

# Neue Rechnerische Auffälligkeitseinstufung in PCI

**Bund-Länder-Konferenz Kardiologie 2024**  
**29. November 2024**

Stefan Gehrig

Institut für Qualitätssicherung und Transparenz im Gesundheitswesen

# Agenda

---

1. Ziele der Umstellung
2. Blick auf die Methodik
3. Rückblick auf Ergebnisse
4. Herausforderungen in der QS bei kleinen Fallzahlen
5. Fazit und Ausblick

# 1. Ziele der Umstellung

## Zusammenfassung Ziele

---

- **Berücksichtigung von statistischer Unsicherheit**
  - Gleichziehen mit wissenschaftlichen und internationalen Standards
  - Weniger STNV bei Leistungserbringern ohne tatsächliches Qualitätsdefizit
  - Vermeidung des systematischen Fokus auf Leistungserbringer mit kleiner Fallzahl
- **Gezielte Aufwandssteuerung bei perzentilbasierten Referenzbereichen**
  - Anteil an rechnerischen Auffälligkeiten genau am gewählten Perzentil, statt darüber

## Zusammenfassung Ziele

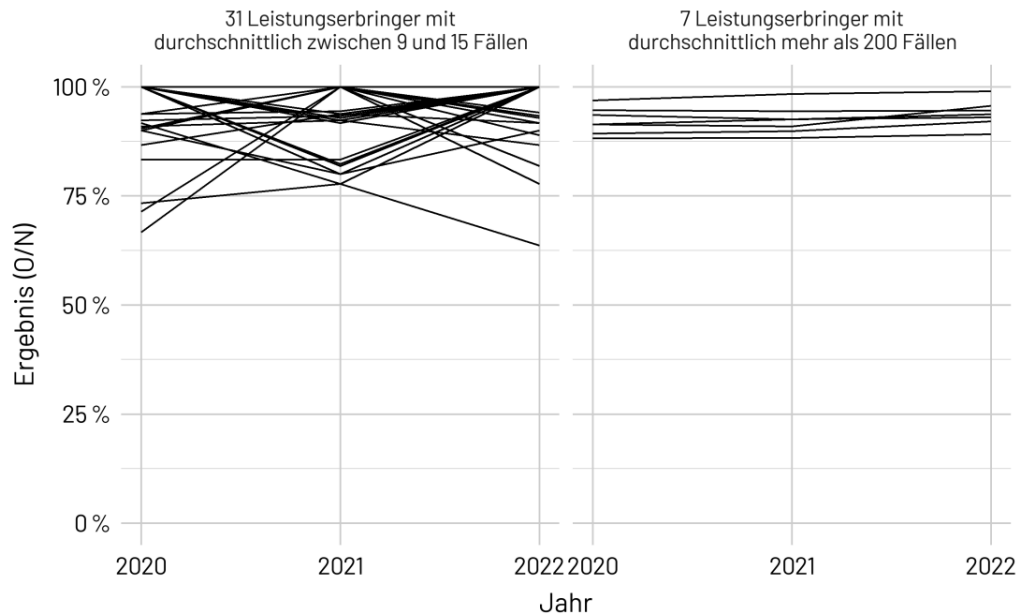
---

- **Berücksichtigung von statistischer Unsicherheit**
  - Gleichziehen mit wissenschaftlichen und internationalen Standards
  - Weniger STNV bei Leistungserbringern ohne tatsächliches Qualitätsdefizit
  - Vermeidung des systematischen Fokus auf Leistungserbringer mit kleiner Fallzahl
- **Gezieltere Aufwandssteuerung bei perzentilbasierten Referenzbereichen**
  - Anteil an rechnerischen Auffälligkeiten genau am gewählten Perzentil, statt darüber

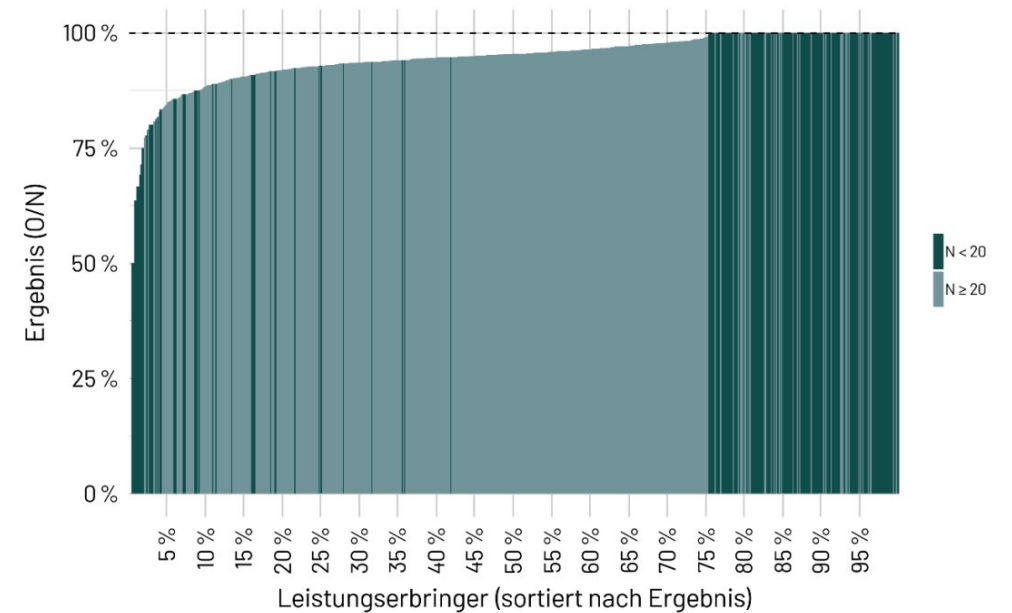
# Berücksichtigung von statistischer Unsicherheit

## Fallzahlabhängige Variabilität führt zu ...

... schwankenden QI-Ergebnissen vor allem bei kleinen LE



... extremen QI-Ergebnissen vor allem bei kleinen LE



## Berücksichtigung von statistischer Unsicherheit

---

- Fallzahlabhängige Variabilität muss für Schlussfolgerungen anhand QI-Ergebnissen berücksichtigt werden
  - Vermeidung eines rein „mechanischen“ Überfokus auf Leistungserbringer mit kleiner Fallzahl (aus Ressourcen- und Patientensicht)
- Nach alter Methodik der rechn. Auffälligkeit passiert Berücksichtigung **nur ad hoc und nicht systematisch:**
  - Bei der Festlegung perzentilbasierter Referenzbereiche (Ausschluss LE mit  $N < 20$ )
  - Bei der Festlegung fester Referenzbereiche (Mitdenken einer Toleranz gegen Zufallsschwankungen)
  - Bei der Auslösung von STNV durch LAGen? S. nächste Folie

# Berücksichtigung von statistischer Unsicherheit

Fallzahl und STNV  
(AJ 2021, QS PCI)

Fallzahl pro Qualitätsindikator (Grundgesamtheit)	Anzahl rechnerisch auffällige Ergebnisse	Anzahl Stellungnahmeverfahren (Anteil)	Anzahl qualitativer Auffälligkeiten (Anteil)
1. Quintil (1)	156	50 (32,1 %)	5 (3,2 %)
2. Quintil (2 – 10)	387	173 (44,7 %)	30 (7,8 %)
3. Quintil (11 – 46)	289	182 (63,0 %)	51 (17,6 %)
4. Quintil (47 – 252)	279	211 (75,6 %)	70 (25,1 %)
5. Quintil (253 – 6.326)	278	234 (84,2 %)	90 (32,4 %)
<b>Gesamt</b>	<b>1.389</b>	<b>850 (61,2 %)</b>	<b>246 (17,7 %)</b>

Bundesqualitätsbericht QS PCI 2022, Tabelle 20

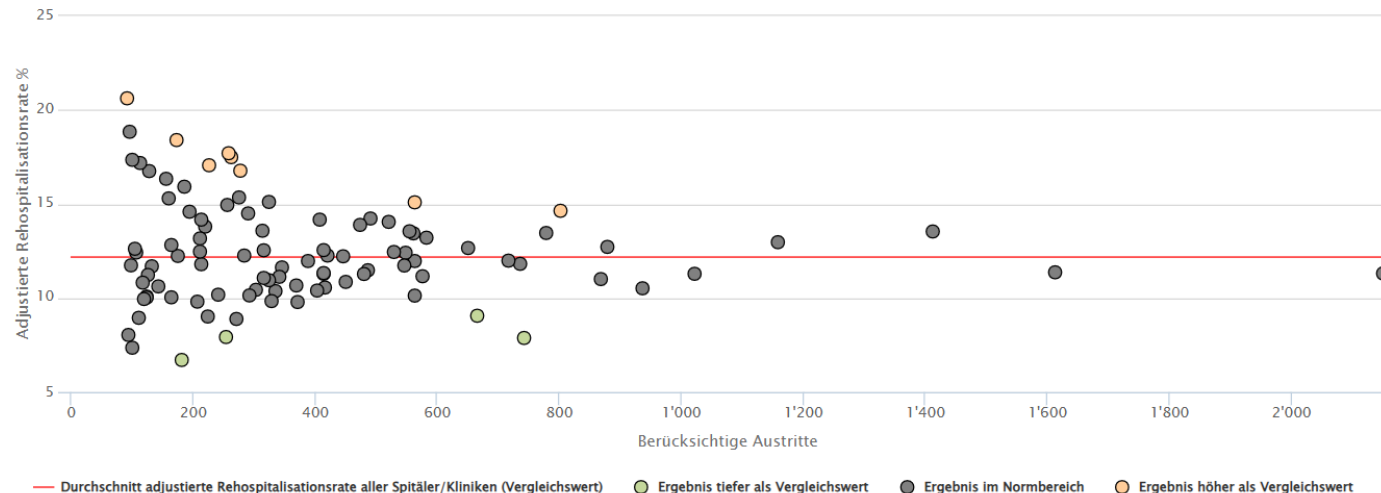
Fallzahl und Begründungen zum Auslassen  
des STNV (AJ 2023, alle QS-Verfahren)

QSEB Freitext-Angabe zu STNV nicht eingeleitet	Häufigkeit (alle QS-Verfahren)
geringe Fallzahl	23
geringe Fallzahl / erstmalige Auffälligkeit	88
geringe Fallzahlen (<=5 Fälle)	58
fehlende Signifikanz / niedriges Signifikanzniveau	251
geringe Fallzahlen oder Ausreißer	1
geringe GG	4
Einzelfall	570
FK: kein STNV aufgrund statistischer, methodischer oder inhaltlicher Gründe	173
FK: Einzelfall	4
...	...



## Berücksichtigung von statistischer Unsicherheit

- Neue rechnerische Auffälligkeitseinstufung ersetzt nicht-systematischen und nicht-einheitlichen Umgang mit fallzahlabhängiger Variabilität durch **etablierten Ansatz aus der Inferenzstatistik** (Hengelbrock et al., 2023)
- Inferenzstatistische Methoden sind **wissenschaftlicher Standard** in klinischer Forschung (z.B. Signifikanztests)
- Inferenzstatistische Methoden sind **internationaler Standard** beim Einrichtungsvergleich
  - Z.B. Empfehlung von „Funnelplots“ (Trichtergrafiken) statt Ergebnisranglisten (Spiegelhalter, 2005)



Beispiel ANQ (Schweiz): Ungeplante Rehospitalisierungen der kardiorespiratorischen Kohorte (Vergleichswert: nationales Mittel)

## Zusammenfassung Ziele

---

- **Berücksichtigung von statistischer Unsicherheit**
  - Gleichziehen mit wissenschaftlichen und internationalen Standards
  - Weniger STNV bei Leistungserbringern ohne tatsächliches Qualitätsdefizit
  - Vermeidung des systematischen Fokus auf Leistungserbringer mit kleiner Fallzahl
- **Gezielte Aufwandssteuerung bei perzentilbasierten Referenzbereichen**
  - Anteil an rechnerischen Auffälligkeiten genau am gewählten Perzentil, statt darüber

## Gezieltere Aufwandssteuerung bei perzentilbasierten Referenzbereichen

---

- Perzentilbasierte Referenzbereiche orientieren sich an der Verteilung der Indikatorergebnisse, um die Anzahl an rechnerischen Auffälligkeiten zu steuern
  - Durch alte Methodik werden Perzentile i.d.R. nicht eingehalten: Überschuss an kleinen LE ( $N < 20$ )
  - Neue perzentilbasierte Referenzbereiche + neue Auffälligkeitseinstufung: **genauest mögliche Einhaltung des Perzentils auf Bundesebene**

## 2. Blick auf die Methodik

## Zusammenfassung Methodik

---

**Neue Methodik:** Ein LE ist dann rechnerisch auffällig, wenn nach Beobachtung seines Indikatorergebnisses seine Wahrscheinlichkeit, den Referenzbereich einzuhalten, den Schwellenwert  $\alpha$  unterschreitet.

- In QS PCI gilt derzeit  $\alpha = 2,5\%$
- $\alpha$  bestimmt, ab wann eine Abweichung vom Referenzbereich nicht mehr plausibel durch zufallsbedingte Variabilität erklärbar ist

## Zusammenfassung Methodik

---

**Neue Methodik:** Ein LE ist dann rechnerisch auffällig, wenn nach Beobachtung seines Indikatorergebnisses seine Wahrscheinlichkeit, den Referenzbereich einzuhalten, den Schwellenwert  $\alpha$  unterschreitet.

- In QS PCI gilt derzeit  $\alpha = 2,5\%$
- $\alpha$  bestimmt, ab wann eine Abweichung vom Referenzbereich nicht mehr plausibel durch zufallsbedingte Variabilität erklärbar ist

**Vergleich alte Methodik:** Jedes beobachtete Indikatorergebnis außerhalb des Referenzbereichs führt zur rechnerischen Auffälligkeit.

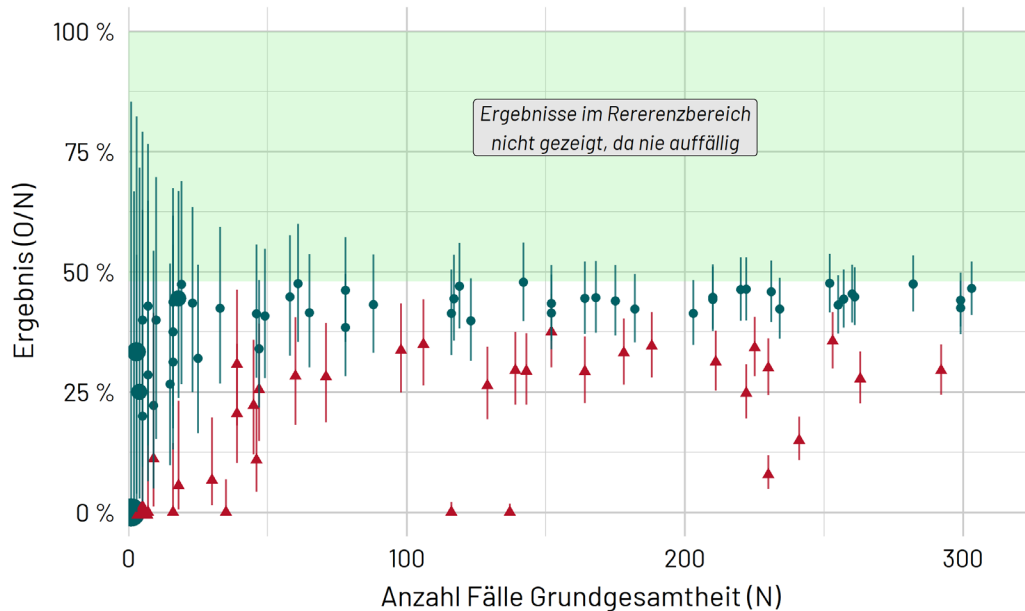
- Ein extremer Spezialfall der neuen Methodik: *keine* Abweichung vom Referenzbereich ist plausibel durch zufallsbedingte Variabilität erklärbar

Für *beide* Methodiken gilt: **Liegt ein QI-Ergebnis im Referenzbereich, ist es nicht auffällig**

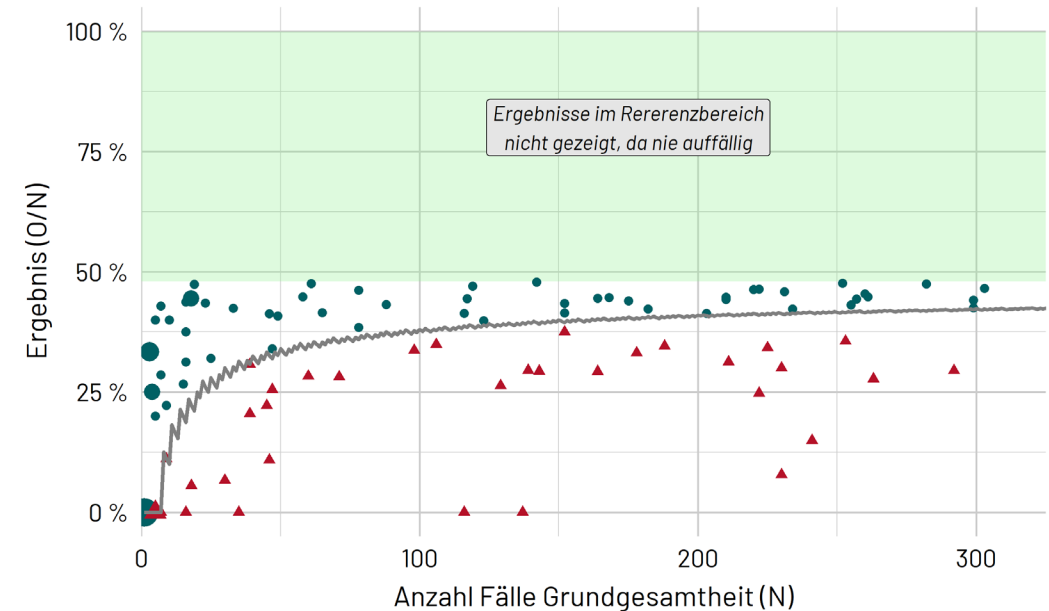
# Zusammenfassung Methodik

## Alternative Darstellung anhand von ...

### ... Unsicherheitsintervallen



### ... Trichterschwellen („Funnel-Boundaries“)



Rechnerisch auffällig    Anzahl Standorte  
 ▲ Ja    ● Nein    ● 2    ● 4    ● 6    ● 8

Beispiel: Fallzahlbereich [1,325] von QI 56000: Objektive, nicht-invasive Ischämiezeichen als Indikation zur elektiven, isolierten Koronarangiographie (AJ 2024)

# Neue Methodik bei perzentilbasierten Referenzwerten

- Perzentilbasierter Referenzwert muss sich an Verteilung der **Intervallgrenzen**, nicht mehr Verteilung der **Indikatorergebnisse**, orientieren

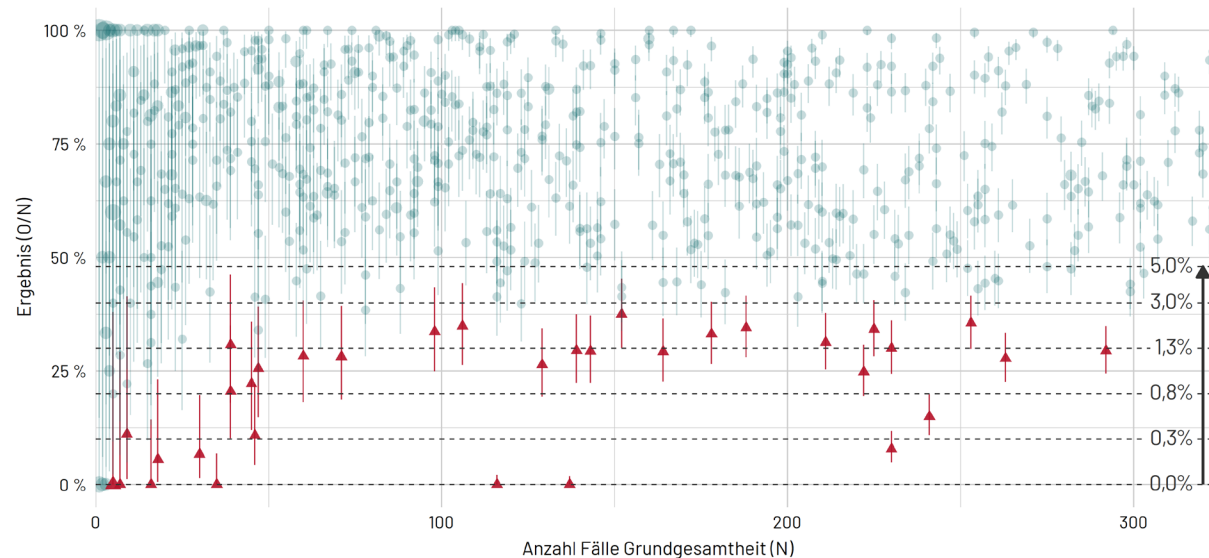


Illustration Referenzwertsuche am 5. Perzentil  
(Q1 56000 in AJ 2024, Fallzahlbereich [1,325])

- Referenzbereiche nominell strenger als mit alter Methodik, aber: Vergleich nun mit **Intervallgrenze**
  - Interpretation:** „Der Wert, bei dem X% auffällig werden“; kein medizinisches Kriterium für unzureichende Qualität (welches hingegen in festen Referenzbereich eingehen könnte), sondern rein aufwandsgesteuert
  - Referenzbereich am Bundeswert „gedeckt“: **Nur unterdurchschnittliches Ergebnis kann auffällig werden**



### 3. Rückblick auf Ergebnisse

## Rückblick auf Ergebnisse: Aufwand für STNV

- **Gezieltere Aufwandssteuerung bei perzentilbasierten Referenzbereichen** hat erwartete Aufwandsreduktion verursacht
  - Alle Klassik-Indikatoren QS PCI:

	Vor Umstellung		Nach Umstellung		
	AJ 2021	AJ 2022	AJ 2023	AJ 2024	
Anzahl rechn. Auff.	1.476	985	931	621	
Ø Anzahl rechn. Auff pro Indikator	73,8	51,8	51,7	51,8	

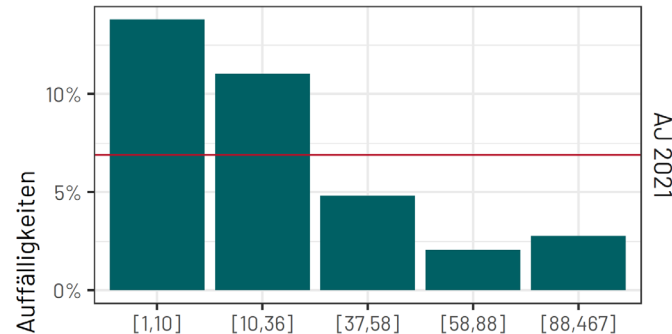
# Rückblick auf Ergebnisse: Fallzahlfokus der rechn. Auffälligkeit

- **Vermeidung des systematischen Fokus auf Leistungserbringer mit kleiner Fallzahl** wurde generell erreicht (variiert zwischen Indikatoren). Beispielhaft zwei Indikatoren:

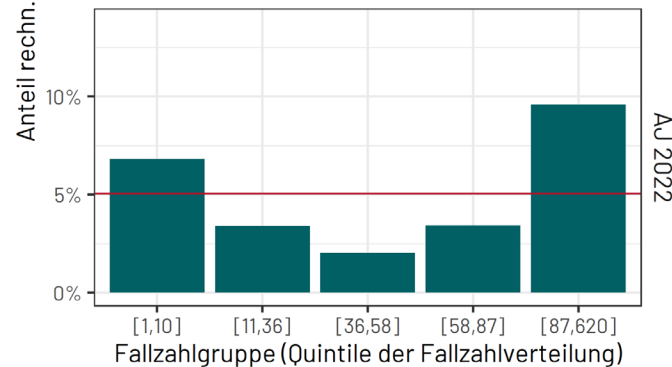
**Vor Umstellung:**

**QI 56014**

Erreichen des wesentlichen Interventionsziels bei PCI mit der Indikation ST-Hebungsinfarkt

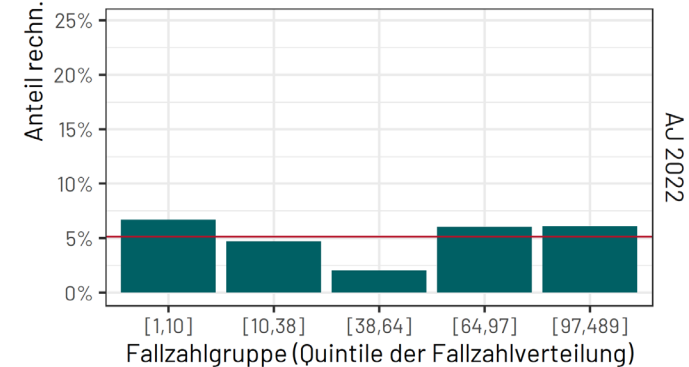
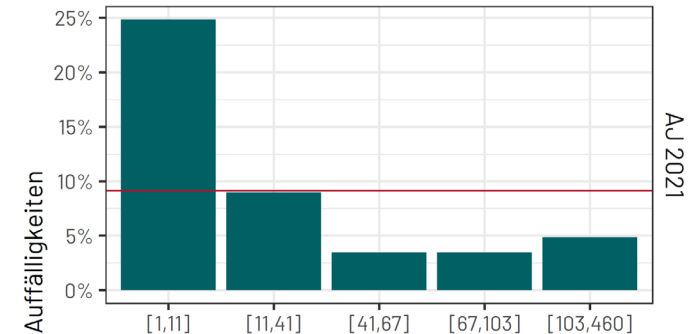


**Nach Umstellung:**



**QI 56022**

MACCE innerhalb von 7 Tagen bei Patientinnen und Patienten mit Erst-PCI bei ST-Hebungsinfarkt



— Durchschnittlicher Anteil im QI und AJ

## Rückblick auf Ergebnisse: Qualitative Auffälligkeiten

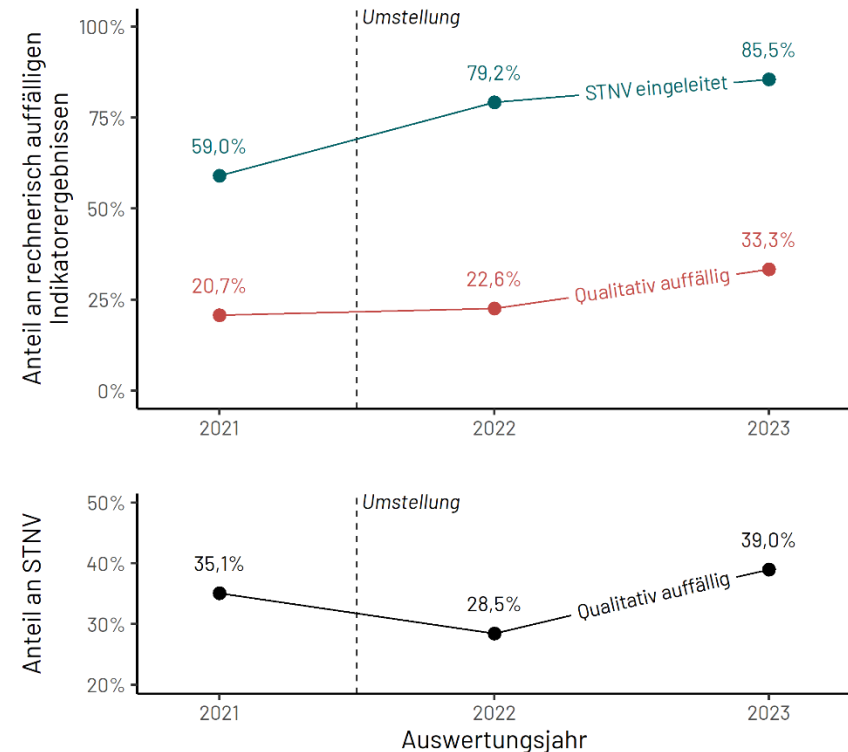
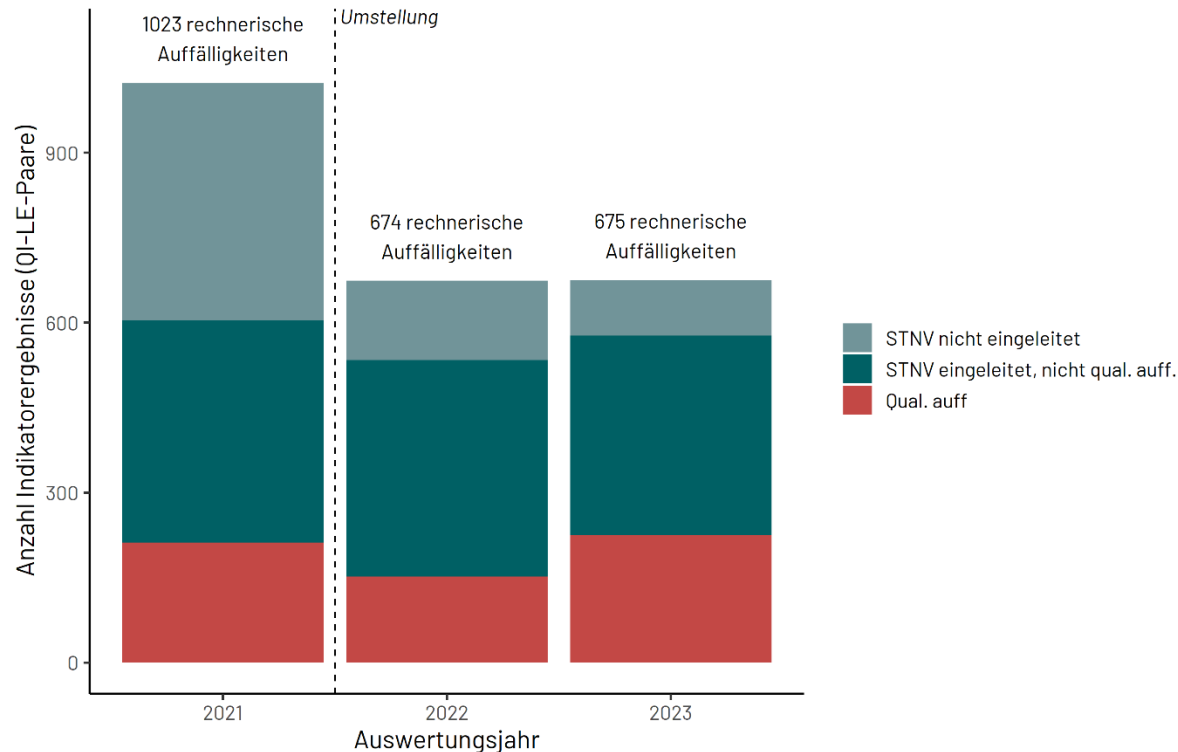
---

- Auswirkungen der neuen Methodik auf Treffsicherheit (z.B. Sensitivität) der rechnerischen Auffälligkeitseinstufung schwer zu evaluieren
  - Ergebnisse des STNV kein „Goldstandard für Qualität“ (vgl. IQTIG, 2020, Kapitel 2.5; IQTIG, 2022b)
  - Keine Ergebnisse der qualitativen Beurteilung für LE ohne rechnerische Auffälligkeit
- Aber Trends in qualitativen Auffälligkeiten legen nahe: „Ausbeute“ verbessert; s. nächste Folie
  - konsistent mit Ziel: **Weniger STNV bei Leistungserbringern ohne tatsächliches Qualitätsdefizit**

# Rückblick auf Ergebnisse: Qualitative Auffälligkeiten

## Zeitliche Betrachtung STNV nach DeQS-RL in QS PCI

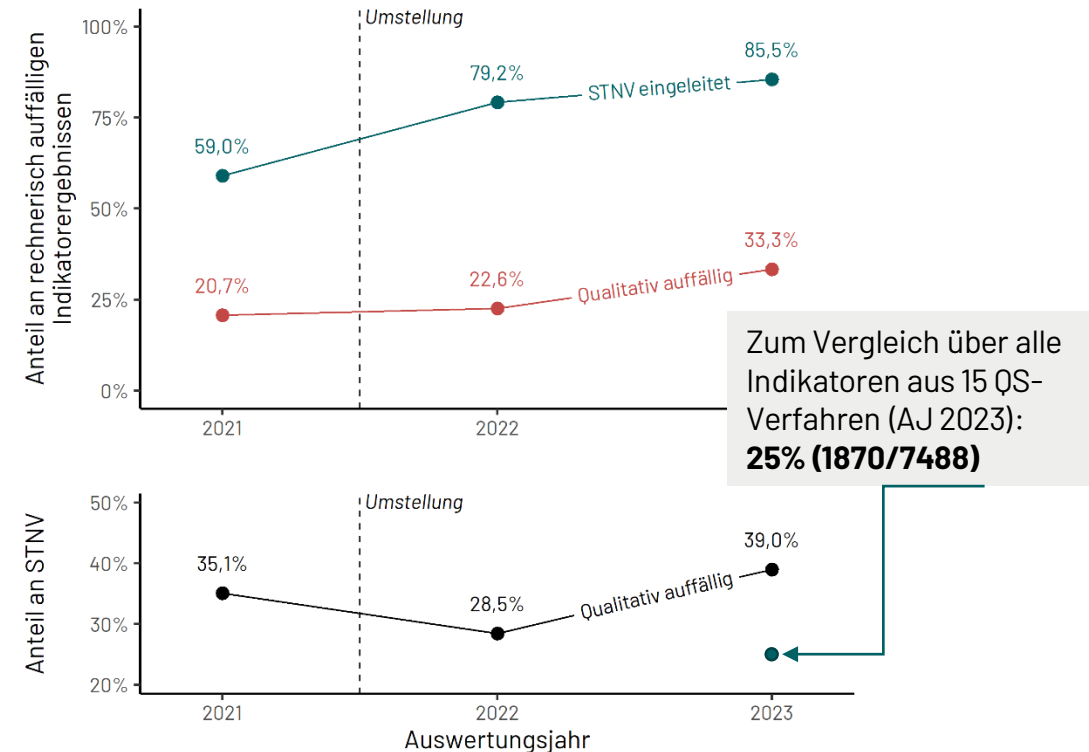
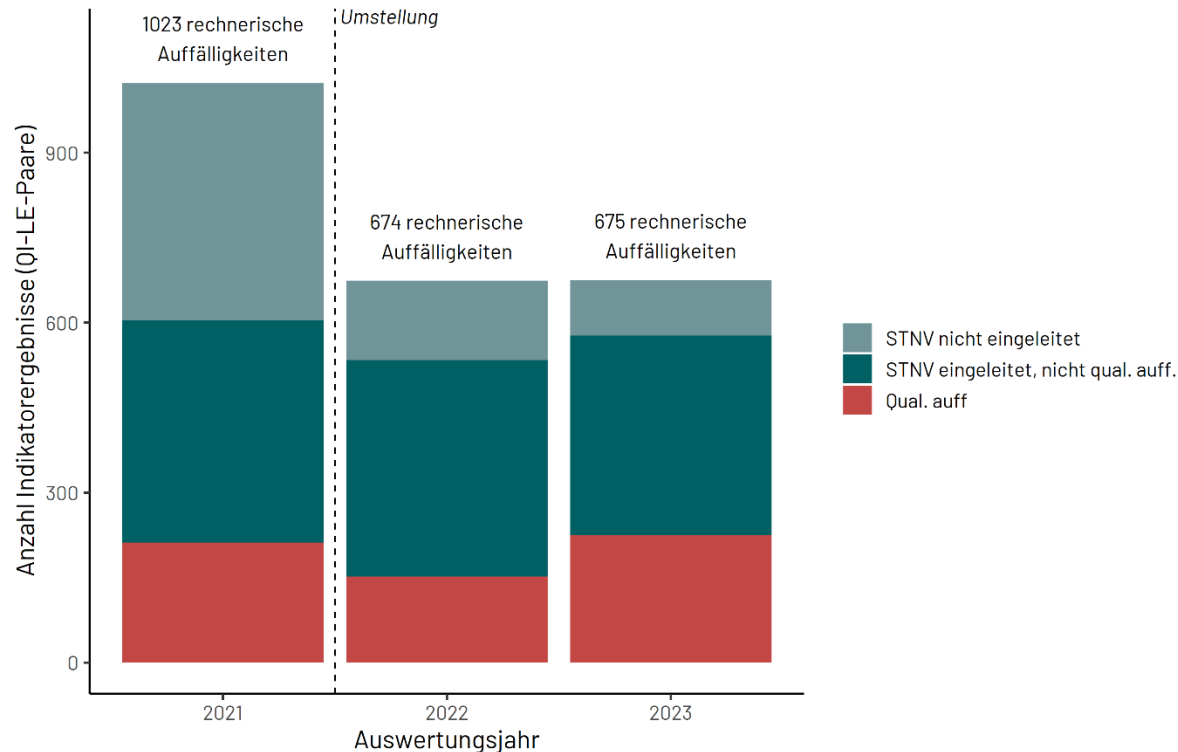
13 Indikatoren mit STNV-Daten für alle betrachteten Jahre: 56000 56001 56003 56004 56005 56006 56007 56008 56009 56010 56011 56014 56016 (ohne Sozialdaten-Indikatoren, da kein STNV für EJ 2019)



# Rückblick auf Ergebnisse: Qualitative Auffälligkeiten

## Zeitliche Betrachtung STNV nach DeQS-RL in QS PCI

13 Indikatoren mit STNV-Daten für alle betrachteten Jahre: 56000 56001 56003 56004 56005 56006 56007 56008 56009 56010 56011 56014 56016 (ohne Sozialdaten-Indikatoren, da kein STNV für EJ 2019)



## 4. Herausforderungen in der QS bei kleinen Fallzahlen

## Herausforderung 1 : Entscheidung unter großer Unsicherheit

---

- LE mit kleiner Fallzahl liefern wenig Information für ihre Bewertung. Dadurch Schlussfolgerungen über ihre Qualität schwierig, unabhängig von der statistischen Methodik.
  - Keine Einstufungsmethodik kann völlig gleiche Bedingungen für alle LE herstellen
- Stattdessen **Abwägung**: Wie Ressourcen fürs STNV verteilen?
  - Abwägung zwischen **Aufwand** (zusätzliche STNV) und potenziellem **Nutzen** (von Qualitätsförderung profitierende Patientinnen und Patienten) unter **Unsicherheit**
    - Kleinere LE: **größere Unsicherheit**
    - Schwellenwert  $\alpha$  ist Werkzeug zur Steuerung dieser Abwägung (IQTIG, 2020)
    - Je mehr Nutzen (weniger Aufwand) das STNV verspricht, desto leichter soll es auslösen (hohes  $\alpha$ )
    - Kann u.a. an Fallzahlverteilung im QS-Verfahren/Indikator ausgerichtet werden



## Herausforderung 2: Frage systematischer Defizite kleiner LE

---

- Es ist denkbar, dass kleinere LE in manchen Bereichen systematische Qualitätsdefizite aufweisen (positiver *Volume-Outcome-Zusammenhang*)
  - Rechnerische Auffälligkeit nach DeQS-RL ist nicht dafür konzipiert diesen Zusammenhang systematisch zu erforschen oder zu beheben
- Instrumente zur **Erforschung**:
  - Volume-Outcome-Analysen
    - Poolen über LE und Jahre mindert das statistische Fallzahlproblem (Gutzeit et al., 2022); Beispiele: TAVI (IQTIG, 2022a) und Herztransplantation (IQTIG, 2023).
  - Folgeabschätzungen hypothetischer Mindestmengen
    - Z.B. vermeidbare Ereignisse, Fallzahlveränderungen, Fahrzeiten/Wegstrecken
- Instrumente zur **Intervention**:
  - Politikmaßnahmen wie Mindestmengenregelungen, Struktur-/Personalvorgaben

## 5. Fazit und Ausblick

## Fazit und Ausblick

---

- Umstellung hat ihre Ziele generell erreicht
  - Schwer zu erforschen ist eine Verbesserung der Treffsicherheit
    - Ergebnisse weisen vorsichtig darauf hin (AJ 2023)
    - Reduziert wurde in jedem Fall die Anzahl an nicht eingeleiteten STNV: Mehr Vergleichbarkeit/Transparenz der STNV-Auslösung über Länder hinweg
- Durch die Umstellung der Methodik wurde auch Weiterentwicklungsbedarf sichtbar
  - Fokus auf „neue“, größere LE und ihre Kollektive führte zu neuen Rückmeldungen hinsichtlich Risikoadjustierung und (Sozialdaten-)Rechenregeln
  - Wahl von Schwellenwert  $\alpha = 2,5\%$  war starker Eingriff in bisherige QS-Routinen und kann für künftige Umstellungen weniger scharf gewählt werden ( $\alpha > 2,5\%$ )

## Fazit und Ausblick

---

- Geringer Informationsgehalt bei kleinen Fallzahlen ist Problem für jede Einstufungsmethodik
  - Um Fallzahlzusammenhänge in der QS zu erforschen und (regulatorisch) darauf zu reagieren sind andere Instrumente nötig
  - IQTIG prüft derzeit Bündelung von Informationen als Teil der G-BA-Beauftragung „Weiterentwicklung des Verfahrens der qualitativen Beurteilung“ (6. März 2024)
    - Ausweitung der Datengrundlage auf > 1 Jahr (s. auch Vach et al., 2022)
    - Clustern von Indikatoren in QI-Sets
- **Wichtig ist:** Die rechn. Auffälligkeit nach DeQS-RL ist ein sehr limitiertes Abbild von Qualität
  - Binäre rechnerischer Einstufung praktisch nützlich zur Auslösung des STNV für die Qualitätsförderung
  - In Auswahlentscheidungen / öffentliche Berichterstattung sollten mehr Informationen eingehen
    - Beispiel Bundes-Klinik-Atlas: 5 statt 2 Kategorien, andere Vergleichswerte, Strukturinformationen, Aggregation über Indikatoren, detaillierte Ergebnisse/Grafiken, ...

## Fazit und Ausblick

---

- Schrittweise Einführung in weiteren QS-Verfahren, um die übergangsweise verwendete Methodik abzulösen
  - Für AJ 2025 Erweiterung auf QS-Verfahren GYN-OP, MC, CHE (s. LAG-IQTIG-Treffen, 13.11.2024)
- Weiterhin freut sich das IQTIG über konstruktive Rückmeldungen zur Methodik und den betroffenen QS-Verfahren
- Wir sind optimistisch, dass die Vorteile der neuen Methodik langfristig für die meisten Akteure sichtbar werden

# Anhang (I)

---

## Referenzen

- Gutzeit, M., Rauh, J., Kähler, M., & Cederbaum, J. (2022). Modelling volume-outcome relationships in health care. *arXiv preprint arXiv:2203.12927*.
- Hengelbrock, J., Rauh, J., Cederbaum, J., Kähler, M., & Höhle, M. (2023). Hospital profiling using Bayesian decision theory. *Biometrics*, 79(3), 2757–2769.
- IQTIG (2020). Weiterentwicklung des Strukturierten Dialogs mit Krankenhäusern: Abschlussbericht zu Stufe 1 und Stufe 2. ([https://iqtig.org/downloads/berichte/2020/IQTIG\\_Weiterentwicklung-Strukturierter-Dialog-Stufe-1-u-2\\_Abschlussbericht\\_2020-02-11.pdf](https://iqtig.org/downloads/berichte/2020/IQTIG_Weiterentwicklung-Strukturierter-Dialog-Stufe-1-u-2_Abschlussbericht_2020-02-11.pdf))
- IQTIG (2022a). Datenauswertungen zu Mindestmengen bei TAVI: Abschlussbericht. ([https://iqtig.org/downloads/berichte/2022/IQTIG\\_Mindestmengen-TAVI-Abschlussbericht\\_2022-09-30\\_barrierefrei.pdf](https://iqtig.org/downloads/berichte/2022/IQTIG_Mindestmengen-TAVI-Abschlussbericht_2022-09-30_barrierefrei.pdf))
- IQTIG (2022b). Methodische Grundlagen: Version 2.0. ([https://iqtig.org/downloads/berichte-2/meg/IQTIG\\_Methodische-Grundlagen\\_Version-2.0\\_2022-04-27\\_barrierefrei.pdf](https://iqtig.org/downloads/berichte-2/meg/IQTIG_Methodische-Grundlagen_Version-2.0_2022-04-27_barrierefrei.pdf))
- IQTIG (2023). Datenauswertung zu Mindestmengen bei Herztransplantationen: Abschlussbericht. ([https://iqtig.org/downloads/berichte/2023/IQTIG\\_Datenauswertungen-bei-Mindestmengen-zu-HTX-Abschlussbericht\\_2023-10-06\\_barrierefrei.pdf](https://iqtig.org/downloads/berichte/2023/IQTIG_Datenauswertungen-bei-Mindestmengen-zu-HTX-Abschlussbericht_2023-10-06_barrierefrei.pdf))
- Spiegelhalter, D. J. (2005). Funnel plots for comparing institutional performance. *Statistics in medicine*, 24(8), 1185–1202.
- Vach, W., Wehberg, S., Güntert, B., Jakob, M., & Luta, G. (2022). Healthcare provider profiling: fixing observation period or fixing sample size?. *BMJ Open Quality*, 11(2), e001588.

## Anhang (II)

---

### Abbildungen teilweise aus:

- IQTIG (2023). Rechnerische Auffälligkeitseinstufung und perzentilbasierte Referenzbereiche: Umstellung der Methodik im Verfahren Perkutane Koronarintervention und Koronarangiographie (QS PCI). ([https://iqtig.org/dateien/berichte/2023/IQTIG\\_QS-PCI\\_Methodik-Auffaelligkeit-Referenzbereiche\\_2023-07-07.pdf](https://iqtig.org/dateien/berichte/2023/IQTIG_QS-PCI_Methodik-Auffaelligkeit-Referenzbereiche_2023-07-07.pdf))

**Gesammelte Materialien** zur Begründung und detaillierten Erläuterung der Methodik finden Sie auf der Webseite des IQTIG (<https://iqtig.org/das-iqtig/wie-wir-arbeiten/grundlagen/biometrische-grundlagen/biometrische-methodik-zur-auffaelligkeitseinstufung/>)

## Anhang (III)

### Anzahl rechn. Auffälligkeiten bei Klassik-Indikatoren in QS PCI, Auswertungsjahre 2021 bis 2024

QI	vor Umstellung			nach Umstellung								
	AJ 2021			AJ 2022			AJ 2023			AJ 2024		
	Anzahl	Anzahl Auff.	Anteil Auff.	Anzahl	Anzahl Auff.	Anteil Auff.	Anzahl 2023	Anzahl Auff.	Anteil Auff.	Anzahl	Anzahl Auff.	Anteil Auff.
56000	1.113	97	8,72%	1.089	54	4,96%	1.068	53	4,96%	1.031	52	5,04%
56001	1.155	103	8,92%	1.111	56	5,04%	1.108	55	4,96%	kein Refb.		
56002	1.217	61	5,01%	ausges.			ausges.			ausges.		
56003	698	57	8,17%	695	35	5,04%	700	35	5,00%	693	35	5,05%
56004	703	48	6,83%	708	35	4,94%	712	36	5,06%	701	35	4,99%
56005	1.227	83	6,76%	1.194	60	5,03%	1.193	58	4,86%	1.157	58	5,01%
56006	684	59	8,63%	689	35	5,08%	697	35	5,02%	688	34	4,94%
56007	997	70	7,02%	973	49	5,04%	976	48	4,92%	971	49	5,05%
56008	1.251	129	10,31%	1.223	121	9,89%	1.216	125	10,28%	1.197	119	9,94%
56009	1.233	89	7,22%	1.203	60	4,99%	1.200	60	5,00%	kein Refb.		
56010	686	67	9,77%	692	35	5,06%	698	35	5,01%	kein Refb.		
56011	1.001	75	7,49%	974	49	5,03%	979	49	5,01%	kein Refb.		
56012	1.209	76	6,29%	1.248	62	4,97%	1.220	62	5,08%	1.195	59	4,94%
56014	725	50	6,90%	733	37	5,05%	731	36	4,92%	728	35	4,81%
56016	1.014	96	9,47%	971	48	4,94%	981	50	5,10%	kein Refb.		
56018	1.193	64	5,36%	1.230	63	5,12%	1.199	60	5,00%	1.173	58	4,94%
56020	946	72	7,61%	1.005	49	4,88%	967	49	5,07%	971	49	5,05%
56022	722	66	9,14%	744	38	5,11%	750	37	4,93%	746	38	5,09%
56024	944	45	4,77%	997	50	5,02%	972	48	4,94%	kein Refb.		
56026	942	69	7,32%	996	49	4,92%	kein Refb.			kein Refb.		
<b>Alle</b>	<b>19.660</b>	<b>1.476</b>	<b>7,51%</b>	<b>18.475</b>	<b>985</b>	<b>5,33%</b>	<b>17.367</b>	<b>931</b>	<b>5,36%</b>	<b>11.251</b>	<b>621</b>	<b>5,52%</b>



# Anhang (IV)

## Beispiel: Perzentilbasierter Referenzbereich vorher vs. nachher (QI 56014, AJ 2024)

