

# Quantifizierung des Verbesserungspotentials bei medizinischen Leistungserbringern

**Johannes Rauh, IQTIG**

**Biometrisches Kolloquium**

**19. November 2024**

# Agenda

---

1. Das IQTIG
2. Eignungskriterien von Qualitätsindikatoren
3. Operationalisierung des Verbesserungspotentials
4. Herausforderungen
5. Ergebnisse und Erfahrungen
6. Zusammenfassung

# Das IQTIG

## Das IQTIG

---



© EVP II Parkside Berlin S.à.r.l

- Das IQTIG wurde 2015 als unabhängiges wissenschaftliches Institut für die gesetzlich verankerte Qualitätssicherung (QS) im Gesundheitswesen gegründet (§136ff SGB V).
- Aufgaben:
  - wissenschaftliche Beratung des G-BAs und des Bundesministeriums für Gesundheit
    - G-BA = Gemeinsamer Bundesausschuss: Gremium der Selbstverwaltung im deutschen Gesundheitswesen.
  - Entwicklung, Umsetzung und Evaluation von Qualitätssicherungsmaßnahmen
  - Darstellung der Versorgungsqualität

# Externe Qualitätssicherung im Gesundheitswesen

---

- Das IQTIG organisiert Qualitätssicherungsverfahren in verschiedenen medizinischen Fachgebieten, u.a.:
  - Transplantationsmedizin
  - Geburtshilfe
  - Herzschrittmacher
  - Orthopädie und Unfallchirurgie (z.B. Hüft- und Knieprothesen)
- Alle Leistungserbringer (*LE*, derzeit vorwiegend Krankenhäuser) sind gesetzlich verpflichtet, zu jedem „relevanten“ Fall Daten zu liefern (*QS-Daten*).
- Weitere Datenquellen:
  - *Sozialdaten* bei den Krankenkassen (Abrechnungsdaten)
  - *Patientenbefragungen*
- Das IQTIG legt *Qualitätsindikatoren* (QIs) fest, die auf diesen Daten berechnet werden.

## Qualitätsindikatoren

---

- Qualitätsindikatoren messen den Grad der Erfüllung von Qualitätsanforderungen.
- Ein Qualitätsindikator (QI) beinhaltet in der Regel:
  - eine Beschreibung der *Grundgesamtheit* an Fällen, und
  - eine Beschreibung von *unerwünschten Ereignissen*.
- Für jeden LE werden einmal im Jahr die Anzahl an unerwünschten Ereignissen  $o$  und die Größe der Grundgesamtheit  $n$  bestimmt.
- Hängt die Wahrscheinlichkeit für unerwünschte Ereignisse von Patienteneigenschaften ab, wird zudem eine erwartete Anzahl an unerwünschten Ereignissen  $e$  geschätzt („Risikoadjustierung“).
- Das Verhältnis  $o/n$  bzw.  $o/e$  (SMR) ist ein **Maß für die Kompetenz** (bzgl. der Qualitätsanforderung)
- Durch Vergleich mit einem **Referenzwert** wird über Anschlusshandlungen (Stellungnahmeverfahren durch Fachkommissionen) entschieden.

## Qualitätsindikatoren – Beispiele

---

### 1. Präoperative Verweildauer bei Hüftfrakturen (QI 54003)

- **Grundgesamtheit:** Patienten mit Hüftfraktur
- **unerwünscht:** Wartezeit > 48 Stunden bis zur OP
- **Referenzwert:** 15 %

### 2. Anwesenheit des Pädiateers bei Frühgeburten (QI 318)

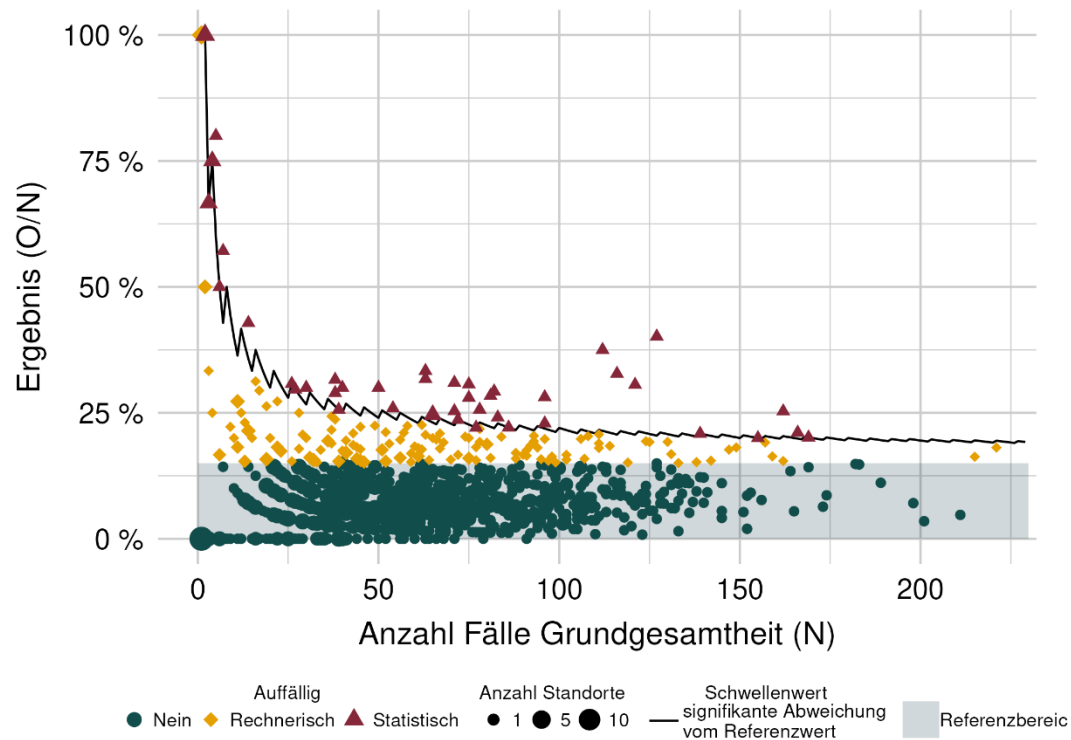
- **Grundgesamtheit:** Frühgeburten (SSW 24+0 bis unter 35+0)
- **unerwünscht:** kein Pädiateer anwesend während der Geburt
- **Referenzwert:** 10 %.

### 3. Sterblichkeit im Krankenhaus bei Pneumonie (QI 50778)

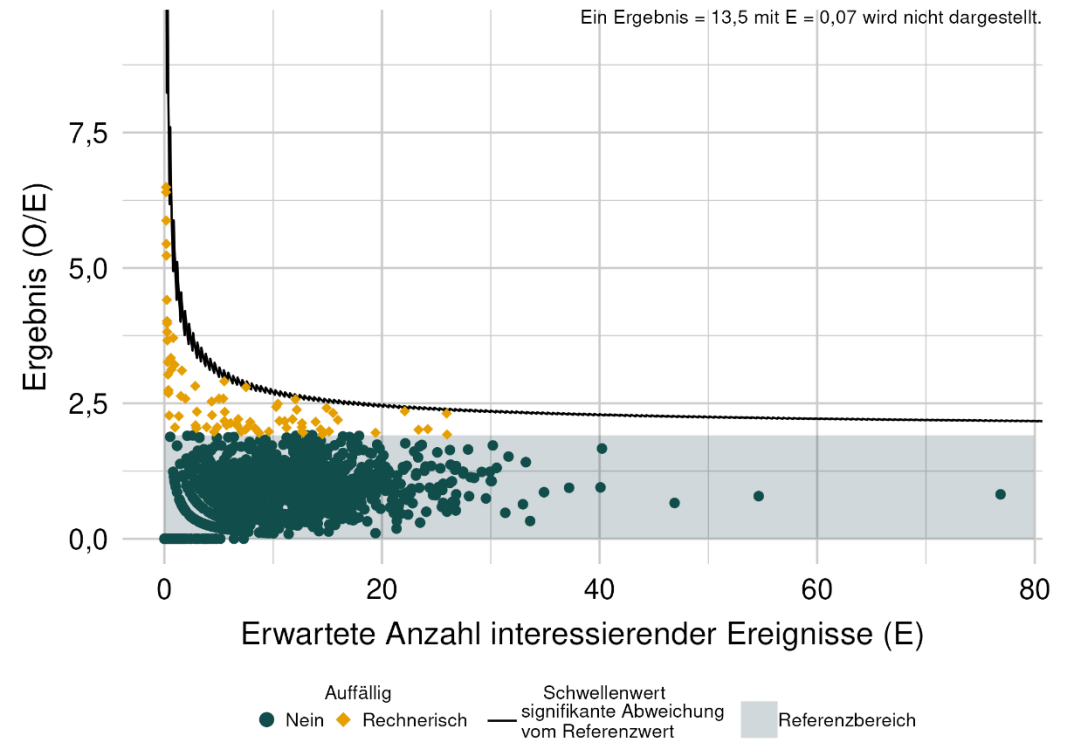
- **Grundgesamtheit:** Patienten mit Lungenentzündung
- **unerwünscht:** Tod im Krankenhaus
- **Risikoadjustierungsmodell:** 8 Risikofaktoren (ca. 14 df, GAM mit Splines)
- **Referenzwert:** 1,69 (auf Ebene von  $o/e$ )

# Beispielergebnisse

## Präoperative Verweildauer bei Hüftfrakturen



## Sterblichkeit im Krankenhaus bei Pneumonie





# Eignungskriterien von Qualitätsindikatoren

## Die Eignungskriterien des IQTIG

---

- Aufbauend auf der Literatur und den Erfahrungen der externen Qualitätssicherung hat das IQTIG Eignungskriterien für Qualitätsindikatoren entwickelt.
- Sie beschreiben die Eignung
  - für Aussagen zur Qualität sowie
  - als Grundlage für qualitätsbezogene Entscheidungen.
- Die Eignung wird dabei sowohl bei der Entwicklung als auch im Regelbetrieb überprüft.
- Seit 2022 findet eine grundlegende Überprüfung des Aufwand-Nutzen-Verhältnisses der aktuellen Verfahren statt (Abschluss Anfang 2025).
  - Ziel war auch die Abschaffung von nicht ausreichend geeigneten Indikatoren.
  - Im Rahmen dieser Überprüfung wurde die Operationalisierung der Eignungskriterien grundlegend überarbeitet.

# Die Eignungskriterien des IQTIG für QI

Eignung		
des Qualitätsmerkmals	der Operationalisierung	des Bewertungskonzepts
Bedeutung für die Patientinnen und Patienten	Objektivität der Messung	Angemessenheit des Referenzbereichs
Zusammenhang mit pat.relevantem Merkmal	Reliabilität der Messung	Klassifikationsgüte
Potenzial zur Verbesserung	Datenqualität	
Beeinflussbarkeit	Validität der Messung	
Brauchbarkeit für den Handlungsanschluss	Angemessenheit der Risikoadjustierung	
	Praktikabilität der Messung	

## Kriterium *Potenzial zur Verbesserung*

---

- Ziel: Umfang des potenziellen Nutzens der *externen Qualitätssicherung dieses Merkmals* benennen
- Bestehen *gegenwärtig Qualitätsunterschiede* bzw. **-defizite**?
  - || ■ **Wie viele Patientenergebnisse** sind verbesserbar bzw. vermeidbar?
    - **Wie viele LE** können mehr erreichen?
  - ( ■ Wie viele LE wurden im Stellungnahmeverfahren als auffällig beurteilt? )
- Besteht das Risiko einer *zukünftigen* Verschlechterung der Versorgung?

# Operationalisierung des Verbesserungspotentials

## Grundidee beim Verbesserungspotential

---

1. Identifiziere in den Daten „gute“ LE, die als Referenz dienen sollen.
2. Schätze die Anzahl an unerwünschten Ereignissen, wenn alle LE so gut gewesen wären, wie die Referenz-LE.
3. Die Differenz der tatsächlich beobachteten Anzahl an unerwünschten Ereignissen mit dieser geschätzten Anzahl bildet die *Anzahl an verbesserbaren Ereignissen*.

### Bemerkungen:

- **Herausforderung:** Berücksichtigung von Unsicherheit bei den beobachteten Raten an unerwünschten Ereignissen (Stichwort „Regression to the mean“)
- **Limitation:** Die Idee beruht auf der Detektion von Heterogenität. Sie funktioniert nicht, wenn alle LE „gleich schlecht“ sind.
- **grundlegende Annahme:** Eine datengestützte Schätzung des Verbesserungspotentials ist nur möglich, wenn die Operationalisierung des Indikators geeignet ist (insbesondere angemessene Risikoadjustierung).

# Statistische Modellierung von Qualitätsindikatoren

---

- Standardmodell ohne Risikoadjustierung: Binomialmodell
  - Anzahl unerwünschter Ereignisse  $O_i \sim \text{Bin}(n_i, \pi_i)$ , mit **Kompetenzparameter**  $\pi_i$  des LE  $i$ .
- Mit Risikoadjustierung hängt die Wahrscheinlichkeit  $\pi_{i,j}$  für ein unerwünschtes Ereignis vom Fall  $j$  und LE  $i$  ab.
  - Standardmodell:  $\text{logit}(\pi_{i,j}) = \beta_0 + x_{i,j}^t \beta + u_i$ , mit
    - $x_{i,j}$  - relevante Falleigenschaften („Risikofaktoren“)
    - $\beta$  - Einfluss der Risikofaktoren („Risikoadjustierungsmodell“)
    - $\beta_0$  - Intercept
    - $u_i$  - Einfluss des LE = **Kompetenzparameter**.
- Bei beiden Modellen stellt sich die Frage nach der Schätzung von Modell/Kompetenzparametern:
  - Feste Effekte? Pönalisiert (z.B. Firth-Korrektur)?
  - Zufällige Effekte?
  - Bayesianische Schätzung? Empirical Bayes oder nicht informative Priori?

## Die ABC™-Methodik (*Achievable Benchmarks of Care*, Weissman et al. 1999)

---

### Für Indikatoren ohne Risikoadjustierung

1. Berechne für jeden LE ein geshrinktes Ergebnis APF (*Adjusted Performance Fraction*, entsprechend einer Beta(1,1)-Priori):  $APF = \frac{o_i+1}{n_i+2}$
2. Erstelle anhand des APF ein Ranking der LE.
3. Schließe so viele der am besten gerankten LE in die Referenz ein, bis ein festgelegter Prozentsatz behandelter Patient:innen (z.B. 10 %) erreicht ist.
4. Berechne für diese Patient:innen anhand der **beobachteten** Daten ein gepooltes Ergebnis (erreichbarer Indikatorwert):

$$ABC = \frac{\sum_{i \text{ Referenz-LE}} o_i}{\sum_{i \text{ Referenz-LE}} n_i}$$

5. Die **Anzahl an verbesserbaren Ereignissen** ist dann  $(ABC \cdot \sum_i n_i) - \sum_i o_i$



## Eigenschaften der ABC<sup>TM</sup>-Methodik

---

Aus Weissman et al. 1999:

*Three characteristics of the ABC<sup>TM</sup> that we deem essential are:*

- *(1) benchmarks represent a measurable level of excellence;*
- **(2) benchmarks are demonstrably attainable;**
- *(3) benchmarks are derived from data in an objective, reproducible and predetermined fashion.*

*From these characteristics it follows that*

- *(4) providers with high performance are selected to define a level of excellence in a predetermined fashion, but*
- **(5) providers with high performance on small numbers of cases do not influence unduly benchmark levels.**

## Beobachtungen zur ABC-Methodik

---

- Die ABC-Methodik versucht Erreichbarkeit zu belegen, indem Referenz-LE explizit angegeben werden.
- Die ABC-Methodik berücksichtigt keine Risikoadjustierung.
- Die ABC-Methodik bedient sich zweier unterschiedlicher Schätzungen für die Behandlungsqualität:
  - Für das Ranking (Identifikation der Referenz-LE) wird geshrinkt.
  - Für die (kontrafaktische) Prädiktion wird nicht geshrinkt.
- Zitat aus Weissman et al. 1999:

*Finally, the APF is used solely for the purpose of ranking providers for ABC™ calculation. The Bayesian numbers are not used in the further calculations of the ABC™. In no event should these numbers be used to compare performance of individual providers.*

## Unsere Methodik

---

- Schätze das Standardmodell:  $\text{logit}(\pi_{i,j}) = \beta_0 + x_{i,j}^t \beta + u_i$ , mit
  - $x_{i,j}$  - relevante Falleigenschaften („Risikofaktoren“)
  - $\beta$  - Einfluss der Risikofaktoren („Risikoadjustierungsmodell“)
  - $\beta_0$  - Intercept
  - $u_i$  - Einfluss des LE, modelliert als **zufällige Effekte**:  $u_i \sim N(0, \tau)$
  - Schätzung mittels REML
- Definiere als Referenz-LE einen LE mit Effekt
$$u_* := \text{Median}\{\hat{u}_i : \hat{u}_i < 0\}$$
- Berechne durch Einsetzen von  $u_*$  die Anzahl an unerwünschten Ereignissen, wenn alle LE so gut gewesen wären, wie der Referenz-LE.

## Beobachtungen zu unserer Methodik

---

- Es gibt nur einen (hypothetischen) Referenz-LE.
- Für die Identifikation des Referenz-LE wird dasselbe Modell verwendet wie für die Berechnung der Anzahl an (un)vermeidbaren unerwünschten Ereignissen
  - Beide Punkte ergeben sich natürlich aus der Berücksichtigung der Risikoadjustierung.
- Bei der ABC<sup>TM</sup>-Methodik enthalten die Referenz-LE einen festen Anteil an Fällen. Unsere Methodik berücksichtigt die Fallzahlen nicht explizit.
- Die ABC<sup>TM</sup>-Methodik verwendet zur Schrumpfung eine nicht-informative B(1,1)-Priori. Unser Ansatz entspricht eher einem empirischen Bayes-Ansatz.
- Oft werden zufällige Effekten dazu verwendet, für Heterogenität zu adjustieren. Hier ist hingegen die Verteilung der zufälligen Effekte selbst von Interesse.

# Herausforderungen

## Herausforderungen

---

- Quantifizierung von Unsicherheit (Konfidenzintervalle)
  - Weniger wichtig beim Benchmarking für einzelne LE?
  - Wichtig bei der Bewertung der Eignung von Indikatoren!
- Schätzung des Verbesserungspotentials bei kleinen Fallzahlen
- Erweiterung auf Indikatoren mit nicht-binären Outcomes; z.B.:
  - Patientenbefragung
  - Strahlenbelastung (Dosisflächenprodukt in  $\text{cGy} \times \text{cm}^2$ )
- Zeitliche Abhängigkeit des geschätzten Verbesserungspotentials
- Umgang mit Indizes (fehlende Additivität)
- Robustheit gegenüber Verletzungen von Modellannahmen

Betrifft analog auch  
die ABC™-Methodik

## Herausforderung: Umgang mit Indizes

---

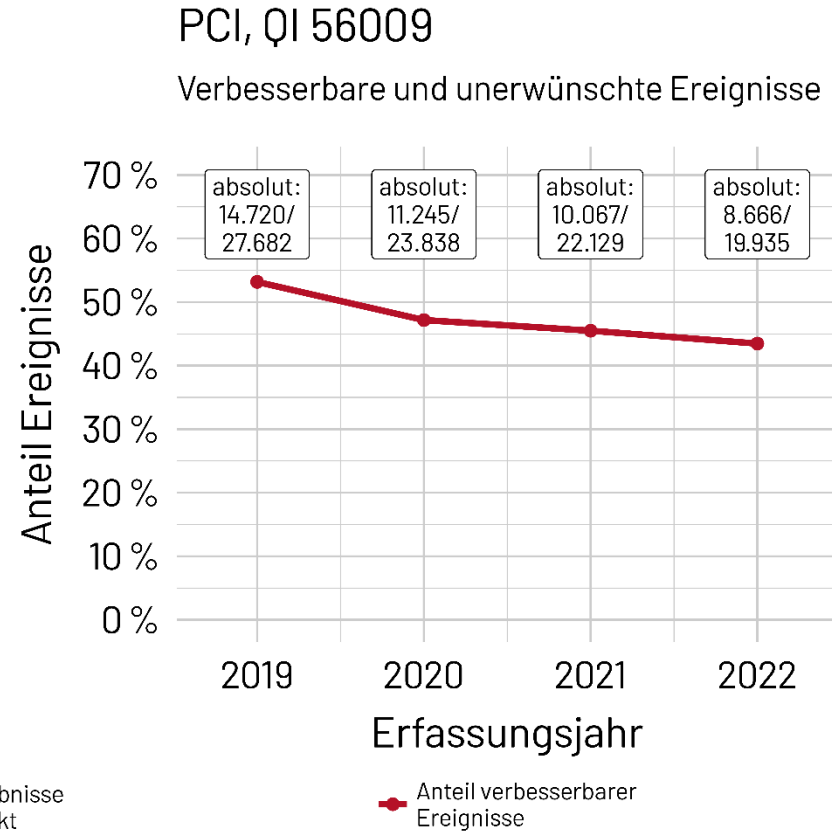
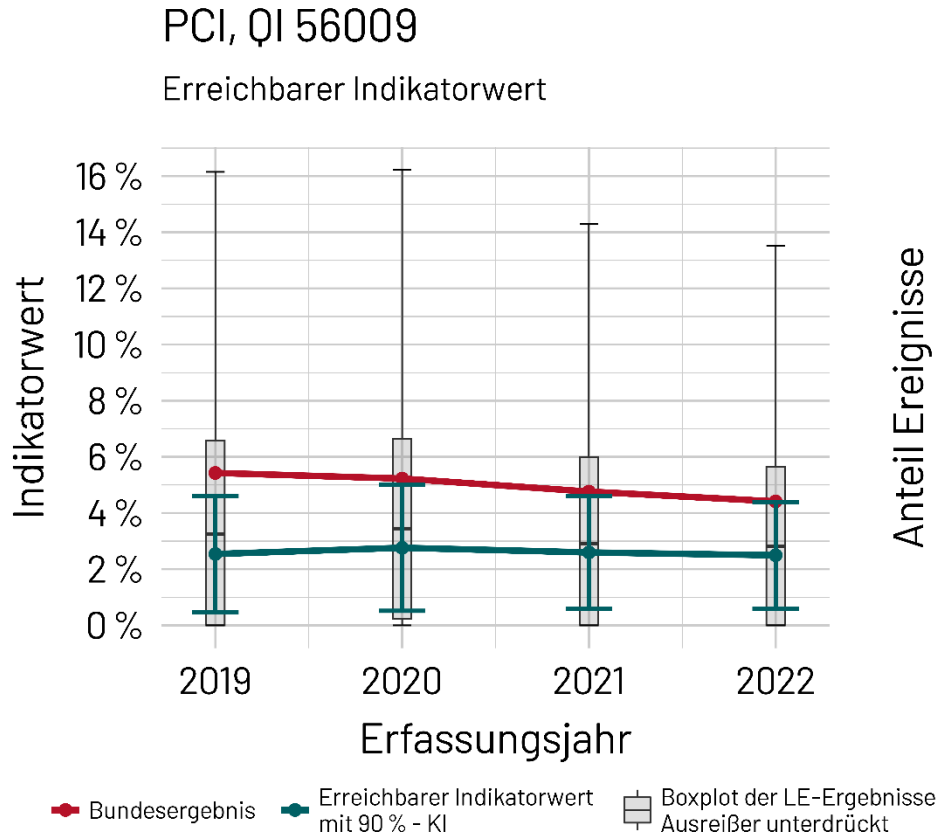
**Behauptung:** Datengetriebene Methodiken fürs Verbesserungspotential können nicht additiv sein, in folgendem Sinne:

- Bei kombinierten Endpunkten (z.B. „Tod oder Komplikation“) gilt:  
geschätzte Anzahl vermeidbarer Tode + geschätzte Anzahl vermeidbarer Komplikationen >  
geschätzte Anzahl an vermeidbaren unerwünschten Ereignissen insgesamt
- Bei Risikoadjustierung (z.B. nach Geschlecht) gilt analog:  
geschätzte Anzahl vermeidbarer unerwünschter Ereignisse bei Frauen  
+ geschätzte Anzahl vermeidbarer unerwünschter Ereignisse bei Männern >  
geschätzte Anzahl an vermeidbaren unerwünschten Ereignissen insgesamt
- **Statistisch:** Die Annahme eines einzelnen Kompetenzparameters ist hier ggf. zu stark.
- **Inhaltlich** muss man entscheiden, was man zusammen betrachten will – und was man als erreichbar fordern will.

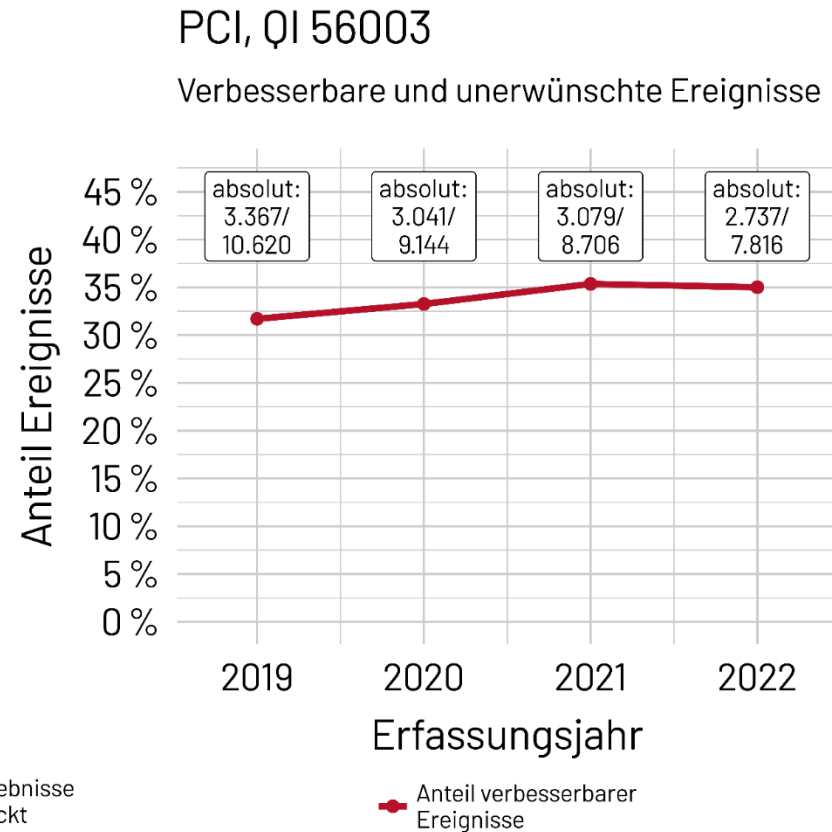
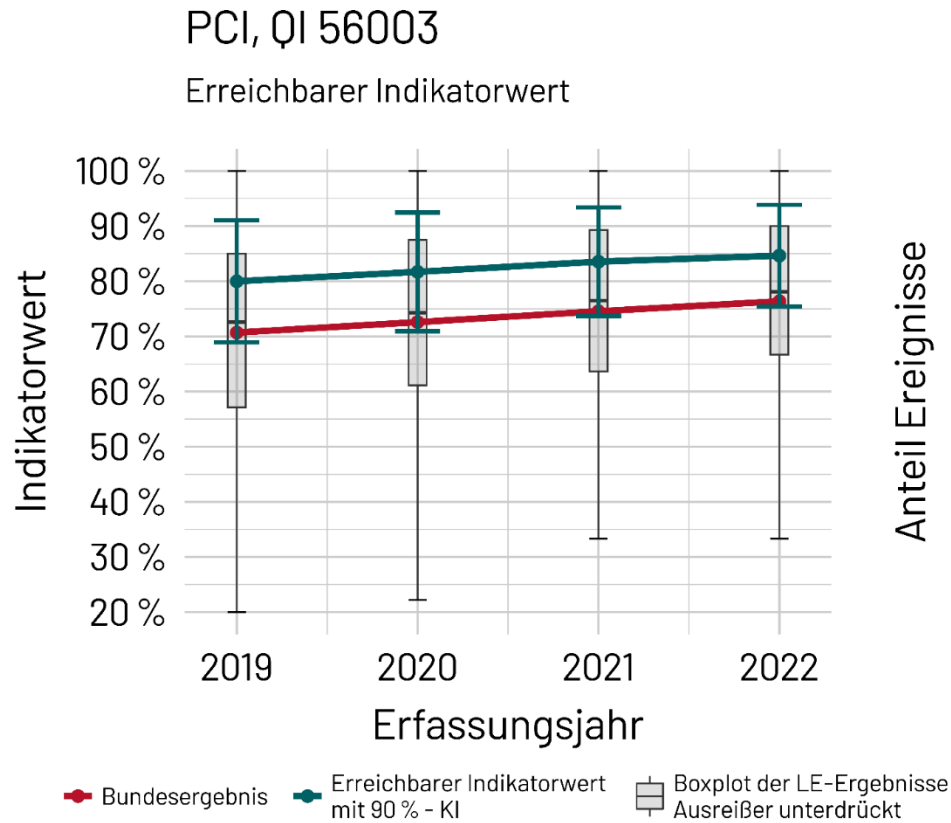
# Ergebnisse und Erfahrungen



## Beispiel: Isolierte Koronarangiographien mit Kontrastmittelmenge über 150 ml



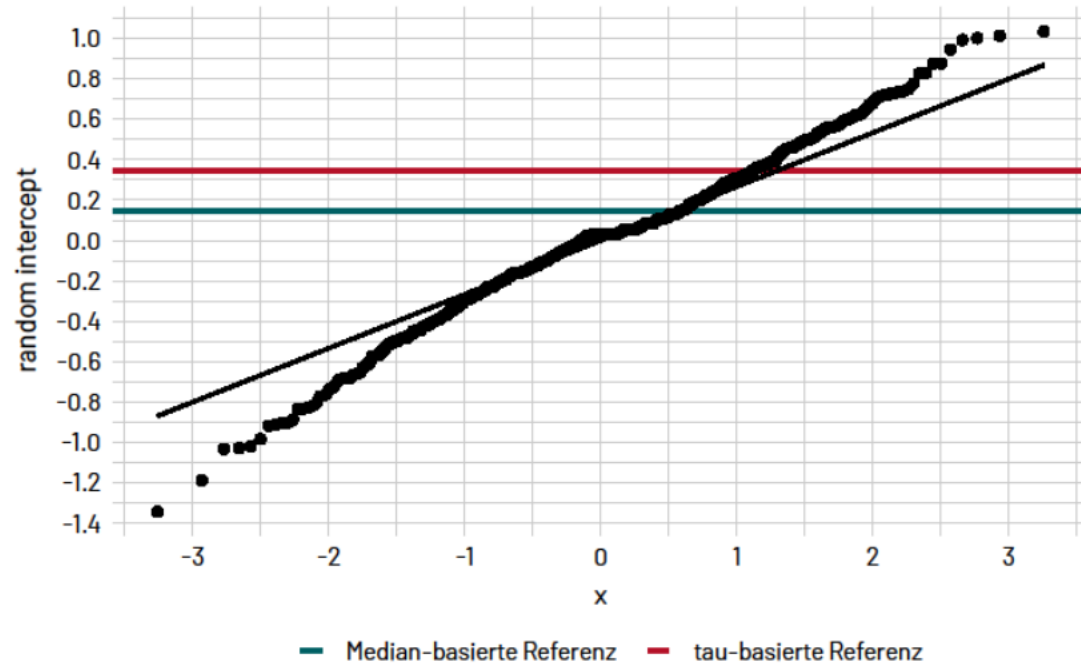
## Beispiel: „Door-to-balloon“-Zeit bis 60 Minuten bei Erst-PCI mit der Indikation ST-Hebungsinfarkt



# Geschätzte Kompetenzparameter

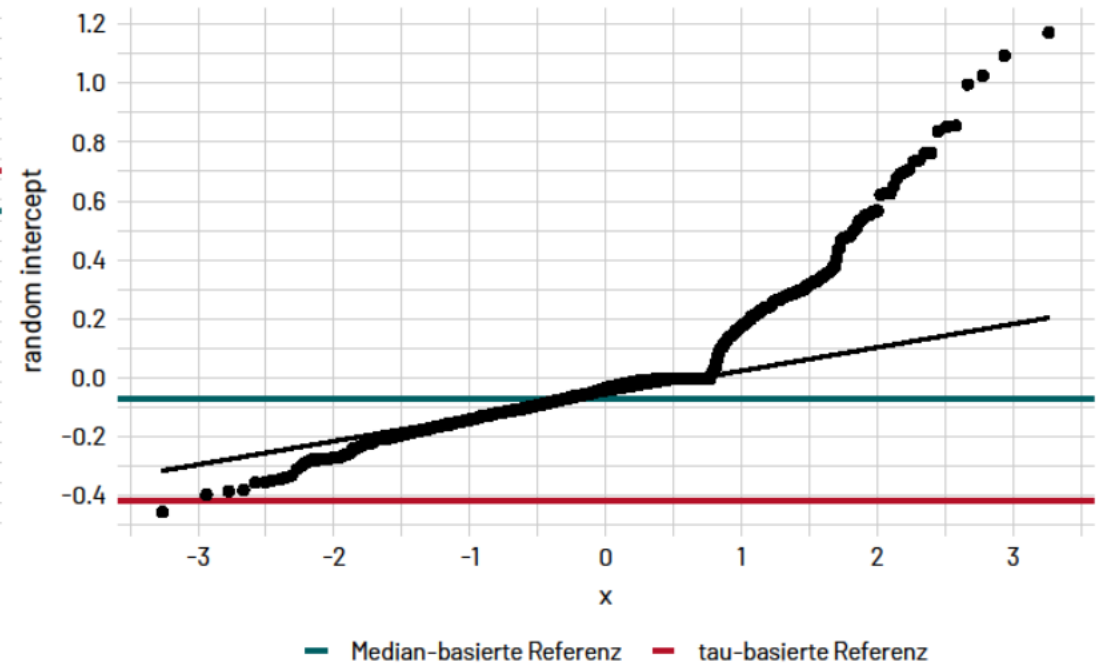
Beispiel 1 (positiv = gut)

tau = 0,52; Median-basierte Referenz: 0,15; tau-basierte Referenz: 0,35



Beispiel 2 (negativ = gut)

tau = 0,62; Median-basierte Referenz: -0,07; tau-basierte Referenz: -0,42



## Erfahrungen

---

- Die geschätzten Zahlen wurden im Großen fachlich nicht infrage gestellt.
- Kritikpunkte an der Methodik:
  1. Warum werden die Referenzwerte der Indikatoren nicht berücksichtigt?
    - Die Referenzwerte sind Vergleichswerte für Indikatorergebnisse für die Entscheidung über Qualitätsförderungsmaßnahmen bei LE.
    - In der Praxis entsprechen sie eher Minimalanforderungen.
    - Zudem sind sie uneinheitlich festgelegt.
    - Die Referenzwerte eignen sich nur bedingt dazu, den erreichbaren Indikatorwert normativ festzulegen.
  2. Warum wurden keine Zeitreihen verwendet, um Deckeneffekte zu detektieren?
    - Auch bei Indikatoren, bei denen sich die Ergebnisse nicht mehr verbessern, kann es noch ein Verbesserungspotential geben. In diesem Fall müsste allerdings der Druck zu Qualitätsverbesserungen erhöht werden.

# Zusammenfassung

## Zusammenfassung

---

- Die Schätzung des Verbesserungspotentials ist wichtig für die Bewertung der Eignung von Qualitätsindikatoren im Kontext von Kosten-Nutzen-Betrachtungen.
- Wir haben hierfür eine eigene Methodik entwickelt, die sich an der ABC™-Methodik orientiert.
  - Die Methodik ist datengetrieben und beruht auf der Schätzung von Heterogenität.
- Statistische Herausforderungen und offene Fragen:
  - Umgang mit Indizes
  - Interpretation der Zeitabhängigkeit
  - Berücksichtigung von Setzungen und fachlicher Erwartung?
  - Umgang mit kleinen Fallzahlen
  - Sensitivität bei Verletzungen von Modellannahmen

## Referenzen

---

- Weissmann et al. Achievable benchmarks of care: the ABC™s of benchmarking. *Journal of Evaluation in Clinical Practice*, 5, 3, 269-281 (1999).
- Zu den Eignungskriterien und ihrer Operationalisierung:
  - [Methodische Grundlagen](#) V2.1 (Entwurf), 2024
  - Empfehlungen zur Weiterentwicklung von Verfahren der datengestützten gesetzlichen Qualitätssicherung: [Abschlussbericht für die ersten drei Verfahren](#), [Abschlussbericht für die nächsten sechs Verfahren](#) (dritter Abschlussbericht erscheint 2025).

**Vielen Dank für die Aufmerksamkeit!**

**Werbung:** DAGStat-Session „Bayesianische Methoden im regulativen Kontext“