

Vernetzte interdisziplinäre Nutzung von Routinedaten aus der Rehabilitation

Kaluscha R¹, Muche R², Jacobi E^{1,3}

¹Forschungsinstitut für Rehabilitationsmedizin an der Universität Ulm

²Abteilung Biometrie und Medizinische Dokumentation, Universität Ulm

³Universitätsrehabilitationsklinik Bad Wurzach
rainer.kaluscha@uni-ulm.de

Einleitung und Fragestellung

Im rehabilitationswissenschaftlichen Forschungsverbund Ulm gelang mit der Forschungsdatenbank „Patientenkonto“ die Vernetzung von Versorgung, Kostenträger und Forschung. Im „Patientenkonto“ befinden sich z.Z. 100.000 anonymisierte Entlassungsberichte aus 6 Reha-Kliniken, die sowohl strukturierte Daten wie auch Freitexte umfassen [1]. Die Deutsche Rentenversicherung Baden-Württemberg übernimmt die Anonymisierung und ergänzt aus ihrem Versicherterkonto follow-up-Daten.

In Zusammenarbeit von Informatikern, Biometrikern und Rehabilitationsmedizinern wurde untersucht, ob bereits aus diesen Routinedaten Prognosemodelle, z.B. zur Vorhersage vorzeitiger Berentung, entwickelt werden können.

Material und Methoden

In den beteiligten Reha-Kliniken wird routinemäßig der vereinheitlichte Entlassungsbericht der Deutschen Rentenversicherung erstellt und zusammen mit einem Fragebogen zur Qualitätssicherung für Arzt und Rehabilitand elektronisch an den Kostenträger, d.h. in diesem Fall die Deutsche Rentenversicherung Baden-Württemberg, übermittelt. Dieser ergänzt aus seinem Routinedatenbestand im Versicherterkonto Angaben zu abgelehnten und bewilligten Rentenansprüchen, Arbeitsunfähigkeitszeiten von mehr als 6 Wochen sowie ggfs. das Sterbedatum und übermittelt die Daten anonymisiert in die Forschungsdatenbank. Dort werden diese Daten weiter aufbereitet, indem z.B. mit computerlinguistischen Methoden Informationen aus den Freitexten extrahiert werden [2]. So entsteht etwa aus einschlägigen textuellen Formulierungen wie „Kein Genussmittelabusus“ eine binäre Variable „Raucher“, die dann in die statistische Modellierung einbezogen werden kann [3].

Da die Datensätze über lange Zeit gesammelt wurden, ist eine Versionierung zwingend: so erfolgte während der Erhebung die Umstellung der Diagnosekodierung von ICD-9 auf ICD-10; gesetzliche Änderungen erforderten 2001 eine Anpassung des Entlassungsberichtes, etc.

Für jede Auswertung muss daher neben üblichen Einschlusskriterien wie Alter, Geschlecht und Krankheitsbild auch festgelegt werden, welche Daten in welcher Kodierung vorhanden sein müssen, ob ggfs. umkodiert und gepooled werden kann, etc. Dabei sind sowohl inhaltliche wie auch methodische Aspekte zu berücksichtigen.

Ergebnisse

Unter Nutzung der Routinedaten gelang in enger Zusammenarbeit aller Beteiligten bereits die Erstellung einiger Prognosemodelle zur Vorhersage vorzeitiger Berentung und zur Identifikation von Risikogruppen [3,4,5]. Die Prognosegüte wurde dabei nach logistischer Regressionsmodellierung u.a. anhand der AUC nach ROC-Analyse nach einer Modellvalidierung mit Bootstrap-Methoden angegeben [4]. Diese ist mit Werten zwischen 0.85 und 0.9 für die Nutzung von Routinedaten, die nicht für eine solche Fragestellung erhoben wurden, als sehr gut zu bezeichnen. In diese Modellierung flossen auch Merkmale ein, die sich nur im Freitext finden, z.B. die Reha-Motivation, Familienstand oder evtl. Migrationshintergründe. Die entsprechenden grenzwert-abhängigen Kenngrößen wie Sensitivität und Spezifität liegen dabei zwischen 80 und 85%.

Diskussion

Da es sich um Routinedaten handelt, ist eine gute Kenntnis des Umfeldes, in dem die Daten entstanden sind, Voraussetzung für ihre adäquate Nutzung und eine Einschätzung der Datenqualität. Durch enge Zusammenarbeit und intensiven Austausch der unterschiedlichen Datenproduzenten und -nutzer kann dies gelingen. So erfordert etwa die Entscheidung, ob zur Fallzahlerhöhung eine Umkodierung aus einer anderen Version des Datensatzes sinnvoll ist, sowohl medizinisch-inhaltliches wie auch statistisch-methodisches Wissen. Auch der Umgang mit fehlenden Werten, die bei solchen Routinedaten natürlich vorkommen, muss differenziert erfolgen: in manchen Fällen dokumentiert der Praktiker unauffällige Befunde nicht gesondert und die Annahme einer unauffälligen Ausprägung wäre plausibel, in anderen Fällen ist tatsächlich von fehlenden Angaben auszugehen.

Hilfreich ist es auch, wenn bei Datenproduzenten und -nutzern die beteiligten Disziplinen komplementär vertreten sind: so gestaltet es sich z.B. als schwierig, wenn der Biometriker aus der Forschung eine sozialmedizinische Beurteilung eines Reha-Mediziners aus der Klinik interpretieren oder wenn der Mediziner mit der EDV-Abteilung des Kostenträgers die Realisierung des Datentransfers klären soll. Hier können dann z.B. Informatiker aus der Forschung als „Dolmetscher“ helfen. So ist die Vernetzung der verschiedenen Disziplinen notwendig.

Zusammenfassend hat sich das vorgestellte Konzept bewährt und bereits zu interessanten Ergebnissen geführt. Der Datenbestand wird auch künftig für weitere Fragestellungen genutzt werden.

Literatur

- [1] Kaluscha R, Jacobi E: Erschließung von Routinedaten mittels Vernetzung und Computerlinguistik. *Inf. Biom. Epidemiol. Med. Biol.* 34/3 (2003).
- [2] Kaluscha R: Informationsgewinnung aus Freitexten in der Rehabilitationsmedizin. Dissertation: Medizinische Fakultät, Universität Ulm (2005).
Online: <http://vts.uni-ulm.de/doc.asp?id=5265>
- [3] Wolf U: Drohende Frühberentung wegen Minderung der Erwerbsfähigkeit nach stationärer Rehabilitation bei Patienten mit Affektionen des Bewegungsapparates. Masterarbeit: Public Health, Medizinische Fakultät, Universität Ulm (2005). Online: <http://vts.uni-ulm.de/doc.asp?id=5447>
- [4] Muche R, Ring Ch, Ziegler Ch: Entwicklung und Validierung von Prognosemodellen auf Basis der logistischen Regression. Shaker Verlag, Aachen (2005).
- [5] Moßhammer D: Logistische Regressionsmodellierung anhand routinemäßig erhobener Daten zur Vorhersage drohender Erwerbsminderung bei Patienten der LVA Baden-Württemberg, die wegen Adipositas stationär rehabilitiert wurden. Magisterarbeit Public Health, Universität Ulm (2004).