

**IQTIG**

Institut für  
Qualitätssicherung  
und Transparenz im  
Gesundheitswesen

# **Risikoadjustierung: Vorgehen des IQTIG bei der Entwicklung**

## **IQTIG-Methodensymposium**

**11. Mai 2026**

**Dr. Maria Golde, Stefan Gehrig**

## Angabe zu Interessenkonflikten

---

*Zum folgenden Vortrag bestehen keine Interessenkonflikte.*

## Risikoadjustierung: Vorgehen des IQTIG

---

1. Identifikation von Einflussfaktoren
2. Einschätzung der Verantwortlichkeit
3. Einschätzung der Relevanz
4. Einschätzung der Operationalisierbarkeit
5. Modellentwicklung
6. Gesamtbewertung der Risikoadjustierung

# 1. Identifikation von Einflussfaktoren

---

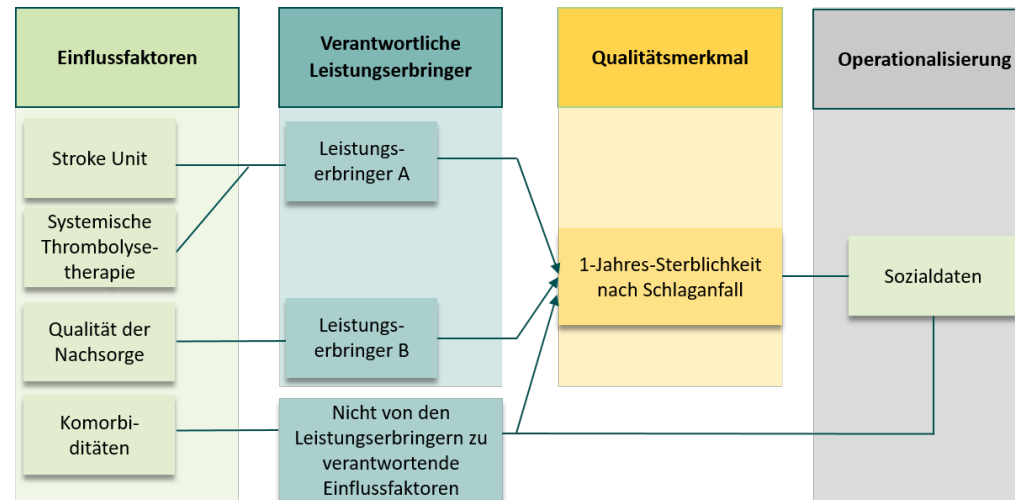
## Welche Faktoren haben Einfluss auf das untersuchte Ergebnis / den untersuchten Prozess?

- Literaturrecherche
  - Populationsbeschreibung bei Baseline in RCTs („Table 1“)
  - Prognosescores/Faktoren in Prognosemodellen in Kohortenstudien
  - Leitlinien
- Nennung durch Fachexpert:innen
- Sammlung, ggf. Systematisierung (Pufulete et al., 2022; IQWiG 2025)

## 2. Einschätzung der Verantwortlichkeit

### Liegt der Einflussfaktor in der Verantwortung der betreffenden Leistungserbringer?

- Beratung durch Fachexpert:innen
- Wenn nein: Berücksichtigung als potenzieller Faktor in der Risikoadjustierung



## 3. Einschätzung der Relevanz

---

- Prävalenz des Einflussfaktors
- Stärke des Effekts
- Konsistenz des Effekts

## 4. Einschätzung der Operationalisierbarkeit

---

- Vollständigkeit der Informationen in der Primärdokumentation / den Sozialdaten
- Objektivität/Reliabilität der Erhebung in der Versorgung
- Datenqualität

## 5. Modellentwicklung: Hintergrund

- Wie werden die identifizierten Einflussfaktoren für das QI-Ergebnis berücksichtigt?
- Theoretische Definition risikoadjustierter Indikator (für Leistungserbringer **LE1**):

$$\frac{\text{Mittleres Behandlungsergebnis für die Population des LE1 unter Behandlung bei LE1}}{\text{Mittleres Behandlungsergebnis für die Population des LE1 unter Behandlung beim Referenz-LE}} = \frac{\text{„O“}}{\text{„E“}}$$

Die **Population des LE1** ist anhand der Ausprägungen aller Einflussfaktoren definiert (auch ungemessene)

Als Referenz-LE dient i.d.R. die „bundesdurchschnittliche“ Behandlung im Vorjahr

## 5. Modellentwicklung: Hintergrund

---

- Gibt es viele Einflussfaktoren mit vielen Ausprägungen (z.B. Alter 18, 19, 20, ...), bietet sich zur Schätzung von „E“ ein **statistisches Modell** an
- Oft eine multiple logistische Regression mit Schätzung auf Bundesdatensatz des Vorjahres:

$$\text{Logit}(E(Y|X_1, X_2, \dots, X_M)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_M X_M$$

- Das Modell ordnet den für die Population des LE1 gemessenen Faktoren  $X_1, X_2, \dots, X_M$  ein erwartetes Ergebnis beim Referenz-LE zu:

$$X_1, X_2, \dots, X_M \longrightarrow E(Y|X_1, X_2, \dots, X_M)$$

Technische Anmerkung: Hier sieht man auch, weshalb ungemessene Einflussfaktoren Verzerrungspotenzial bergen:  $E(Y|X_1, X_2, \dots, X_M)$  entspricht dann nicht dem Nenner der Vorfolie.

## 5. Modellentwicklung: Hintergrund

---

- **Zielparameter** der Risikoadjustierung:  $O/E$
- **Schätzmethode** per Outcome-Modell mit  $\beta_1, \beta_2, \dots, \beta_M$ 
  - Bei Modellentwicklung helfen klinische und statistische Perspektiven

*„A combination of clinical judgment and empirical modeling produces better models than does either approach alone.“*

lezzoni, L. I.(2013). Risk Adjustment for Measuring Health Care Outcomes.

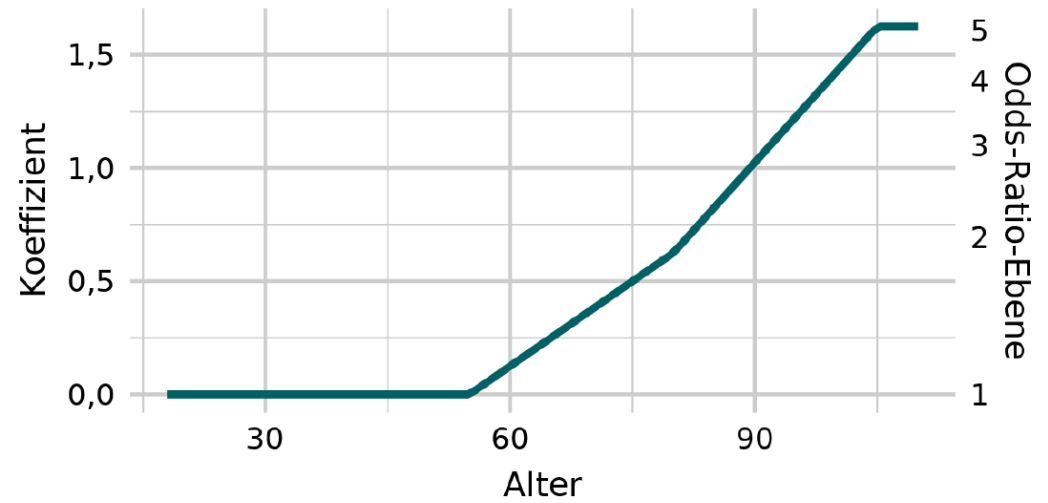
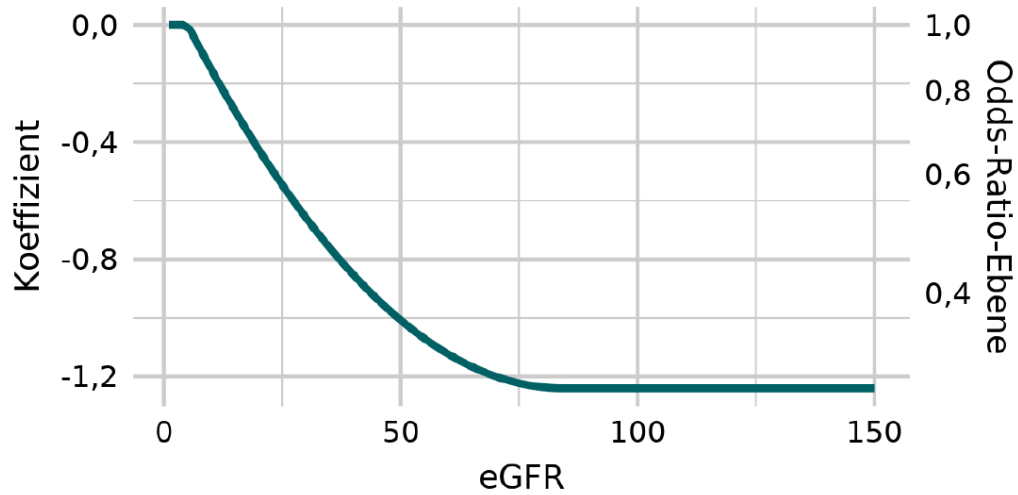
- Klinisch: z.B. Plausibilität, Vollständigkeit, Interpretierbarkeit
- Statistisch: z.B. Prognose-/Modellgüte, gut begründete Verteilungsannahmen und Schätzverfahren, Abwägung von Bias-Variance-Tradeoffs

## 5. Modellentwicklung: Stetige Einflussfaktoren (z.B. Alter, eGFR, BMI, Blutdruck, ...)

---

- Stetige und nicht-lineare Modellierung oft klinisch plausibel und empirisch gestützt
- Modellflexibilität muss abgesichert sein gegen Überanpassung
- Verlauf des geschätzten Effekts, v.a. im Randbereich mit wenig/keinen Daten sollte kritisch geprüft werden
- Flexibler Effekt lässt sich schwer an den (vielen) Koeffizienten  $\beta_1, \beta_2, \dots, \beta_M$  ablesen
- Verzicht auf Diskretisierung, stattdessen flexible Schätzverfahren (z.B. Splines)
- Pönalisierungsverfahren und Out-of-sample-Validierung
- Fachliche Plausibilisierung des Verlaufs, ggf. Trunkierung des Verlaufs am Rand
- Vereinfachung geschätzter Effekt (stückweise polynomiale Approximation) und Bericht von Effekt-Grafiken für Modelltransparenz

## 5. Modellentwicklung: Stetige Einflussfaktoren (z.B. Alter, eGFR, BMI, Blutdruck...)



QI 56022: MACCE nach Erst-PCI bei ST-Hebungsinfarkt (AJ 2025)

[https://iqtig.org/downloads/auswertung/aj2025/pci/DeQS-RL\\_PCI\\_AJ2025\\_MODELL\\_V01\\_2025-06-18.pdf](https://iqtig.org/downloads/auswertung/aj2025/pci/DeQS-RL_PCI_AJ2025_MODELL_V01_2025-06-18.pdf)

## 5. Modellentwicklung: Paradoxe Effekte (z.B. protektiver Effekt einer Vorerkrankung)

---

- Bekanntes Phänomen (z.B. Iezzoni, 2013; Becker et al., 2016): Widerspruch zwischen Effektrichtung/-stärke im geschätzten Modell und klinischer Hypothese
- Zielkonflikt zwischen empirischer Prognosegüte und Plausibilität / Augenscheinvalidität?

*„[...] if clinicians do not feel comfortable with the resulting models, much of the models' value is lost.“*

Iezzoni, L. I. (2013). Risk Adjustment for Measuring Health Care Outcomes.

- Entfernung vs. Aufnahme solcher Effekte ins Modell ist statistisch und klinisch zu diskutieren und zu dokumentieren

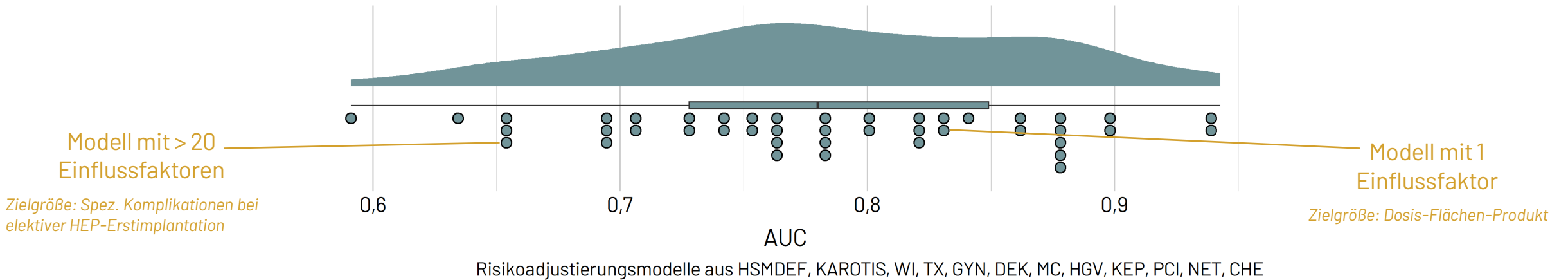
## 5. Modellentwicklung: Paradoxe Effekte (z.B. protektiver Effekt einer Vorerkrankung)

---

- Mögliche Erklärungsansätze:
  - (1) Unvollständige/fehlerhafte/selektive Dokumentation des Einflussfaktors
  - (2) „Klinische Hypothese“ hat andere Bezugsgruppe  $\neq$  Grundgesamtheit des QI (Interaktionseffekt)
  - (3) Assoziation mit anderen beobachteten Einflussfaktoren: Modellschätzer ist *konditional* zu interpretieren, die „klinische Hypothese“ hingegen oft *marginal*
  - (4) Assoziation mit unbeobachteten Einflussfaktoren
  - (5) Effekt agiert über indirekten Mechanismus
    - z.B. durch Vorerkrankung mehr Arztkontakt, frühere Diagnostik, besser medikamentös eingestellt
- Paradoxe Effekte sind u.U. informativ über Datenqualität oder angenommenes Wirkmodell
  - Je nach Erklärung ergeben sich andere Konsequenzen für die Modellentwicklung

## 5. Modellentwicklung: Grenzen reiner Prognosegütemaße

- Variation der Prognosegüte zwischen Modellen *verschiedener Indikatoren* ist erwartbar: Manche Ereignisse in manchen Populationen schwerer vorhersagbar als andere
- Beispiel AUC in Risikoadjustierungsmodellen (in-sample, AJ 2026):



- Perfekte Prognose durch Fallmerkmale widerspräche relevanter Rolle von Behandlungsqualität
- Prognosegüte auch durch ungeeignete Faktoren (z.B. beeinflussbar)
- Ziel des Modells ist Adjustierung des Qualitätsindikators, nicht individuelle Prognose

## 5. Modellentwicklung: Fazit & Ausblick

---

- Modellentwicklung für Zuordnung  $X_1, X_2, \dots, X_M \longrightarrow E(Y|X_1, X_2, \dots, X_M)$  ist „statistischster“ Teil der Risikoadjustierung – **doch auch hier klinische Perspektive wichtig**
  - Plausibilisierung von Modellannahmen und Ergebnissen der Modellwahl
- Eignung als Risikoadjustierungsmodell ergibt sich **nicht rein aus Prognosegüte**
  - Schätzfehler im „O/E“ hängt nicht nur von Prognosefehler ab
  - Doch oft nützlich während Modellentwicklung und für Modelltransparenz (IQTIG publiziert Maße)
- IQTIG arbeitet kontinuierlich an **Verbesserung und Harmonisierung** der Modellierungspraxis
  - Themen z.B.: Variablenselektion, fehlende Werte, hierarchische Modellierung, ..., teils Überlapp mit aktuellen Forschungsthemen der Literatur (z.B. Ullmann et al. 2024)
  - Neue Methoden aus (kausalem) Machine Learning vielversprechend (Vansteelandt, 2021; Susmann et al., 2025), aber Interpretierbarkeit als Herausforderung (Spoden et al., 2026) (?)

## 6. Gesamtbewertung der Risikoadjustierung

---

- Auf Basis der Einschätzungen zu:
  - Relevanz der Einflussfaktoren
  - Operationalisierbarkeit der Einflussfaktoren
  - Adäquate statistische Berücksichtigung der Einflussfaktoren (Modellentwicklung)
- Abstufungen der Gesamtbeurteilung der Risikoadjustierung:
  - vollständig angemessen
  - eingeschränkt angemessen
  - nicht angemessen
  - nicht erforderlich

## Mögliche Schwerpunkte der Diskussion

---

- Algorithmus möglich? – Wie standardisiert kann die Beurteilung erfolgen?
- Wie gut muss die Risikoadjustierung für welchen Zweck sein?
- Welchen Einfluss auf die Gesamteinschätzung sollten klinische vs. statistische Perspektive bzw. die Effekte im Modell haben? Welche Aspekte dieser Perspektiven besonders?
- Wie wichtig ist die Nachvollziehbarkeit/Lesbarkeit der statistischen Modelle für die Risikoadjustierung in der QS? (Stichwort: Machine Learning)
- Ist eine systematische Literaturrecherche nötig oder reicht eine orientierende? Ist die Suche nach Übersichtsarbeiten oder nach Primärliteratur effektiver?