

## Verknüpfung von Krankenkassen- und Registerdaten: Projekt WiZen

Jochen Schmitt<sup>1</sup>, Olaf Schoffer<sup>1</sup>, Martin Rößler<sup>1</sup>, Christoph Bobeth<sup>1</sup>, Veronika Bierbaum<sup>1</sup>, Michael Gerken<sup>2</sup>, K. Kleihues-van Tol<sup>3</sup>, Patrik Dröge<sup>4</sup>, Thomas Ruhnke<sup>4</sup>, Andreas Klöss<sup>4</sup>, Christian Günster<sup>4</sup>, Monika Klinkhammer-Schalke<sup>2,3</sup>

1) ZEGV, 2) TZR, 3) ADT, 4) WiDO

8. DNVF-Forum Versorgungsforschung

12.5.2021



Gefördert durch



**Gemeinsamer  
Bundesausschuss**  
Innovationsausschuss

Förderkennzeichen: 01VSF17020



# Zertifizierung in der Onkologie

- Nationaler Krebsplan:
  - Handlungsfeld 2, Ziel 5: Schaffung einheitlicher Konzepte und Bezeichnungen für die Qualitätssicherung, Qualitätsförderung und Zertifizierung onkologischer Behandlungseinrichtungen
- Zentren: zertifizierte Netzwerke (stationäre und ambulante Einrichtungen) mit enger Zusammenarbeit aller Fachrichtungen zur Behandlung von Krebspatienten
  - Unterschiedliche Zertifizierungen



- 2020 sind mehr als 1.400 Zentren in Deutschland durch DKG zertifiziert

## Bisherige Evidenz

Überlebensvorteile in Zentren bzw. nach Zertifizierung bzgl. Kolonkarzinom (GKV-Daten)<sup>1</sup>, bzgl. Kolorektalem Karzinom (KKR-Daten)<sup>2</sup> sowie bzgl. Mammakarzinom (Zentren-Daten)<sup>3</sup>

1: Trautmann, Reißfelder, Pecqueux, Weitz, Schmitt: "Evidence-based quality standards improve prognosis in colon cancer care". EJSO, 2016

2: Völkel, Draeger, Gerken, Fürst, Klinkhammer-Schalke: "Langzeitüberleben von Patienten mit Kolon- und Rektumkarzinomen: Ein Vergleich von Darmkrebszentren und nicht zertifizierten Krankenhäusern". Gesundheitswesen, 2018

3: Kreienberg, Wöckel, Wischnewsky: "Highly significant improvement in guideline adherence, relapse-free and overall survival in breast cancer patients when treated at certified breast cancer centres: An evaluation of 8323 patients". Breast, 2018



# Studiendesign und Ziel

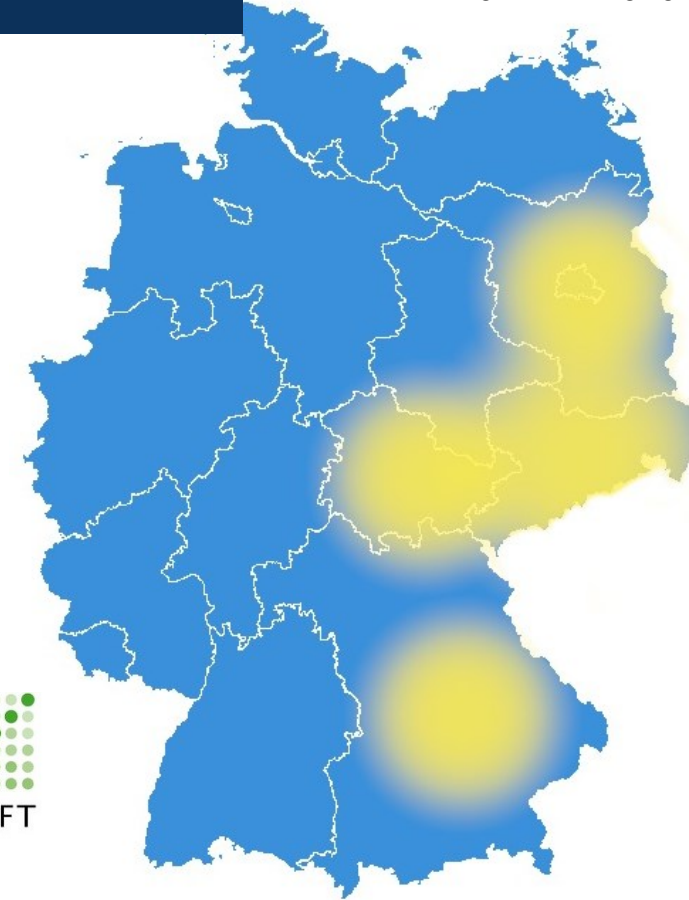
- **Studiendesign:** Sekundärdatenbasierte Kohortenstudie
- **Hypothese:** Patienten mit Behandlung in zertifizierten Zentren haben bessere Behandlungsergebnisse im Vergleich mit nicht zertifizierten Kliniken
- **Ziel:** Aggregierter Vergleich zertifizierter Zentren mit nicht zertifizierten Kliniken hinsichtlich vielfältiger Outcomes:
  - Gesamtüberleben, 1- bis 5-Jahres-Überlebensraten, 30-Tage Mortalität
  - Rezidivfreie Überlebenszeit, Fernmetastasierungs-, Nachresektions- und Komplikationsraten bei OP
  - Abbildung der Versorgungswege und regionaler Ausprägungen

# WiZen – Datengrundlage

- AOK-Routinedaten  
(ca. 20 Mio. Versicherte)
  - Daten regionaler klinischer Krebsregister  
(KKR)
  - Gelinkte Daten AOK und KKR
- 
- Ergänzt durch Daten der DKG zur  
Zertifizierung
  - Ergänzt durch Daten strukturierter  
Qualitätsberichte nach §137 SGB
  - Daten über 9 Jahre: 2009-2017



Förderkennzeichen:  
01VSF17020





# Breite: Zeitraum und Entitäten

- Im Zeitraum 2009-2017 werden inzidente Fälle von 8 Entitäten untersucht, und zwar
  - Kolorektales Karzinom (C18/19 und C20)
  - Pankreaskarzinom (C25)
  - Mammakarzinom (C50, +/- D05)
  - Gynäkologische Tumoren – Endometrium, Cervix, Ovar (C53, C54, C56)
  - Prostatakarzinom (C61)
  - Lungenkrebs (C34, +/- C33/D38.1)
  - Kopf-Hals-Tumoren (C00-14 und C30-32)
  - neuroonkologische Tumoren (C47, C70-C72, C75.1/.2/.3, C83.3, C85, D32, D33, D35.2/.3/.4)

# Methoden

- Deskription, Survivalanalysen und Mehrebenenmodelle
  - Adjustierung für patientenseitige Risikofaktoren (Alter, Geschlecht, Schweregrad, Sekundärmalignom) und Strukturmerkmale der Krankenhäuser (z.B. Bettenzahl)
- Festlegung von Falldefinitionen für Erkrankung und Behandlung anhand von Expertenkonsultationen
- Schwerpunkte der Auswertung
  - Daten 1 (AOK-Daten): Komorbiditäten, Therapiearten/-verläufe
  - Daten 2 (KKR-Daten): Schweregrade, Rezidive
  - Daten 3 (AOK+KKR-Daten): Abbildung des Schweregrades in AOK-Daten

# Ergebnisse Kolon-, Pankreas- und MammaCa

## ■ Kohorte:

- Daten des WIdO (bundesweite AOK-Daten 2009-2017)
- Falldefinition unter **Einbezug** systematischer **klinischer Expertise** (insg. 8 x 2-4 Experten)

Stationäre Entlassdiagnose ICD-10: **Kolon- C18/C19, Pankreas- C25, MammaCa C50/D05**

vorab mindestens 3 Jahre ohne entsprechende Diagnose außer vorhergehende 12 Monate ambulant, durchgehende Versicherung inkl. Verstorbene, Wohnort in Deutschland, Mindestalter 18 Jahre, bei Kolon: Therapie innerhalb eines Quartals der Entlassdiagnose

- Zertifizierung gemäß DKG  
„zertifiziert“, falls Krebsdiagnose in zertifizierter Klinik und Diagnose nach Zertifikatserteilung

## ■ Methodik der Survivalschätzungen:

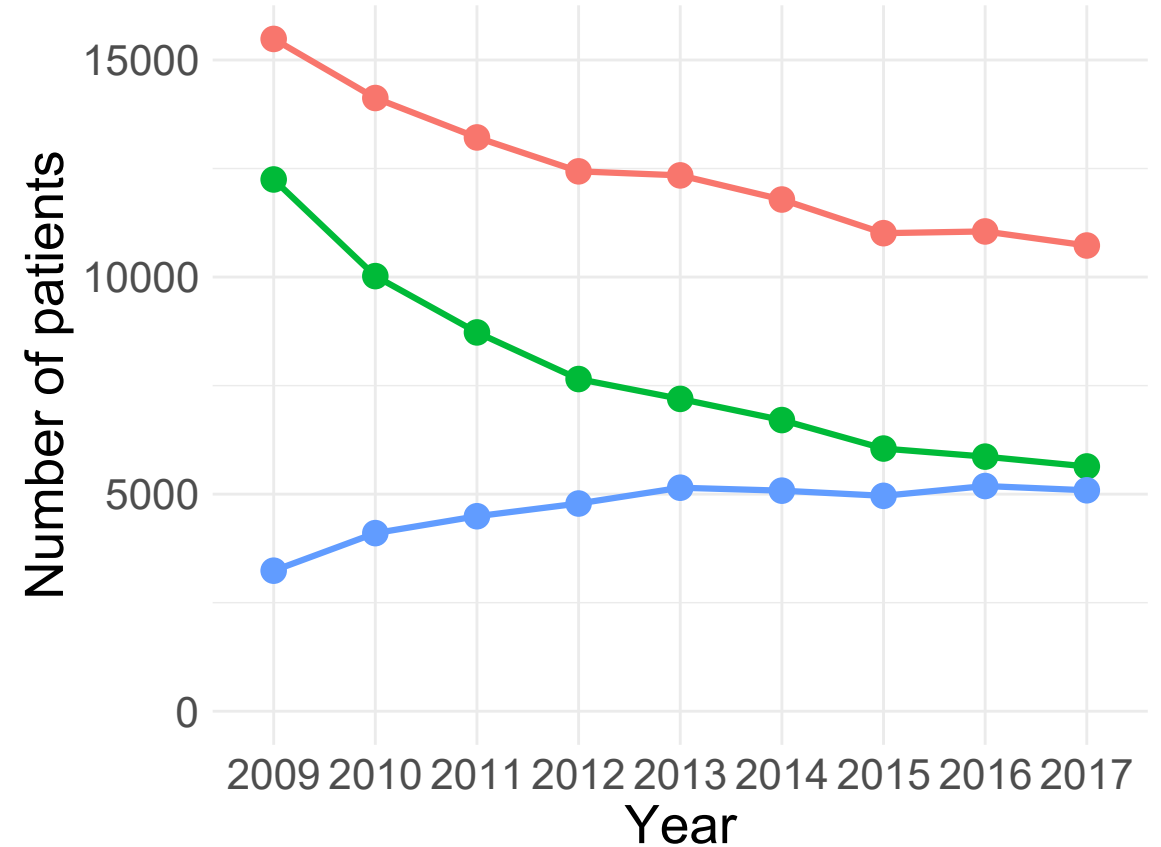
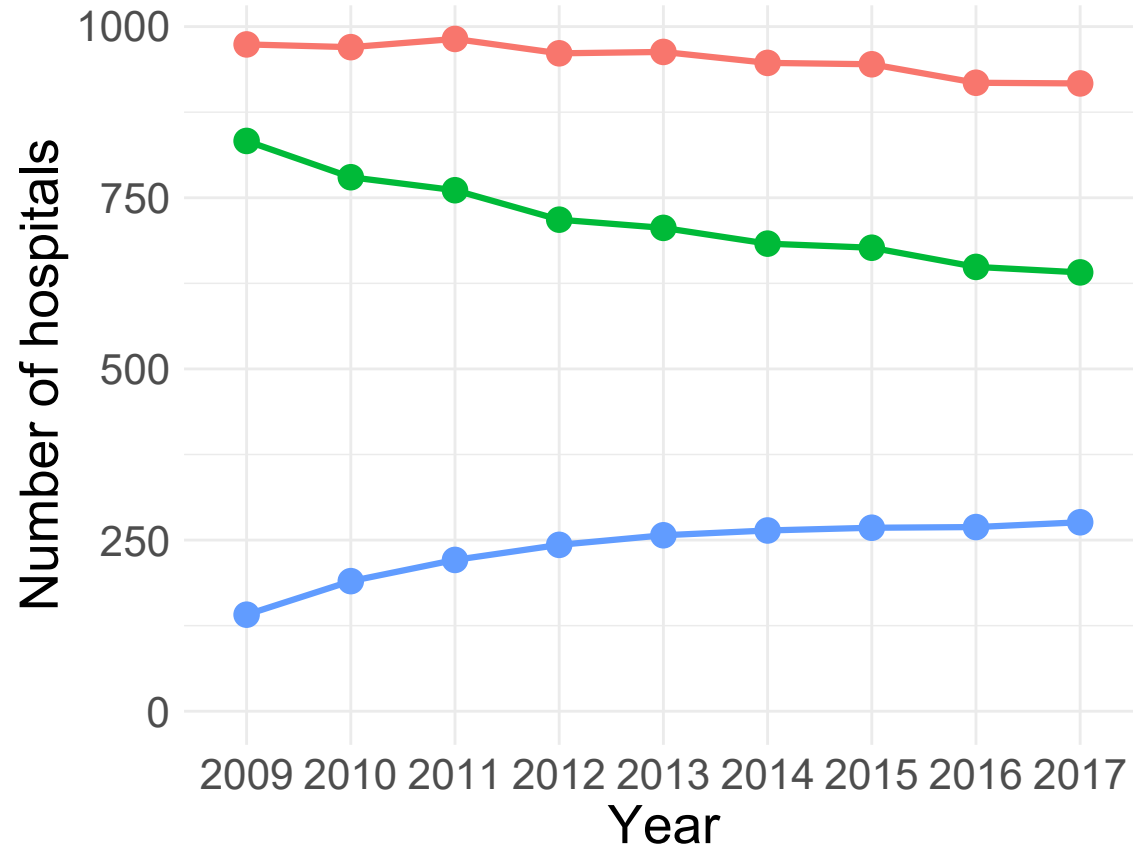
- Relatives Überleben stratifiziert nach Zertifizierungsstatus
- Cox-Regression (Hazard ratio des Zertifizierungsstatus) mit Shared frailty adjustiert nach Patientencharakteristika, Klinikcharakteristika und Jahr der Behandlung

# KolonCa: Untersuchte Population

Variable	Certified: no	(n=70,098)	Certified: yes	(n=42,079)
Age in years, Median (Q1;Q3)	75	(68;82)	75	(67;81)
Sex, n (%)				
female	35,203	(50.2%)	20,374	(48.4%)
male	34,895	(49.8%)	21,705	(51.6%)
Hospital beds, n (%)				
1-299	36,247	(51.7%)	3,599	(8.6%)
300-499	20,191	(28.8%)	14,479	(34.4%)
500-999	9,332	(13.3%)	17,712	(42.1%)
1000+	4,328	(6.2%)	6,289	(14.9%)



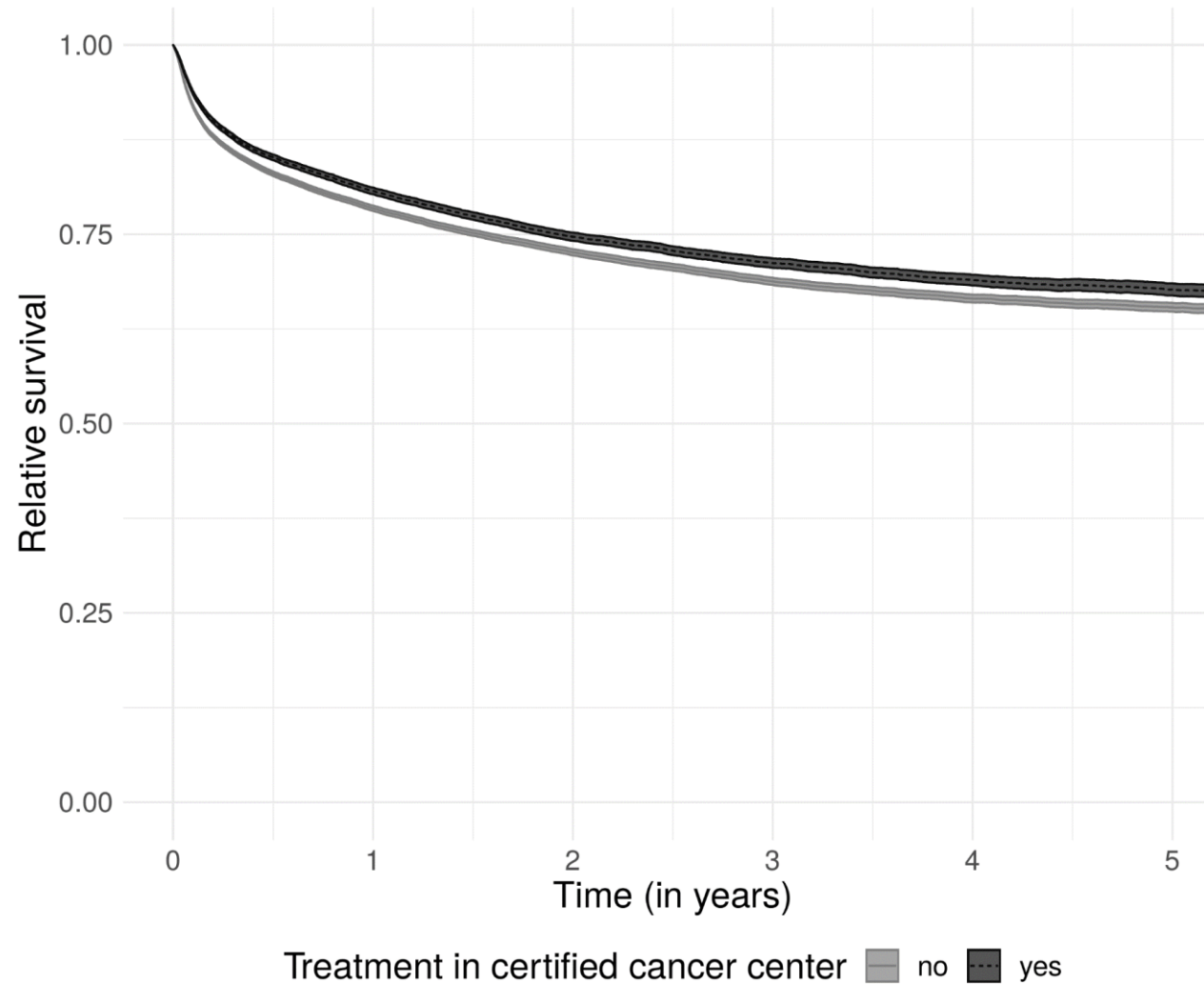
# KolonCa: Zertifizierung im Zeitverlauf



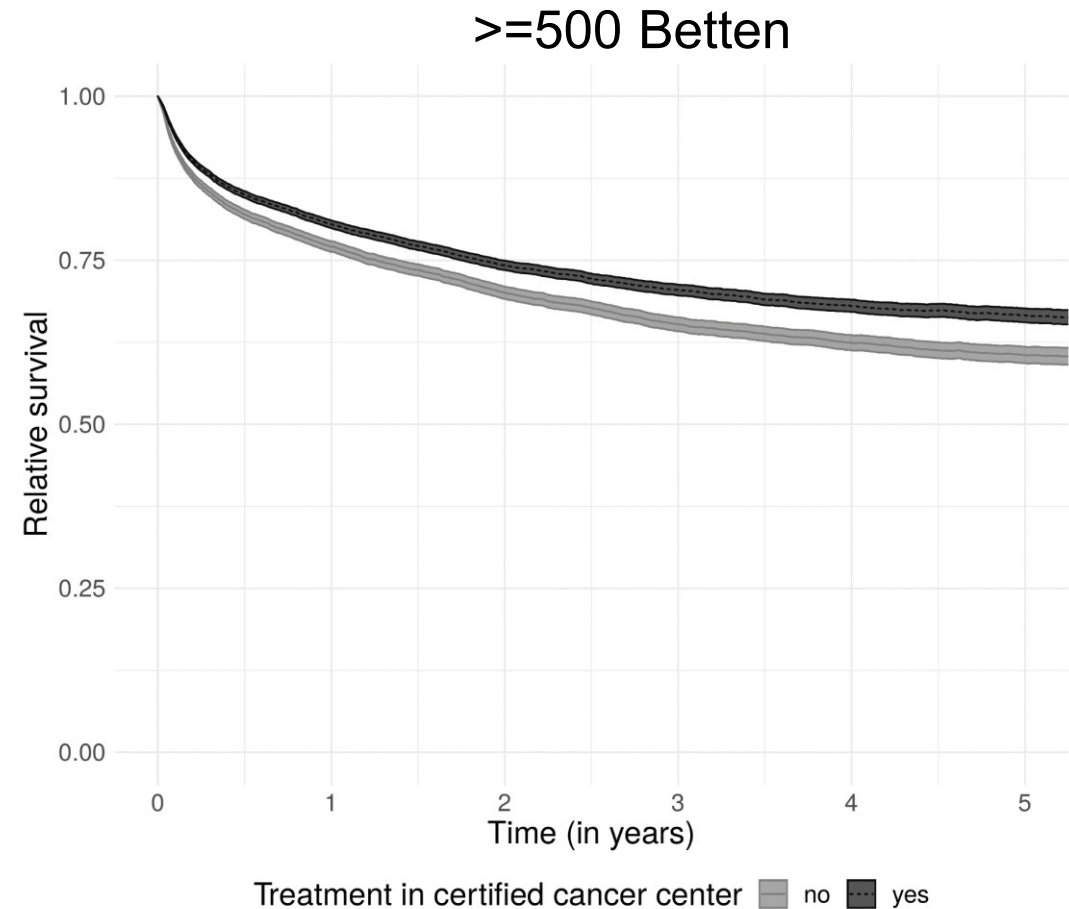
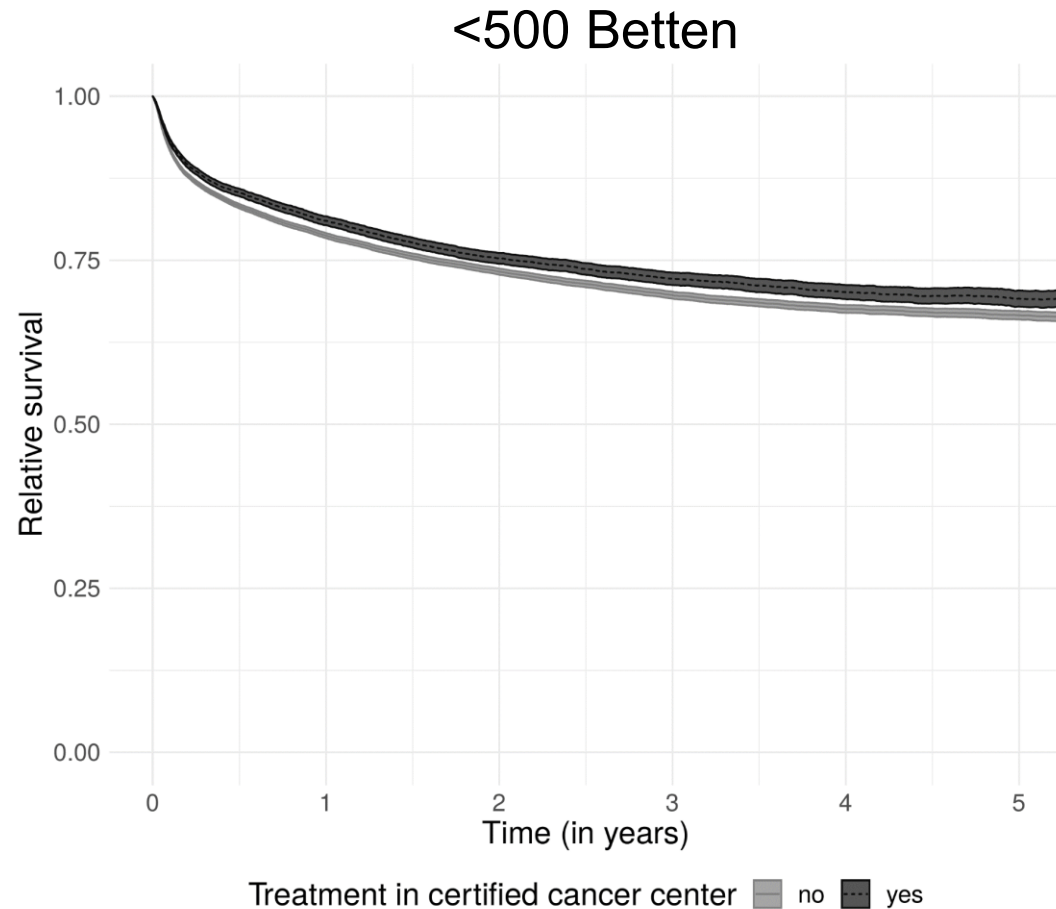
Group —●— All —●— Certified: no —●— Certified: yes



# KolonCa: Überleben und Zertifizierung



# KolonCa: Relatives Überleben und Bettenzahl

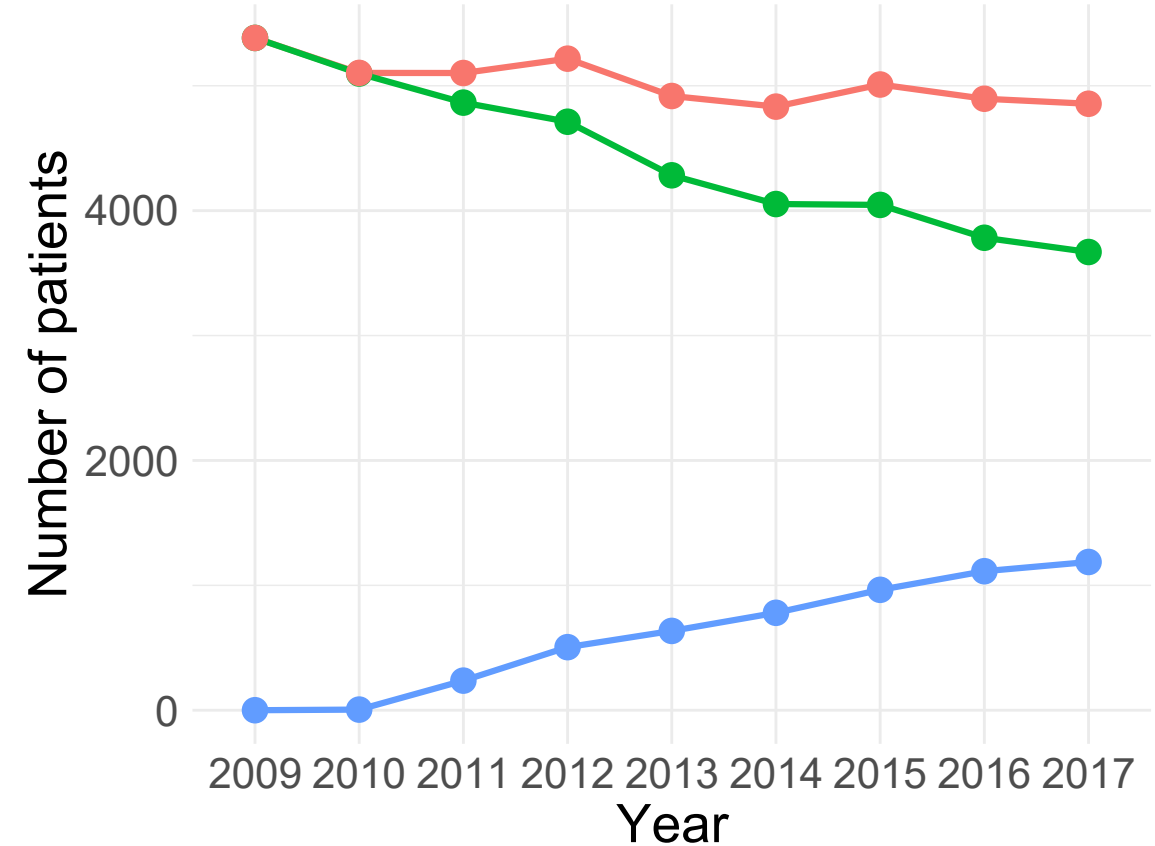
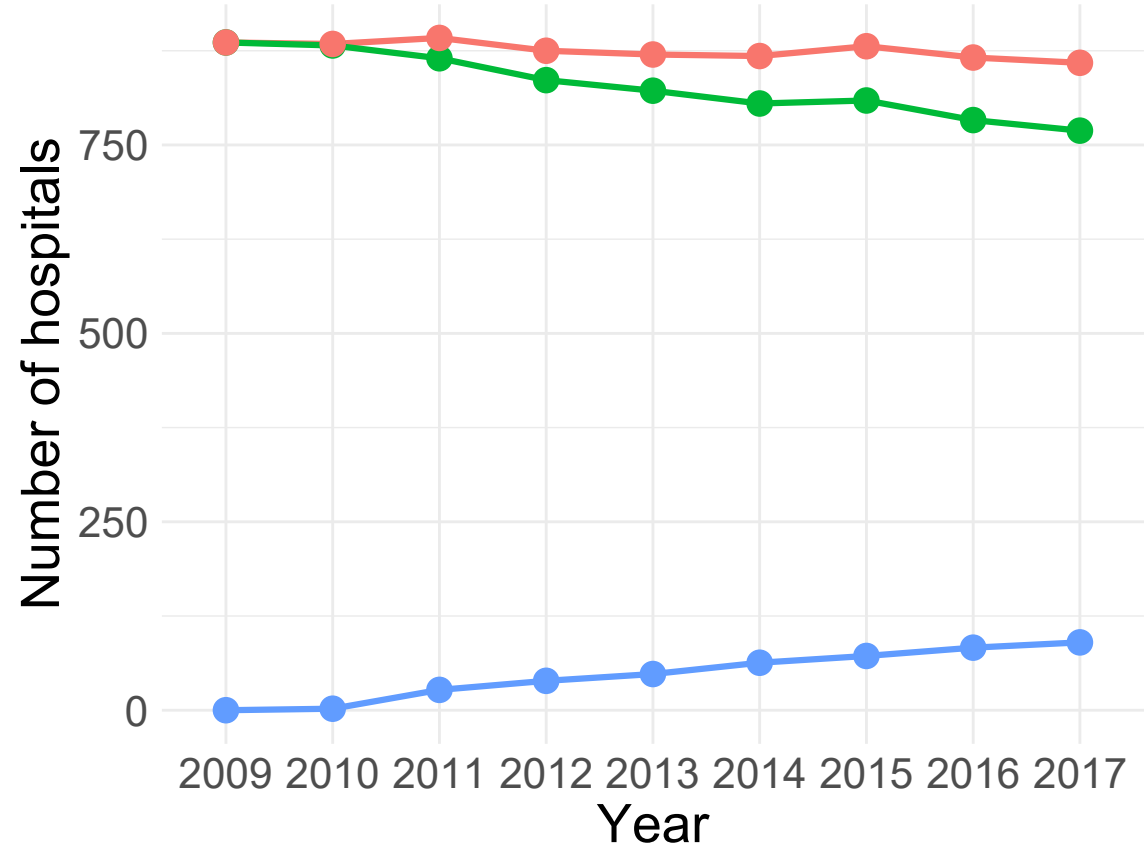


■ Aber: Bettenzahl  $\neq$  Menge der behandelten Patienten

# PankreasCa: Untersuchte Population

Variable	Certified: no	(n=39,892)	Certified: yes	(n=5,426)
Age in years, Median (Q1;Q3)	74	(67;81)	73	(64;79)
Sex, n (%)				
female	20,859	(52.3%)	2,754	(50.8%)
male	19,033	(47.7%)	2,672	(49.2%)
Hospital beds, n (%)				
1-299	11,836	(29.7%)	154	(2.8%)
300-499	12,684	(31.8%)	659	(12.1%)
500-999	9,549	(23.9%)	2,866	(52.8%)
1000+	5,823	(14.6%)	1,747	(32.2%)

# PankreasCa: Zertifizierung im Zeitverlauf

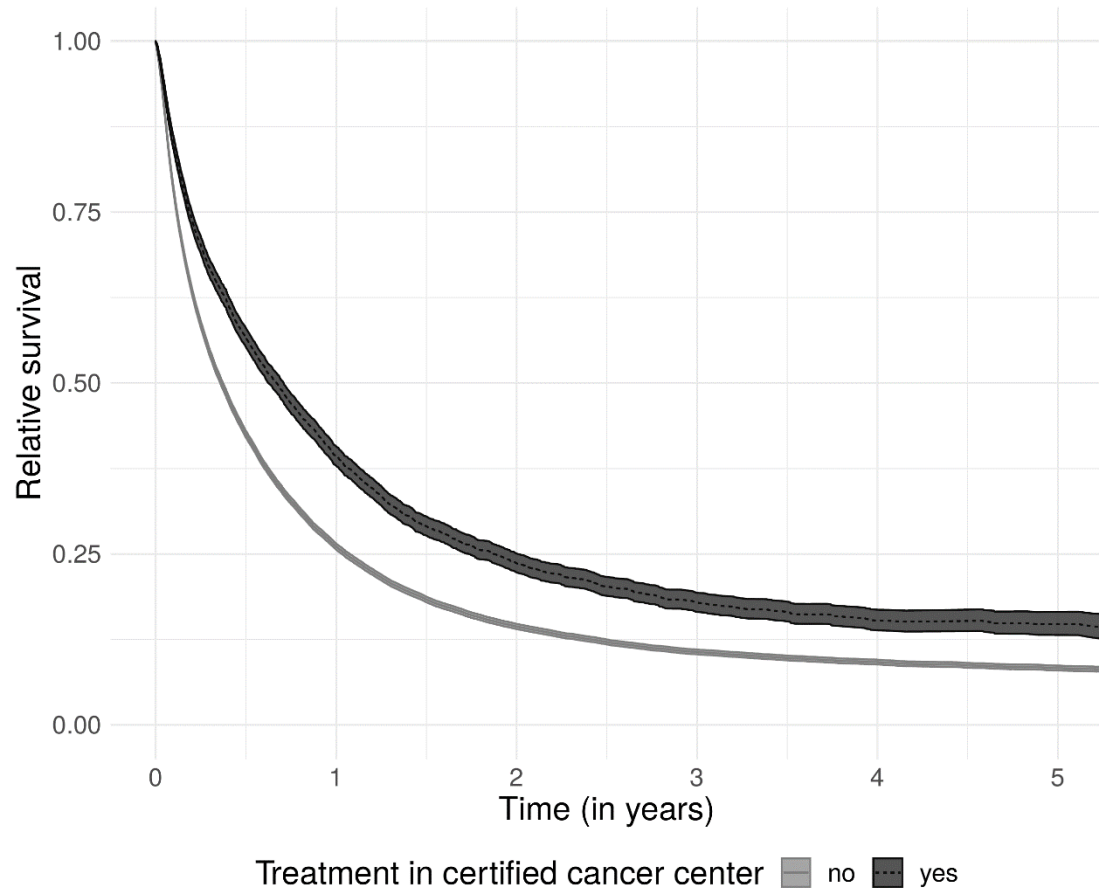


Group — All — Certified: no — Certified: yes



# PankreasCa: Überleben und Zertifizierung

Cox regression with shared frailty



Variable	HR	CI
Certified center (ref: no)	-	-
yes	0.89***	(0.85,0.93)
Age (ref: 18-59)	-	-
60-79	1.46***	(1.41,1.51)
80+	2.47***	(2.38,2.57)
Sex (ref: female)	-	-
male	1.03**	(1.01,1.05)
Hospital beds (ref: 1-299)	-	-
300-499	0.95*	(0.92,0.99)
500-999	0.88***	(0.84,0.92)
1000+	0.82***	(0.77,0.87)
Number of patients	45,318	
Number of hospitals	1,051	
SD(RE)	0.15	

HR=Hazard ratio, CI=95%-confidence interval,

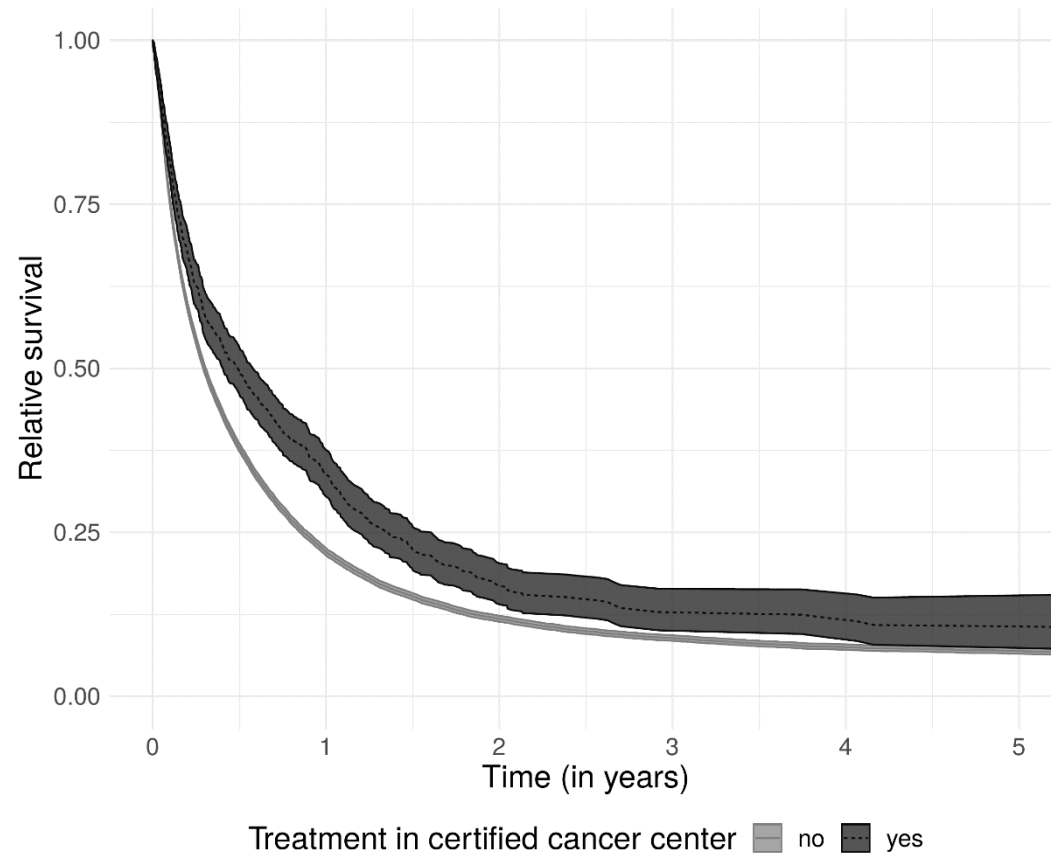
significance levels: \*5%, \*\*1%, \*\*\*0.1. Further covariates included in the regression but not shown in the table



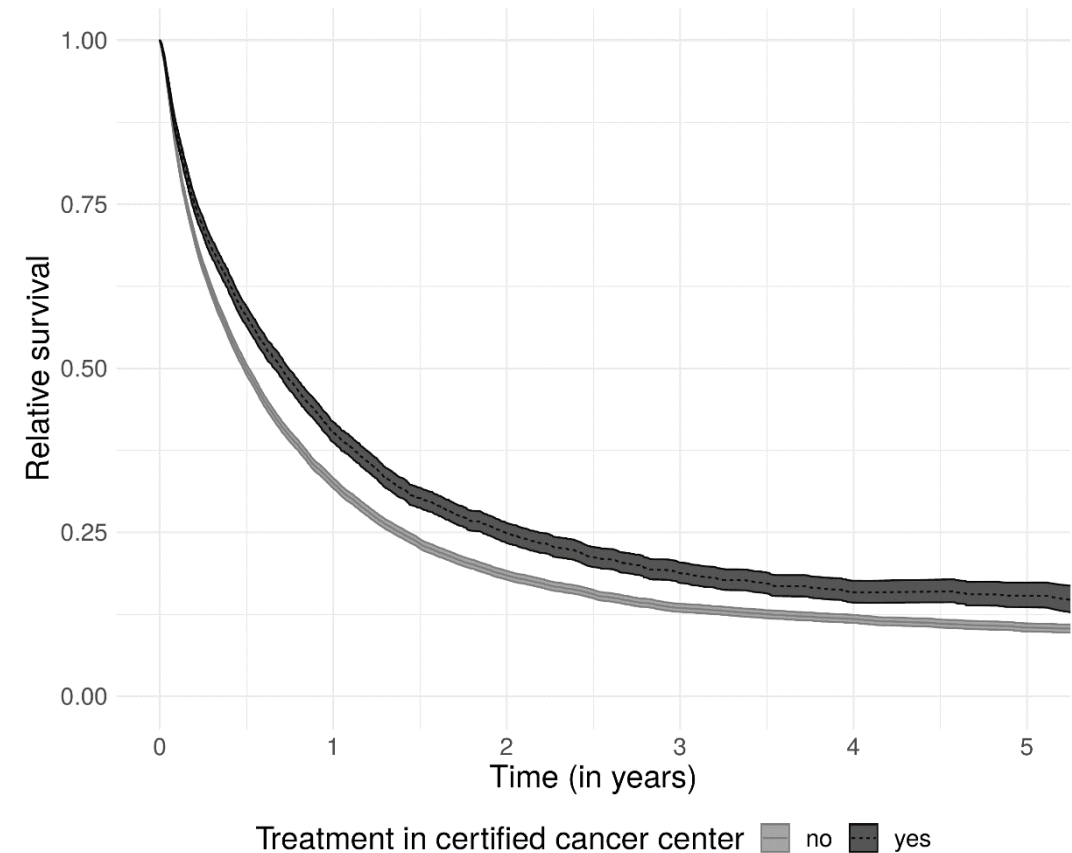


# PankreasCa: Relatives Überleben und Bettenzahl

<500 Betten



>=500 Betten

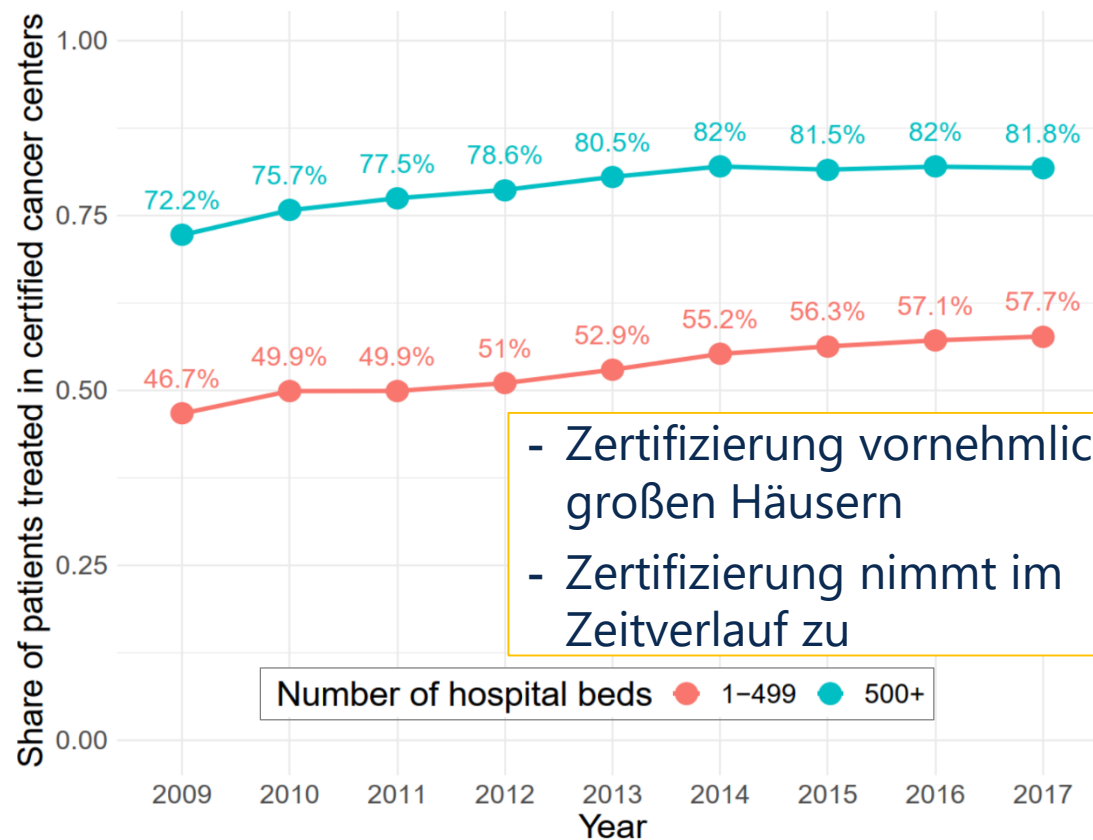


■ Aber: Bettenzahl  $\neq$  Menge der behandelten Patienten

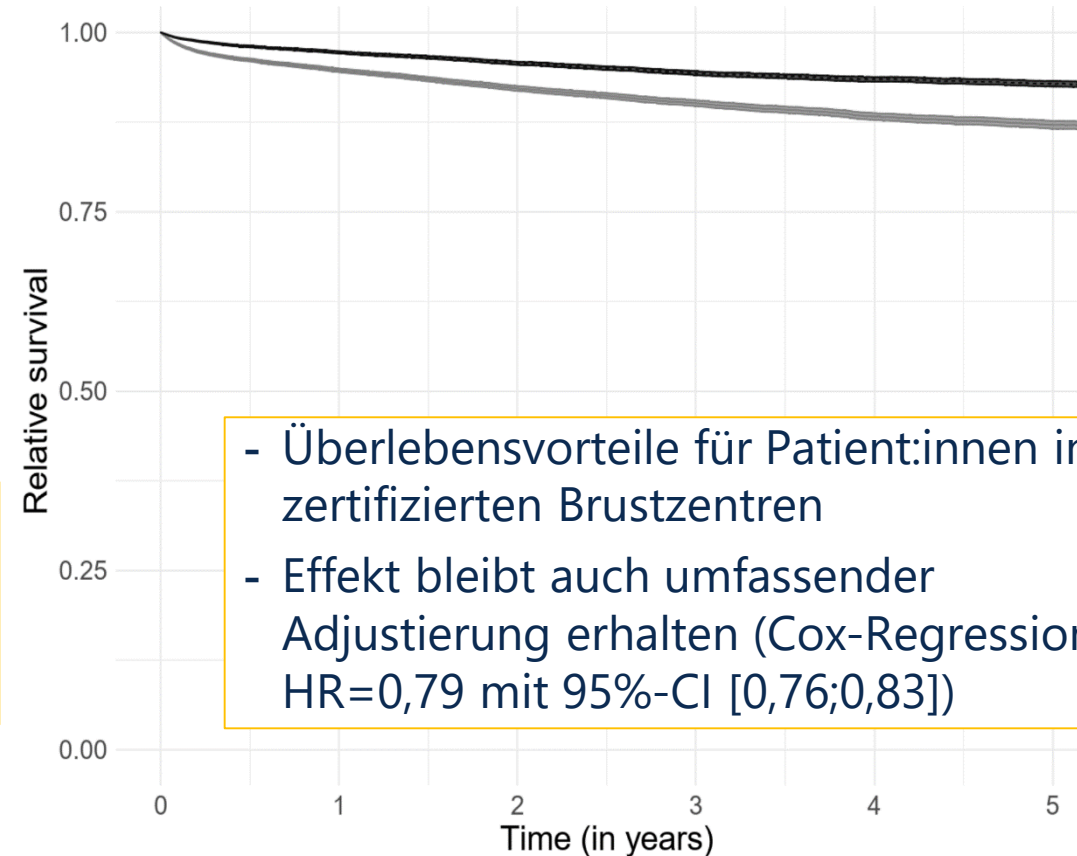


# Ergebnisse MammaCa

- Behandlung in
  - 280 zertifizierten Zentren: n=90.430
  - 734 nicht zertifizierten Kliniken: n=51.941



- Zertifizierung vornehmlich in großen Häusern
- Zertifizierung nimmt im Zeitverlauf zu



- Überlebensvorteile für Patient:innen in zertifizierten Brustzentren
- Effekt bleibt auch umfassender Adjustierung erhalten (Cox-Regression: HR=0,79 mit 95%-CI [0,76;0,83])

Treatment in certified cancer center ■ no ■ yes





# WiZen – Stärken/Limitationen

- Stärken
  - > 1000 Krankenhäuser, hohe Fallzahlen (z.B.  $\approx 140.000$  MammaCa)
  - Viele relevante Confounder berücksichtigt
    - Komorbiditäten,
    - onkologische Zweiterkrankungen, Klinikmerkmale etc.
  - Längsschnittdaten: komplette Krankheitsverläufe 2009-2017
- unbekannte Faktoren
  - Histologie und genaue Erkrankungsschwere nur in KKR-Daten
  - Keine Blutwerte, Genetik, Biomarker
  - Kein Lebensstil und sozioökonomischer Status
  - Volume nicht gut abbildbar
- Keine Randomisierung möglich → Vorsicht mit kausaler Schlussfolgerung

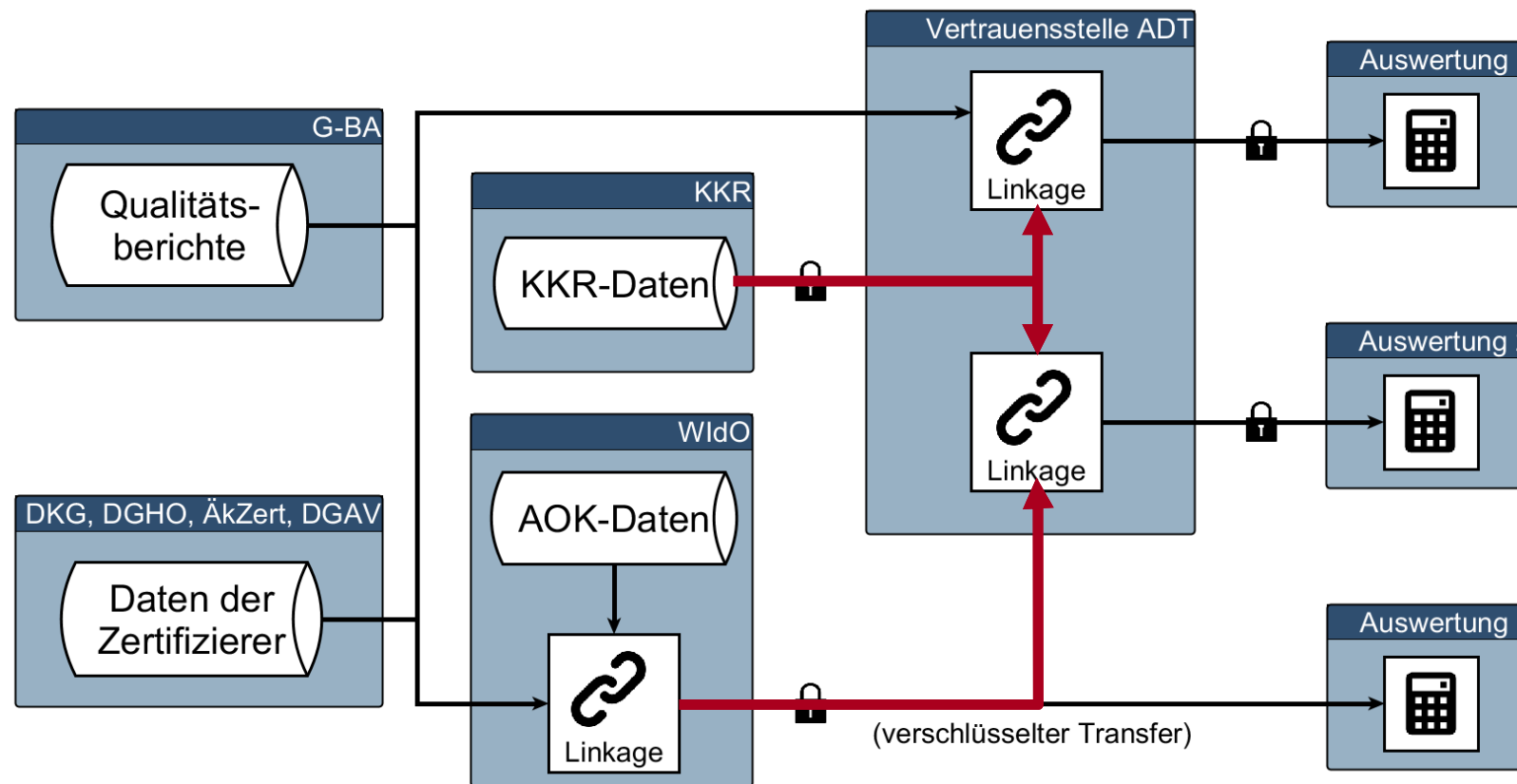


# Datenlinkage



# Datenfluss (vereinfacht)

- Wichtig ist die Trennung von Datenmanagement und -auswertung
- Beim Linkage werden sensible Daten in unserer Vertrauensstelle bei der ADT entfernt



# Daten-Linkage in WiZen

- **Population:** AOK-Versicherte wohnhaft in Regionen der KKR
- **Verlässlichkeitsstufen von Linkage-Variablen**
  - Direkte Identifikatoren
    - KV-Nummer (Achtung, zeitlicher Trend hinsichtlich Verfügbarkeit)
  - Indirekte Identifikatoren
    - Variablen mit hoher Verlässlichkeit: Geburtsdatum, Geschlecht, 5-stellige PLZ, ...
    - Unvollständige Variablen: z.B. Therapie (in KKR nicht zu 100% erfasst)
    - Approximierte Variablen: Diagnosedatum (exakt in KKR, in GKV-Daten nur approximativ)
  - Ausgeschlossene Variablen: z.B. Todesdatum (Risiko von Selektionsbias)
- **Ziel:** Informationen über „das Beste aus beiden Welten“

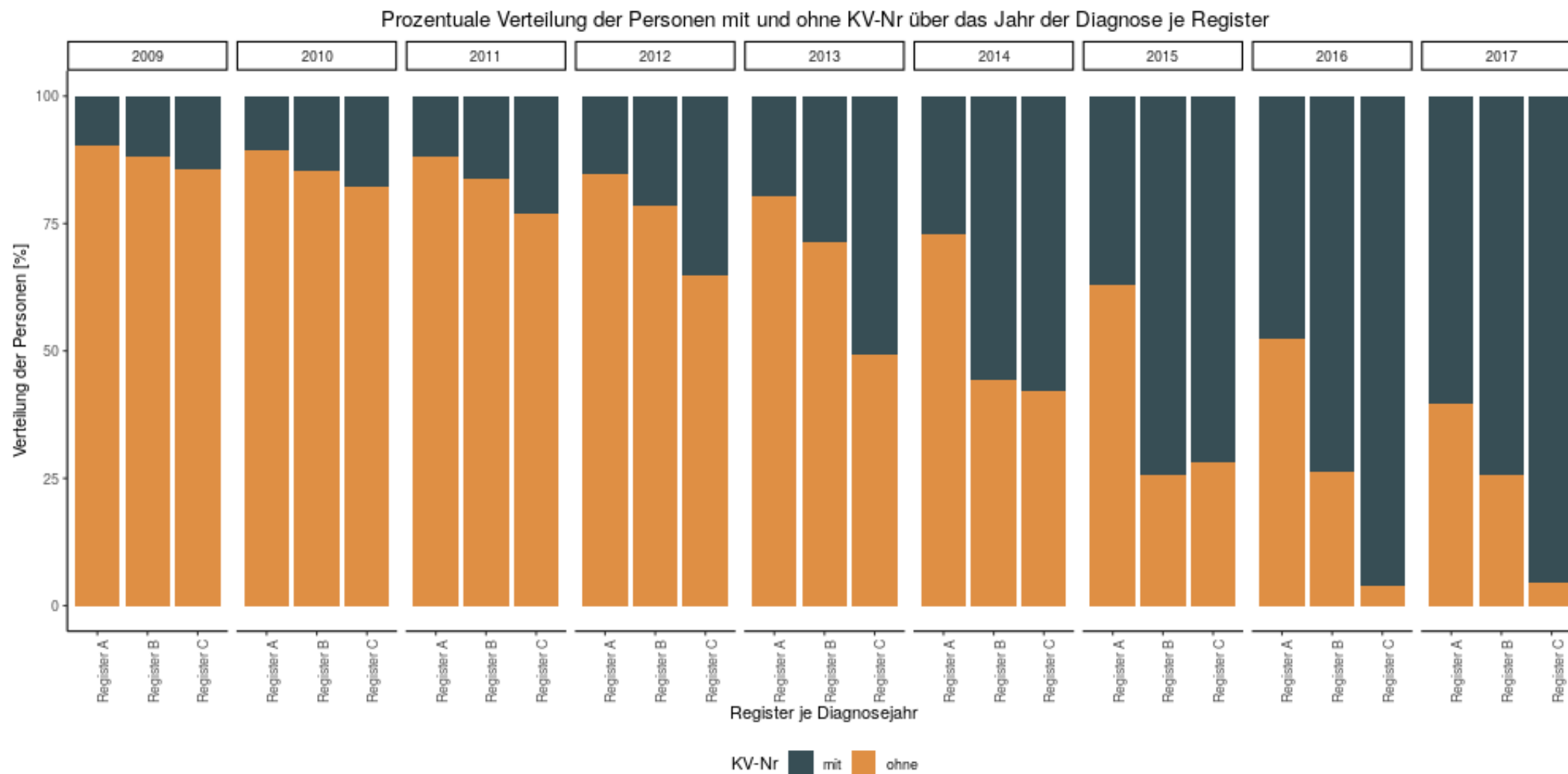
- Stärken der GKV-Daten
  - Mehr als 1000 Krankenhäuser, hohe Fallzahlen
  - Viele Confounder: Komorbiditäten, Klinikmerkmale...
  - Längsschnittdaten: komplette Krankheitsverläufe
  - abrechnungsrelevante Daten vollständig

+

- Stärken der KKR-Daten
  - Erkrankungsspezifische prognoserelevante Faktoren
  - Pathologischer Befund, Stadium (TNM), Rezidive
  - Allgemeinzustand des Patienten: ECOG, Karnofsky

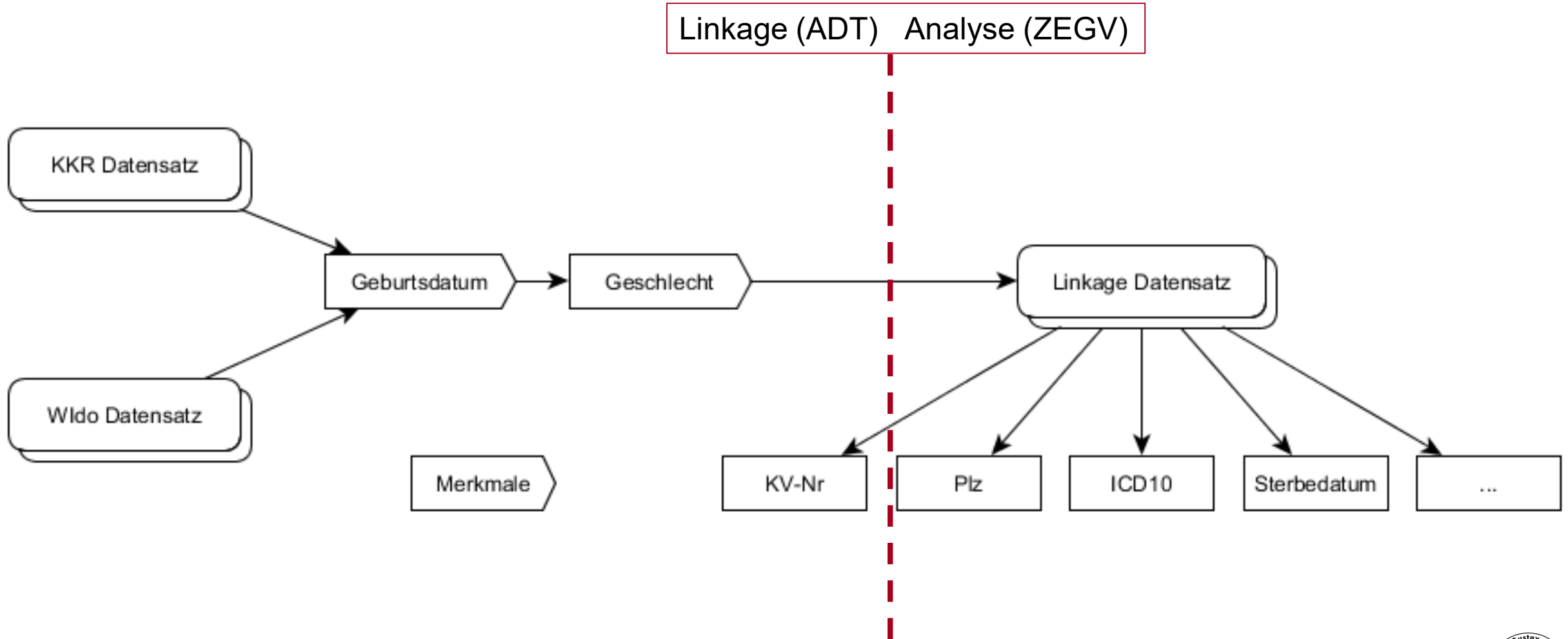


# Erfassung der KV-Nummer nach Register und Jahr



# Eigenschaften des Linkage-Datensatzes

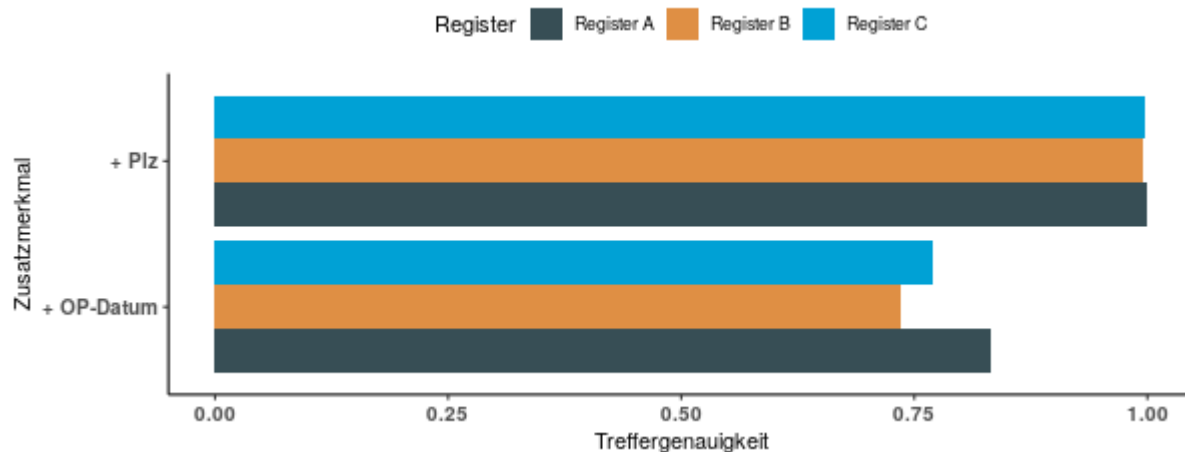
Linkage (ADT) Analyse (ZEGV)



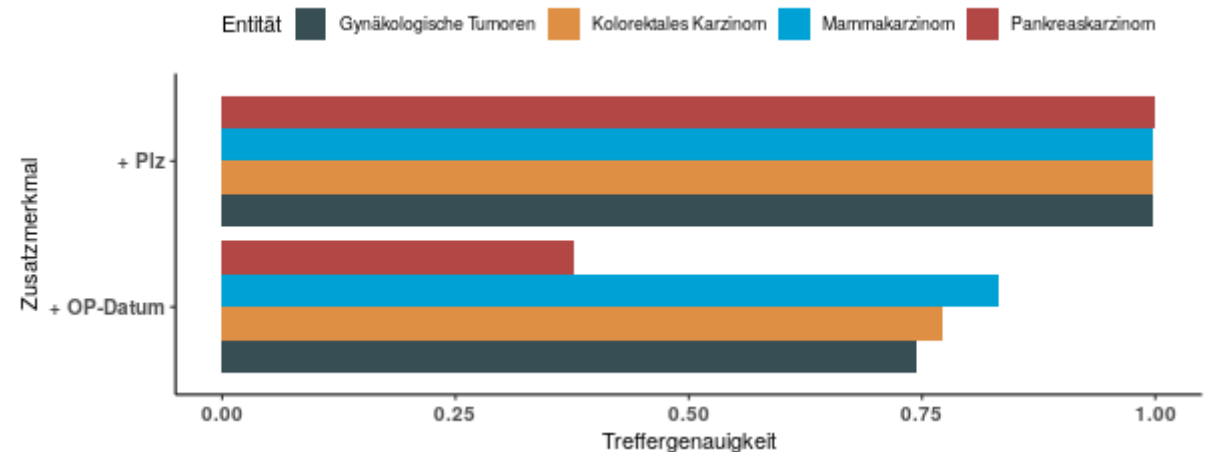
# Treffergenauigkeit Linkage bei Abgleich KV-Nr.

- Betrachtet wird der Datensatz mit gelinkter KV-Nummer als eineindeutiger Zuordnung -> Treffergenauigkeit entspricht Anteil der korrekt zugeordneten Personen
- Alle Daten sind gelinkt auf Alter + Geschlecht + Zusatzmerkmal

## Treffergenauigkeit nach Register



## Treffergenauigkeit Entität



# Schlussfolgerung zum Linkage

- **Linkage über die KV-Nummer** ist eineindeutig, unterliegt aber zeitlichen Entwicklungen
- **Linkage** über die Variablen Alter, Geschlecht, PLZ gelingt mit sehr hoher Zuverlässigkeit und ist zeitunabhängig
- Stärken der GKV-Daten
  - Mehr als 1000 Krankenhäuser, hohe Fallzahlen
  - Viele relevante Confounder, z.B. **Komorbiditäten**, Klinikmerkmale, (Größe, Träger...)
  - Längsschnittdaten mit kompletten Krankheitsverläufen
  - Erfassung (abrechnungsrelevanter) Daten vollständig
- Stärken der KKR-Daten
  - *Erkrankungsspezifische* prognoserelevante Faktoren:
  - Pathologischer Befund, Stadium (TNM), Rezidive
  - Allgemeinzustand des Patienten: ECOG, Karnofsky
- Ziel: Informationen über „das Beste aus beiden Welten“

# Ausblick und Verwertungspotenzial

## ■ Ausblick

- Ergänzung durch weitere Analysen auf Basis von KKR- sowie verknüpften AOK+KKR-Daten
- Methodische Absicherung: Sensitivitätsanalysen, Risikoadjustierung, Komorbiditäten, ...
- Wiederholung der Analysen für weitere Entitäten

## ■ Verwertungspotenzial

- Aussagen zur Wirksamkeit einer Zertifizierung
- Detaillierte Darstellung der Versorgungssituation von Tumorpatienten
- Langfristig
  - Verbesserung der Zusammenarbeit zwischen Primärversorgern, Sekundärversorgern und Zentren
  - Unterstützung der Patienten bei der Wahl der Kliniken
  - Basis für eine effektive und qualitativ hochwertige Patientenversorgung für die wichtigsten Tumorentitäten

## Verknüpfung von Krankenkassen- und Registerdaten: Projekt WiZen

Jochen Schmitt<sup>1</sup>, Olaf Schoffer<sup>1</sup>, Martin Rößler<sup>1</sup>, Christoph Bobeth<sup>1</sup>, Veronika Bierbaum<sup>1</sup>, Michael Gerken<sup>2</sup>, K. Kleihues-van Tol<sup>3</sup>, Patrik Dröge<sup>4</sup>, Thomas Ruhnke<sup>4</sup>, Andreas Klöss<sup>4</sup>, Christian Günster<sup>4</sup>, Monika Klinkhammer-Schalke<sup>2,3</sup>

1) ZEGV, 2) TZR, 3) ADT, 4) WiDO

8. DNVF-Forum Versorgungsforschung

12.5.2021



Gefördert durch



**Gemeinsamer  
Bundesausschuss**  
Innovationsausschuss

Förderkennzeichen: 01VSF17020