EKN Epidemiologisches Krebsregister Niedersachsen

Niedersächsisches Landesgesundheitsamt

Konzept zur Nutzung von Krebsregisterdaten in prospektiven Kohortenstudien im Epidemiologischen Krebsregister Niedersachsen (EKN)

J. Kieschke[1], U. Schlanstedt-Jahn[2], I. Urbschat[1], M. Hoopmann[3]

[1] Registerstelle des EKN, [2] Vertrauensstelle des EKN, [3] Niedersächsisches Landesgesundheitsamt

Einleitung

Die Durchführung von prospektiven Kohortenstudien konnte in Deutschland bislang nur selten mit vorhandenen Registerdaten gekoppelt werden. Einem solchen Vorgehen stehen oft datenschutzrechtliche Vorgaben entgegen, selbst bei einem allgemeinen Interesse, z.B. wenn nach Unglücksfällen eine Gefährdung der Bevölkerung abgeschätzt werden müsste. Auch wäre der Aufwand z.B. 50.000 Einwohner prospektiv zu verfolgen recht aufwändig gewesen.

Mit dem Kontrollnummernsystem der epidemiologischen Krebsregister steht ein Verfahren zur Verfügung, das zur Risikoabschätzung einen Registerabgleich einer potentiell belasteten breiten Bevölkerungsgruppe im Vergleich zu einer Kontrollgruppe mit akzeptablem Aufwand ermöglicht. Für Niedersachsen gilt, dass im Krebsregister mit Einwilligung auch Daten von nicht an Krebs erkrankten Personen erfasst und verarbeitet werden dürfen. Bei einer größeren Kohorte mit sämtlichen Einwohnern mehrerer Kleinstädte wäre die Einholung entsprechender Einwilligungen aber kaum praktikabel, eine unvollständige Teilnahmerate würde die Aussagekraft jedoch deutlich einschränken. Daher wurde ein Konzept erarbeitet, das als eine Art "ökologischer Kohortenansatz" unter Einhaltung datenschutzrechlichter Belange auch ohne vorliegende Einwilligung einen Abgleich mit dem EKN auf pseudonymisierter Ebene ermöglichen soll.

Material + Methode

Das Kontrollnummernkonzept

Um einen Personenabgleich auf pseudonymisierter Ebene zu ermöglichen, wird auf das Kontrollnummernkonzept der epidemiologischen Krebsregister zurückgegriffen [1]. Dabei werden zur Erhöhung der Fehlertoleranz die Personen identifizierenden Angaben in ihrer Schreibweise standardisiert, wozu auch die Bildung phonetischer Codes für die Namensangaben gehört. Anschließend werden die Angaben in eine bestimmte Anzahl von Einzelattributwerten zerlegt und diese jeweils per Einwegverschlüsselung (MD5) in nicht dechiffrierbare "Pseudonyme", die sog. Kontrollnummern, umgewandelt. Um Deanonymisierungsversuche mittels Probeverschlüsselungen zu verhindern, erfolgt nachträglich eine symmetrische Überverschlüsselung (IDEA).

Zur Zusammenführung korrespondierender Meldungen wird im EKN ein probabilistisches Record Linkage Verfahren angewandt. Dabei wird entsprechend des ermittelten jeweiligen Übereinstimmungsmusters der verschiedenen Einzelattribute (z.B. Nachname unterschiedlich, Vorname und Geburtstag übereinstimmend ...) ein Gewicht als Maß für die Wahrscheinlichkeit der Zusammengehörigkeit zweier Datensätze vergeben. Mit demGewicht wird versucht, die Subgruppe von zusammengehörenden Datensätzen von den nicht zusammengehörenden Datensätzen zu trennen.

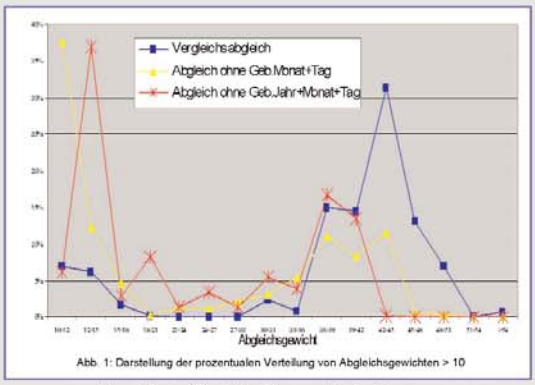

Abb. 1: Darstellung der prozentualen Verteilung von Abgleichsgewichten > 10

Abbildung 1 zeigt die prozentuale Verteilung von ermittelten Gewichten. Die blaue Kurve zeigt das Ergebnis eines durchschnittlichen Abgleichs. Es ergeben sich zwei Gipfel der Verteilungskurven, wobei der linke nur zum Teil dargestellt wird. Im Prinzip repräsentiert die rechte Verteilungskurve die zusammengehörenden Meldungen und die linke die nicht zusammengehörende. Wie gut die Trennschärfe der beiden Kurven ist hängt vor allem von der Qualität der vorliegenden Daten ab. So zeigen die gelbe und die rote Kurve Ergebnisse, bei denen nur Teile oder gar keine Informationen zum Geburtsdatum einbezogen wurden. Mit geringerem Informationsgehalt wird erwartungsgemäß die Trennung der beiden Gipfel unschärfer.

Während bei eindeutig getrennten Kurvenverläufen im Prinzip ein vollständig automatischer Abgleich möglich wäre, wird es in der Praxis einen Überschneidungsbereich geben, in der die Zuordnungen interaktiv getroffen und ggf. sogar nach recherchiert werden sollte. Da nur diese Fälle Personalresourcen beim Abgleich erfordern, sind auch große Datenmengen mittels dieses Verfahrens mit relativ geringem Aufwand bearbeitbar. Derzeit werden im EKN pro Quartal etwa 20.000 Neumeldungen mit dieser Methode abgeglichen.

Ergebnisse: Anwendungsbeispiel

Studienverlauf

Im Anwendungsbeispiel sollen als Teiluntersuchung zur Gesundheitsfolgenabschätzung nach einem Gefahrgutunfall prospektiv die Inzidenz- und Mortalitätsraten in der betroffenen Region und in mehreren Vergleichsregionen ermittelt werden. Abbildung 2 zeigt eine schematische Darstellung des Studienablaufs:

Das Einwohnermeldeamt erstellt eine Datei mit den Personen identifizierenden Angaben der Einwohner zum fraglichen Stichtag (Name, Vorname, Geburtsdatum, Wohnort als Gemeindekennziffer und Geschlecht). Für spätere Vergleichszeitpunkte (z.B. 5 Jahre nach dem Unfall) wird eine entsprechende Vergleichsdatei mit den aktuellen Einwohnerangaben erhoben.

Aus diesen Angaben erstellt die Vertrauensstelle die Kontrollnummern, wobei Geburtsmonat und –jahr, Geschlecht und Wohnort als Gemeindekennziffer unverschlüsselt erhalten bleiben. Über diese Kontrollnummern wird ein sogenannter Projektschlüssel gelegt. Die Vertrauensstelle vernichtet alle von den Meldebehörden über das NLGA für die Verschlüsselung erhaltenen Daten unmittelbar nach der Verschlüsselung, sodass diese allenfalls vorübergehend in der Vertrauensstelle vorhanden sind.

Werden die Daten im EKN nicht aktuell für Auswertungsfragen gebraucht, lagern sie in der Studienzentrale im NLGA, wobei die Daten mit einem PGP-Schlüssel zusätzlich überverschlüsselt werden. Damit ist gewährleistet, dass keine Stelle außerhalb des Krebsregisters die Kontrollnummern lesen und verwenden kann.

Das NLGA übergibt diese „Schlüsseldatei" und die aktuellen Einwohnerdaten der Meldebehörde der Vertrauensstelle im 5-jährigen Abstand. Die Vertrauensstelle leitet diese nach Entschlüsselung (Projektschlüssel, PGP-Verschlüsselung) an die Registerstelle zwecks Abgleich weiter. Als Ergebnis würde die Registerstelle für die betroffene Region und die ausgewählten Vergleichsregionen - in 5-Jahres-Altersgruppen zusammengefasst - die altersspezifischen Krebserkrankungsraten für beide Geschlechter getrennt berechnen. Um durch Wanderungsbewegungen entstandene Lost-Of-Follow-Up-Fälle zu erkennen und ausklammern zu können, werden Unterauswertungen auf den Bevölkerungsanteil begrenzt werden, der sowohl zum Stichtag des Unfalls wie auch zum Auswertungszeitpunkt in der Gemeinde gewohnt hat. Da als Ergebnis nur altersspezifische Krebserkrankungsraten an die Studienzentrale übermittelt werden, ist eine Deanonymisierung dort nicht möglich.

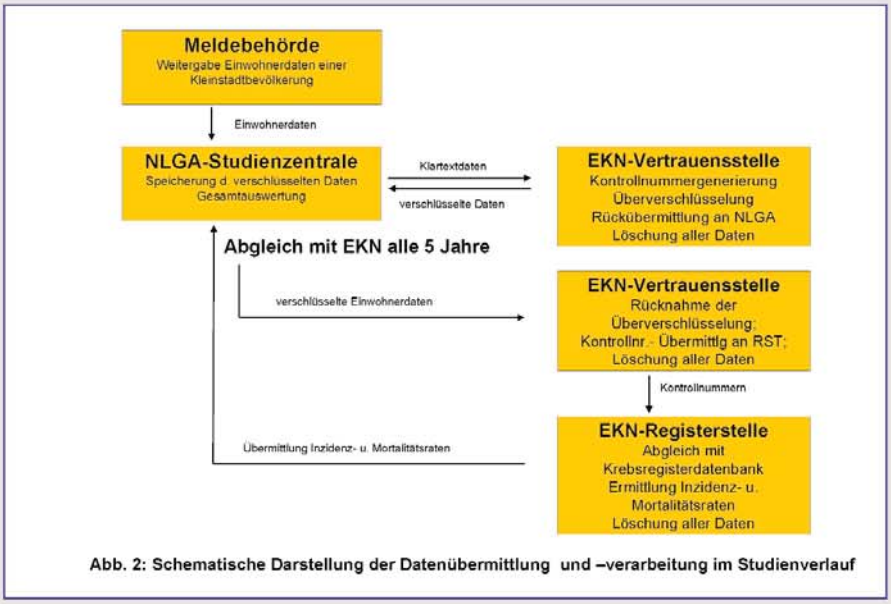

Abb. 2: Schematische Darstellung der Datenübermittlung und –verarbeitung im Studienverlauf

Zusammenfassung und Diskussion

Ausgangspunkt der Überlegungen ist ein Gefahrgutunfall in Niedersachsen. Zur Klärung der Gefährdung soll in einer klassischen individuellen Kohortenbetrachtung bei mutmaßlich hoch exponierten Personen (z.B. Einsatzkräfte) und besorgten Bürgern ("Selbstmelder") das Erkrankungsrisiko in Abhängigkeit von der Expositionshöhe (Biomonitoring, Ausbreitungsrechnungen, Fragebögen) abgeschätzt werden. Um eine Aussage zur Risikoerhöhung bei der kaum betroffenen "durchschnittlichen" Wohnbevölkerung zu ermöglichen, wurde ein bimodales Krebsverfolgungskonzept angedacht, bei dem ergänzend die Inzidenz- und Mortalitätsraten für die betroffene Region und von Vergleichsregionen prospektiv ermittelt werden sollen. Durch Nutzung des Kontrollnummernkonzeptes ist dabei die Durchführung auch ohne Einwilligung möglich. Hierbei ist eine höhere Fallzahl als im individuellen Kohortenansatz erreichbar. Beide Module ergänzen sich somit.

Nachdem die entsprechenden Absprachen mit dem Datenschutzbeauftragen abgeschlossen sind, wird demnächst die Studie beginnen. Eine Übertragbarkeit dieses Konzeptes auch für den Abgleich mit anderen Datenquellen ist zu diskutieren. So käme es auch in Betracht beim Abgleich zwischen Krebsregistern und Teilnehmerinnen am Mammographie-Screening, um die Rate an Intervallkarzinomen zu bestimmen.

Literatur

[1] Appelrath H.-J., Michaelis J., Schmidtmann I., Thoben W.: Empfehlung an die Bundesländer zur technischen Umsetzung der Verfahrensweisen gemäß Gesetz über Krebsregister (KRG), Informatik Biometrie und Epidemiologie in Medizin und Biologie 27 (2) 1996

Kontakt: www-krebsregister-niedersachsen.de