Emmert M, Schlesinger M. Hospital Quality Reporting in the United States: Does Report Card Design and Incorporation of Patient Narrative Comments Affect Hospital Choice? *Health Serv Res.* 2017;52:933–58. doi:10.1111/1475-6773.12519.

14. 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.
15. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic re-views and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62:1006–12. doi:10.1016/j.jclinepi.2009.06.005.
16. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred re-ported items for systematic review and meta-analysis protocols (PRISMA-P) 2015 state-ment. *Syst Rev*. 2015;4:1. doi:10.1186/2046-4053-4-1.
17. Higgins J, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interven-tions*. 5th ed.; 2011.
18. Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consum-ers use quality-of-care information? A systematic review. United States; 2009 Jan.
19. Knapp P, Gardner P, McMillan B, Raynor DK, Woolf E. Evaluating a combined (fre-quency and percentage) risk expression to communicate information on medicine side effects to patients. *Int J Pharm Pract*. 2013;21:226–32. doi:10.1111/j.2042-7174.2012.00254.x.
20. Smerecnik CMR, Mesters I, Kessels LTE, Ruiter RAC, Vries NK de, Vries H de. Under-standing the positive effects of graphical risk information on comprehension: measuring attention directed to written, tabular, and graphical risk information. *Risk Anal*. 2010;30:1387–98. doi:10.1111/j.1539-6924.2010.01435.x.
21. Schram A, Sonnemans J. How individuals choose health insurance: An experimental ana-lysis. *European Economic Review*. 2011;55:799–819. doi:10.1016/j.euroecorev.2011.01.001.
22. Greene J, Peters E. Medicaid consumers and informed decisionmaking. *Health Care Fi-nanc Rev*. 2009;30:25–40.
23. Peters E, Dieckmann NF, Vastfjall D, Mertz CK, Slovic P, Hibbard JH. Bringing meaning to numbers: the im-pact of evaluative categories on decisions. *J Exp Psychol Appl*. 2009;15:213–27. doi:10.1037/a0016978.
24. Papadimitriou C, Magasi S, Demark H, Taylor C, Wolf MS, Heinemann AW, Deutsch A. Eliciting patient and caregiver perspectives to improve the public reporting of rehabilitati-on quality measures. *Rehabil Nurs*. 2013;38:24–31. doi:10.1002/rnj.63.
25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic re-views and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6:e1000097. doi:10.1371/journal.pmed.1000097.
26. Cerully JL, Parker AM, Rybowski L, Schlesinger M, Shaller D, Grob R, et al. Improving Patients' Choice of Clinician by Including Roll-up Measures in Public Healthcare Quality Reports: an Online Experiment. *J Gen Intern Med*. 2019;34:243–9. doi:10.1007/s11606-018-4725-y.
27. Donelan K, Rogers RS, Eisenhauer A, Mort E, Agnihotri AK. Consumer comprehension of surgeon perfor-mance data for coronary bypass procedures. *Ann Thorac Surg*. 2011;91:1400-5; discussion 1405-6. doi:10.1016/j.athoracsur.2011.01.019.
28. Hibbard JH, Greene J, Sofaer S, Firminger K, Hirsh J. An experiment shows that a well-designed report on costs and quality can help consumers choose high-value health care. *Health Aff (Millwood)*. 2012;31:560–8. doi:10.1377/hlthaff.2011.1168.

29. Kanouse DE, Schlesinger M, Shaller D, Martino SC, Rybowski L. How Patient Comments Affect Consumers' Use of Physician Performance Measures. *Med Care*. 2016;54:24–31. doi:10.1097/MLR.0000000000000443.
30. Kenny P, Goodall S, Street DJ, Greene J. Choosing a Doctor: Does Presentation Format Affect the Way Consumers Use Health Care Performance Information? *Patient*. 2017;10:739–51. doi:10.1007/s40271-017-0245-9.
31. Martino SC, Grob R, Davis S, Parker AM, Finucane ML, Cerully JL, et al. Choosing Doctors Wisely: Can Assisted Choice Enhance Patients' Selection of Clinicians? *Med Care Res Rev*. 2017;1077558717743822. doi:10.1177/1077558717743822.
32. Geraedts M, Hermeling P, Cruppe W de. Communicating quality of care information to physicians: a study of eight presentation formats. *Patient Educ Couns*. 2012;87:375–82. doi:10.1016/j.pec.2011.11.005.
33. Greene J, Sacks RM. Presenting Cost and Efficiency Measures That Support Consumers to Make High-Value Health Care Choices. *Health Serv Res*. 2018;53 Suppl 1:2662–81. doi:10.1111/1475-6773.12839.
34. Hibbard JH, Greene J, Daniel D. What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. *Med Care Res Rev*. 2010;67:275–93. doi:10.1177/1077558709356300.
35. Mazor KM, Dodd KS, Kunches L. Communicating hospital infection data to the public: a study of consumer responses and preferences. *Am J Med Qual*. 2009;24:108–15. doi:10.1177/1062860608330827.
36. Sander U, Kolb B, Christoph C, Emmert M. Text Comprehensibility of Hospital Report Cards. *Gesundheitswesen*. 2016;78:828–34. doi:10.1055/s-0034-1396848.
37. Sander U, Kolb B, Taheri F, Patzelt C, Emmert M. Do laymen understand information about hospital quality? An empirical verification using risk-adjusted mortality rates as an example. *Z Evid Fortbild Qual Gesundheitswes*. 2017;127-128:21–9. doi:10.1016/j.zefq.2017.09.010.
38. Zwijnenberg NC, Hendriks M, Damman OC, Bloemendal E, Wendel S, Jong JD de, Rademakers J. Understanding and using comparative healthcare information; the effect of the amount of information and consumer characteristics and skills. *BMC Med Inform De-cis Mak*. 2012;12:101. doi:10.1186/1472-6947-12-101.
39. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. *Med Care Res Rev*. 2007;64:169–90. doi:10.1177/10775587070640020301.
40. Damman OC, Hendriks M, Rademakers J, Spreeuwenberg P, Delnoij DMJ, Groenewegen PP. Consumers' interpretation and use of comparative information on the quality of health care: the effect of presentation approaches. *Health Expect*. 2012;15:197–211. doi:10.1111/j.1369-7625.2011.00671.x.
41. Damman OC, Jong A de, Hibbard JH, Timmermans DRM. Making comparative performance information more comprehensible: an experimental evaluation of the impact of formats on consumer understanding. *BMJ Qual Saf*. 2016;25:860–9. doi:10.1136/bmjqs-2015-004120.
42. Werner RM, Konetzka RT, Polsky D. Changes in Consumer Demand Following Public Reporting of Summary Quality Ratings: An Evaluation in Nursing Homes. *Health Serv Res*. 2016;51 Suppl 2:1291–309. doi:10.1111/1475-6773.12459

## 5.2 Abbildungsverzeichnis

Abbildung 1: Darstellung der Literaturrecherche mittels PRISMA Flow Diagram [25] .....	7
Abbildung 2: Sample screenshot from the SelectMD experimental website .....	8
Abbildung 3: Presentation formats a-d .....	10
Abbildung 4: Examples Of Cost Data Presented To Respondents With No, A Weak Or Strong Quality Signal ....	11
Abbildung 5: Respondents Selecting High-Value Providers, By Type Of Cost Data And Quality Signal .....	12
Abbildung 6: Presentation formats used for practice quality attributes .....	13
Abbildung 7: Overview page, SelectMD, Conventional Comments Arm. Figure A3. ....	15
Abbildung 8: Experimental Report Card Design on Weisse Liste .....	16
Abbildung 9: Screenshot of Hospital Report Card.....	18
Abbildung 10: Selection of the Quantitatively Dominant Hospital and Quantitatively Dominated Hospital.....	19
Abbildung 11: Format 1-8 .....	20
Abbildung 12: An Example of the Display from the Cost Presentation Experiment, with the Alternate Cost Displays below .....	21
Abbildung 13: An Example of the Display from the Hospital Readmission Presentation Experiment, with the Alternate Readmission Displays below .....	23
Abbildung 14: The header from each version of the quality doctor report .....	24
Abbildung 15: First Page from a Word report with inconsistent indicators.....	25
Abbildung 16: Online-Darstellung der RAM für fünf Krankenhäuser (Variante 1).....	27
Abbildung 17: Aspects of comparative healthcare information shown to each subgroup .....	29
Abbildung 18: Studie 1: Unordered—Both Quality and Nonquality Hospital Information Is Presented Unordered	30
Abbildung 19: Studie 1. Choice of the Highest-Quality Hospital by Presentation Format and Numeracy .....	31
Abbildung 20: Studie 2: Death Rate Only Is Made Easier to Evaluate with Black and White Symbols .....	31
Abbildung 21: Studie 2. Choice of the Lowest-Death-Rate Hospitals by Presentation Format and Numeracy .....	32
Abbildung 22: Studie 3: Patient-to-Nurse Ratio Is Presented as “Higher Is Better” with Easier-to-Evaluate Symbols .....	32
Abbildung 23: Studie 3. Choice of the Best Hospitals by Presentation Format and Numeracy .....	33
Abbildung 24: Example of experimental format: a combination of bar chart and star ratings, a rank ordering of providers, stars based on relative performance, three stars, and no inclusion of an overall rating .....	35
Abbildung 25: Experimental Formats .....	37
Abbildung 26: Percentages of respondents correctly selecting the top three best-performing nursing homes from a set of 20 nursing homes, for the different between-subjects manipulations of presentation approaches .....	38
Abbildung 27: Percentages of respondents correctly selecting the top three of best-performing nursing homes from a set of five nursing homes, for the different between-subjects manipulations of presentation approaches..	38
Abbildung 28: Example of the Star Ratings Available for Nursing Homes on the Nursing Home Compare Website .....	40
Abbildung 29: Features of the presentation of information on health care report cards from previous studies.....	46

## 5.3 Tabellenverzeichnis

Tabelle 1: Finale Suchsyntax für die Datenbanksuche auf PubMed und The Cochrane Library .....	5
Tabelle 2: Ein- und Ausschlusskriterien .....	6
Tabelle 3: Effects of Quality Data Presentation on Decision Quality .....	9
Tabelle 4: Surgery Results: Surgeon Preference for Display Type.....	10
Tabelle 5: Engagement, Perceived Usefulness of Comparative Performance Information, and Quality of Choices When Presented With and Without Patient Comments.....	13

<i>Tabelle 6: Logit models for evaluating presentation styles: choosing the worst alternative (model 1) and finding the choice information difficult to understand (model 2)</i> .....	14
Tabelle 7: Logistic regression models predicting likelihood of selection of the best performing hospital based on the complexity of decision making and different design types (N=1270) .....	17
Tabelle 8: Logistic regression models predicting likelihood of selection of the quantitatively dominant hospital and quantitatively dominated hospital based on different design types (N= 2700) .....	19
<i>Tabelle 9: Physicians' assessment, acceptance and comprehension of eight presentation formats</i> .....	21
Tabelle 10: Cost Presentation Experiment: Choice of Hospital by Cost Presentation Approach .....	22
Tabelle 11: Readmissions Presentation Experiment: Choice of Hospital by Readmission Presentation Approach .....	23
Tabelle 12: Comprehension Indices and Usefulness by Experimental Group .....	24
Tabelle 13: Ergebnisse zur Verständlichkeit der Texte von Qualitätsvergleichen .....	26
Tabelle 14: Zusammenhang zwischen den Ergebnissen des Instruments zur Bestimmung des Konzeptverstehens und der Auswahl eines Krankenhauses mit der niedrigsten RAM .....	28
Tabelle 15: Hierarchical regression models with regression coefficients (Beta) predicting the outcome measures [N = 349] .....	30
Tabelle 16: Choice of the hospital with the lowest risk-adjusted mortality rate (RAMR) .....	34
Tabelle 17: Ergebnisse zur Verständlichkeit in Abhängigkeit bestimmter Designelemente .....	36
Tabelle 18: Main findings from the analyses of variance testing between-subjects manipulations of presentation formats on the different outcome variables (realistic version of 20 providers) .....	39
Tabelle 19: The Log-Odds of Being Admitted to a Nursing Home with 2, 3, 4, and 5 Stars before (Coefficients on the Uninteracted Terms) and after the Star-Based Ratings Were Released in 2008 (Coefficients on the Interaction Terms), All Compared to Being Admitted to a 1-Star Facility .....	41
Tabelle 20: Empfehlungen für ein verständliches Public Reporting - Individualebene .....	42
Tabelle 21: Empfehlungen für ein verständliches Public Reporting – Krankenhausebene .....	44
Tabelle 22: Empfehlungen für ein verständliches Public Reporting - Pflegeheimebene .....	44
Tabelle 23: Effects of Presentation Formats for Quality-of-Care Information on Consumers .....	45



#### 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)