

# Pilotstudie zur Methodik und Machbarkeit eines Abgleichs der Daten der Epidemiologischen Krebsregister in Deutschland auf Mehrfachübermittlungen

K Kraywinkel<sup>1</sup>, T Bachteler<sup>2</sup>, B Barnes<sup>1</sup>, S Hentschel<sup>3</sup>, J Kieschke, S Luttmann<sup>5</sup>, M Meyer<sup>6</sup>, V Krieg<sup>7</sup>, A Richter<sup>8</sup>, G Seebauer<sup>9</sup>, B Steller<sup>10</sup>

<sup>1</sup> Zentrum für Krebsregisterdaten am Robert Koch-Institut, <sup>2</sup>Institut für Soziologie der Universität Duisburg-Essen

<sup>3-9</sup> Epidemiologische Krebsregister der Länder Hamburg, Niedersachsen, Bremen, Bayern, NRW, Schleswig-Holstein und Rheinland-Pfalz

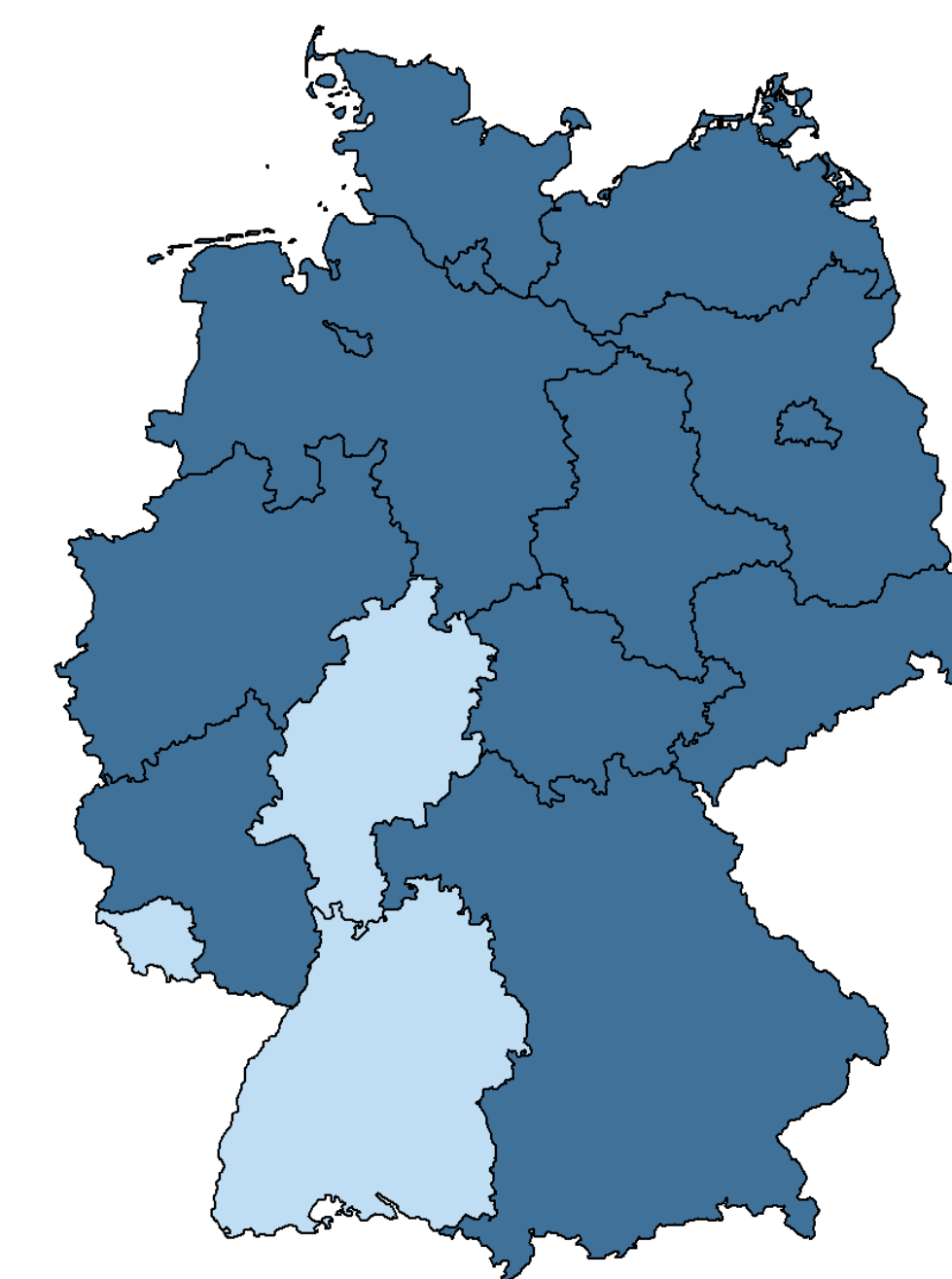
<sup>10</sup> Gemeinsames Krebsregister der Länder Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen-Anhalt, Sachsen und Thüringen

## Hintergrund

Das Bundeskrebsregisterdatengesetz (BKRG) sieht künftig einen regelmäßigen länderübergreifenden Datenabgleich zur Identifizierung von Mehrfachübermittlungen (MFÜ) desselben Erkrankungsfalls aus verschiedenen Landeskrebsregistern auf der Basis von Kontrollnummern vor. Bisher war unklar, in welchem Ausmaß bei der Zusammenführung der Daten aus den epidemiologischen Krebsregistern (EKR) im Zentrum für Krebsregisterdaten (ZfKD, früher: Dachdokumentation Krebs) am Robert Koch-Institut derartige Mehrfachübermittlungen anfallen und zu einer Überschätzung der Inzidenzraten führen. Zu erwarten wären Mehrfachübermittlungen in erster Linie bei Umzügen von Personen mit einer Krebserkrankung in die Bezugsregion eines anderen Krebsregisters, bei Vorhandensein mehrerer Wohnsitze für einen Patienten oder bei unrichtigen Angaben zum Wohnort durch einen Melder.

In einem vom Bundesministerium für Gesundheit (BMG) finanzierten Pilotprojekt wurde deshalb vom Zentrum für Krebsregisterdaten in Kooperation mit acht epidemiologischen Krebsregistern (Abbildung 1) diese Fragestellung erstmals systematisch untersucht. Das Projekt sollte zum einen eine erste Schätzung der Häufigkeit von Mehrfachübermittlungen liefern und zum anderen die Machbarkeit und Effizienz verschiedener Verfahren im Rahmen eines solchen Abgleichs prüfen.

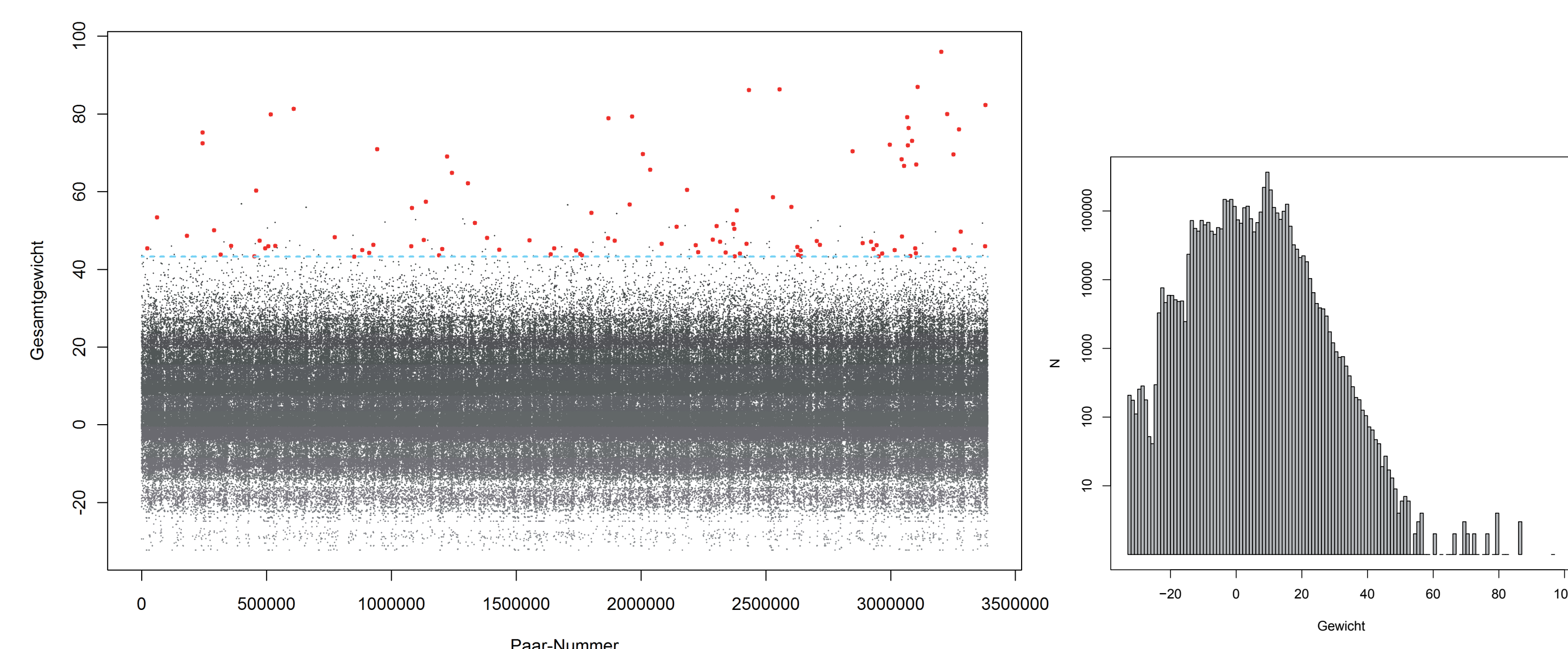
Abbildung 1:  
Teilnehmende Epidemiologische Krebsregister (dunkelblau)



## Methoden

Abgeglichen wurden jeweils die Neuerkrankungen aus 2007 mit den Bestandsdaten der übrigen Register ab 1998 bzw. dem Beginn der Registrierung. Bei dem in »R« programmierten probabilistischen Abgleich wurden anhand der als Kontrollnummern pseudonymisierten Angaben zu Namensbestandteilen sowie dem Geburtsdatum und Geschlecht Übereinstimmungsgewichte für alle Paare mit einem Mindestmaß an Übereinstimmungen berechnet (Abbildung 2). Dieses Verfahren entspricht im Prinzip der Zuordnung von Meldungen zu Personen (Record Linkage) in den epidemiologischen Krebsregistern.

Abbildung 2: Ergebnisse des probabilistischen Abgleichs (Neuerkrankungen aus Hamburg gegen Bestandsdaten aus Niedersachsen) mit Histogramm der Gewichtsverteilung (kleines Bild)

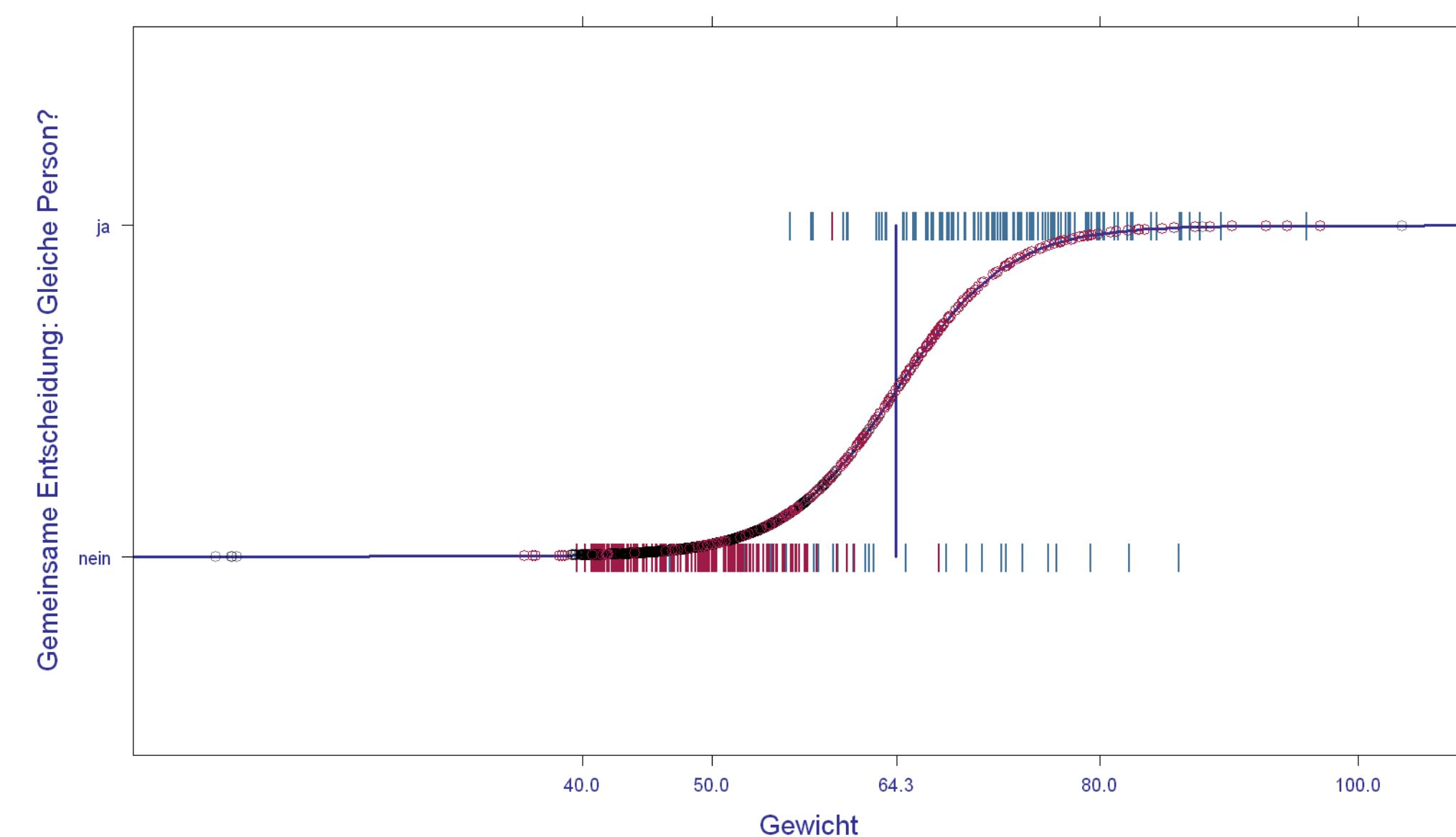


### Erläuterungen

Rote Punkte: Paare mit übereinstimmendem 1. Vor- und Nachnamen und Geburtsdatum

Anschließend wurde für eine Stichprobe von 2.500 potentiell übereinstimmenden Personenpaaren anhand zusätzlicher Informationen aus den Registern und nach gemeinsamer Abstimmung zwischen den betroffenen Registern entschieden, ob eine Übereinstimmung der Person und der Diagnose vorlag und welchem Register der Erkrankungsfall ggf. zuzuordnen war. Diese gemeinsame Entscheidung wurde als Goldstandard genutzt und diente als Grundlage für logistische Regressionsmodelle zur Schätzung der Übereinstimmungswahrscheinlichkeiten der nicht überprüften Fälle (Abbildung 3).

Abbildung 3: Logistisches Regressionsmodell zur Schätzung von Wahrscheinlichkeiten einer Mehrfacherfassung aus den Abgleichs mit dem EKR Niedersachsen



### Erläuterungen

Kurze vertikale Striche: Paare mit vorliegender gemeinsamer Entscheidung  
Rot: Paare mit Übereinstimmung des (jeweils ersten) Vor- und Nachnamens sowie des Geburtsdatums  
Blau: weitere Paare mit hohem Übereinstimmungsgewicht  
Blaue Kurve: Ergebnis des logistischen Regressionsmodells (Zusammenhang zwischen dem Gewicht und der Wahrscheinlichkeit einer gemeinsamen Entscheidung »gleiche Person«)  
Blaue vertikale Linie: Gewicht, bei dem die geschätzte Wahrscheinlichkeit einer gemeinsamen Entscheidung »gleiche Person« bei 50% liegt  
Kreise entlang der Regressionskurve: nicht überprüfte Paare

## Ergebnisse

Der geschätzte Anteil für die Inzidenz relevanter Mehrfachübermittlungen an den Neuerkrankungen eines Jahres betrug für die teilnehmenden Register zwischen 0,01% und 0,37% (Tabelle 1). Etwa 70% davon wurden zwischen benachbarten Registern bzw. Bundesländern identifiziert. In einigen Registern waren mehr als die Hälfte der bestätigten Fälle ursprünglich nur über die Todesbescheinigung gemeldet (DCO-Fälle). Eine modellbasierte Entscheidung im ZfKD erreichte eine Sensitivität von 86% und einen positiv prädiktiven Wert von 83%. Die Entscheidung alleine auf der Basis einer Übereinstimmung des ersten Vornamens und Namens sowie des Geburtsdatums erbrachte eine höhere Sensitivität (92%) bei geringerem ppV (78%). Bei etwa einem Drittel der identifizierten Mehrfachübermittlungen war aufgrund eines übereinstimmenden Erkrankungszeitpunkts eine Zuordnung des »zuständigen« Registers durch das ZfKD nicht möglich.

Tabelle 1: Geschätzte inzidenzrelevante Mehrfachübermittlungen (MFÜ) nach Register und Anteil an den Neuerkrankungen 2007

Register	Bestätigte MFÜ	Nach Hochrechnung geschätzte MFÜ	Anteil der geschätzten MFÜ an allen Neuerkrankungen
Bayern	6	60	0,08%
GKR	32	84	0,08%
Bremen	9	20	0,35%
Hamburg	24	33	0,29%
Niedersachsen	40	108	0,19%
NRW	45	145	0,12%
Rheinland-Pfalz	0	2	0,01%
Schleswig-Holstein	65	93	0,37%
<b>Gesamt</b>	<b>221</b>	<b>545</b>	<b>0,12%</b>

## Diskussion

Obwohl der Einfluss von Mehrfachübermittlungen von Erkrankungsfällen auf die Gesamtschätzung der Inzidenz von Krebserkrankungen in Deutschland relativ gering sein dürfte, erscheint eine Korrektur der Mehrzahl dieser Fälle im Sinne einer weiteren Verbesserung der Qualität krebsepidemiologischer Daten in Deutschland grundsätzlich sinnvoll. Zur Klärung und korrekten Zuordnung einer möglichen Mehrfachübermittlung im Einzelfall reichen jedoch die dem ZfKD zur Verfügung stehenden Informationen in der Regel nicht aus. Hierzu wäre in Zukunft auch in der Routine ein Informationsaustausch mit bzw. zwischen den Registern erforderlich. Die Bestätigung (und ggf. Zuordnung) einer Mehrfachübermittlung in den Registern gelingt vor allem dann, wenn mindestens in einem der beteiligten Register Informationen aus Einwohnermeldeämtern (über Zu- und Wegzüge, Erst- und Zweitwohnsitz) entweder bereits vorliegen oder eine entsprechende Abfragemöglichkeit besteht.