

Medizinischer Erkenntnisgewinn aus vernetzten Datenbanken

Fritz P, Klenk S, Dippon J, Kleinhans A, Albu P, Friedel G, Zakim A, Ott G, Braun N,
Thon P, Winter S, Brinckmann F, Alscher D

Institut für Digitale Medizin Stuttgart, San Francisco

Dr. Peter Fritz Institut für Digitale Medizin,
D-70192 Stuttgart,
Am Kriegsbergsturm 44,
peter.fritz@idem-foundation.org

Abstract: Bei einer Wissensverdopplung der Medizin von unter 10 Jahren ist ein medizinischer Erkenntniszuwachs zwangsläufig auch auf die Analyse historischer Daten angewiesen. Die dem vorliegenden Beitrag zugrunde liegenden Datenbanken umfassen die Krankheitsbilder des Mammakarzinoms, Bronchialkarzinoms, kolorektalen Karzinoms, der verkapselnden Peritonealsklerose und der Riesenzellarteriitis. Folgende Analysen wurden mit diesen Datenbanken ausgeführt: univariate und multivariate Überlebensanalysen, Erstellung und Vergleich von Risikoprofilen, Datamining, Suche nach ähnlichen und identischen Fällen. Beispiele von Erkenntnissen aus der Analyse retrospektiver Daten sind: (1) die histologische Subklassifikation des Mammakarzinoms erlaubt keine Vorhersage des Überlebens, (2) verschleimende und nicht verschleimende kolorektale Karzinome unterscheiden sich hinsichtlich des Überlebens abhängig von der Lokalisation des Tumors (3) die Bestimmung des Glukokortikoidrezeptors erlaubt eine grobe Vorhersage des Therapieansprechens bei der Riesenzellarteriitis (4) die histologische Bewertung erlaubt keine Vorhersage des weiteren Verlaufs bei der sklerosierenden Peritonitis. Den unbestreitbaren Vorteilen der Analyse retrospektiver Daten stehen erhebliche informationstechnologische Probleme gegenüber: (1) meist ungelöste Schnittstellenprobleme (2) die Daten liegen häufig ungeordnet als String Variable vor (3) die Daten werden oft einer wissenschaftlichehn Auswertung trotz Anonymisierung vorenthalten (4) Fehlen von Analyseprogrammen für Nicht-Statistiker

1 Einleitung

Daten- und Gewebe-banken sind heute unverzichtbarer Bestandteil der Alltagsmedizin und der medizinischen Forschung. Ihrer sinnvollen Nutzung durch praktizierende Ärzte und Wissenschaftler ohne vertiefte Kenntnisse in Informationstechnologie und/oder Statistik stehen gravierende Widerstände gegenüber: ethische Bedenken, Schnittstellenprobleme, unzureichenden Datenpflege, rudimentäre Kenntnisse in den technischen und statistischen Grunddisziplinen. Gegenstand der vorliegenden Publikation sind Untersuchungen an Daten- und Gewebebanken des Robert Bosch Krankenhaus und des Instituts für Klinische Pharmakologie. In Zusammenarbeit mit Mathematikern und Informatiker wurden Methoden entwickelt, die einen Teil der oben beschriebenen Probleme verringern. Beispiele der damit erzielbaren Erkenntnisse aus Forschung und Praxis werden angeführt.

2 Material und Methoden

Die eingesetzten Datenbanken wurden seit 1986 am Institut für Klinischen Pharmakologie und am Pathologischen Institut des Robert Bosch Krankenhaus im Rahmen zahlreicher Forschungsprogramme aufgebaut und gepflegt.

2.1 Datenbanken

Die eingesetzten Datenbanken umfassten Datensammlungen zum Mammakarziom, Bronchialkarzinom, Erkrankungen des rheumatologischen Formenkreises und nephrologische Erkrankungen (EPS verkapselnde Peritonealsklerose). Die Datenbank über Mammakarzinome am Robert Bosch Krankenhaus wurde in die OSP Datenbank für Mammakarzinome eingegliedert. In dieser Datenbank wurde eine systematische Erfassung fast aller Mammakarzinome im Grossraum Stuttgart seit 1989 vorgenommen. Für alle Daten- und Gewebebanken wurden Anamnesedaten, Befunddaten, Labordaten, Daten bildgebender Verfahren, soweit erforderlich und morphologische Daten erhoben. Bei einigen Gewebebanken wie dem Mammakarzinom wurden auch molekularbiologische Daten erfasst [RSA08]. Beim Bronchialkarzinom wurden physikalische Meßgrößen aus der Spirometrie zusätzlich mit eingebunden wie etwa die Vitalkapazität oder der Einsekundenwert der Expiration.

2.2 Gewebebanken

Zusätzlich ist in der Teilmenge der Erkrankungsfälle mit invasivem Mammakarzinom am Robert Bosch Krankenhaus Formalin fixiertes Gewebe asserviert. Gleiches trifft für die Fälle mit Bronchialkarzinom zu. Für die nicht onkologischen Krankheitsentitäten der Riesenzellerarteriitis und der verkapselnden Peritonealsklerose (EPS) wurden ebenfalls Daten- und Gewebebanken aufgebaut.

2.3 Verwendete IT Techniken und Statistikprogramme:

Medizinisches Diagnosesystem CLEOS [ZBF08], webbasierte Datenbankprogramme (SQL Server, PHP et. al.), Entwicklungstools (Excel, MS Access et. al.), Data Mining System OCDM [KDF09] sowie verschiedene Statistiksysteme (Hugin Expert, Stratgraph et. al.).

2.4 Biologische Verfahren

Es wurden immunhistochemische Verfahren sowie molekularbiologische Verfahren wie die Austestung von SNP (single nucleotid polymorphism) eingesetzt [BFR09]. Im einzelnen erfolgte ein Nachweis Kernrezeptoren wie dem Östrogenrezeptor, Progesteronrezeptor, Vitamin-D Rezeptor, Glukokortikoidrezeptor. Weiterhin lag von allen im RBK bearbeiteten Krankheitsfällen eine H&E Färbung mit Beurteilung durch einen Pathologen vor.

2.4 Ethikvoten

Für alle aufgebauten Datenbanken wurden Ethikvoten erhoben. Zusätzlich wurde bei Klinikaufnahme eine Einverständniserklärung des Patienten erhoben, ob er Einwände gegen eine wissenschaftliche Auswertung seiner Daten und die Asservierung seines Paraffinblocks für spätere wissenschaftliche Auswertungen hat. Alle Auswertungen wurden für das nicht ärztliche Personal anonymisiert. Für sämtliche Mitarbeiter im ärztlichen und nicht-ärztlichen Bereich liegt zusätzlich eine Schweigeverpflichtung vor.

3 Ergebnisse

3.1 Ergebnisse mit praktischer Bedeutung für den Patienten

Bei onkologischen Fragestellungen sind es nur wenige, dafür aber existentielle Fragen die der Patient an die Medizin, vertreten durch seinen behandelnden Arzt, stellt. Diese sind: (1) wie sind meine Überlebenseaussichten, (2) was ist die optimale operative Therapie, (3) brauche ich eine Chemotherapie und wenn ja, welche, (4) Bin ich an der Institution, an der ich behandelt werde gut aufgehoben.

- Zu allen 4 Grundfragen kann die IT Technologie mit Online Nutzung wesentliche Beiträge leisten, wobei der optimale Nutzen dann entsteht, wenn eine Synthese zwischen Arzt, Patient und IT Technologie erreicht werden kann. Die vier anstehenden Patientenfragen und ihre möglichen IT Antworten sind erwartungsgemäß mit Problemen behaftet. Der Fall eines Mammakarzinoms mit schlechter Prognose, die vom Patienten aus dem Internet abgefragt wird, und zum Suizid der Patientin wegen vermeintlicher Hoffungslosigkeit führt, ist das Horrorszenarium jedes Arztes bei der Entwicklung von IT Systemen. Dies führt zu unterschiedlichen technologischen Vorgehensweisen: Beim OCDM System kann nur der im OSP tätige Arzt nach interner Passwortvergabe tätig werden. Die Finprog Daten des finnischen Krebsregister sind im Internet online abrufbar. Das Spannungsfeld zwischen individueller voller Verfügbarkeit des Patienten über seine Daten und Überforderung des Patienten und zunehmend auch des Arztes durch Datenfülle und Komplexität der Daten wird alle IT Lösungen der nächsten Jahrzehnte begleiten. Ein positives Beispiel der Nutzung eines solchen Prognoseprogramms (OCDM System im OSP Stuttgart) ergibt für eine Patientin mit kleinem Mammakarzinom unter 2 cm, G1 Tumor und positivem Östrogenrezeptor und Progesteronrezeptor ein krankheitsbedingtes Überleben von nahezu 100%).

3.2 Ergebnisse mit Bedeutung für den Arzt

Wie in Abb. 1 und 2 anschaulich erkennbar, ist das Stadium nicht aber der histologische Subtyp ein Prognosefaktor für das Mammakarzinom. Die Häufigkeit des Lokalrezidivs beim Mammakarzinom hängt von der Randfreiheit (R Status), nicht aber vom histologischen Subtyp ab. Ein dissolut wachsendes Tumorstadium erhöht nicht die Wahrscheinlichkeit eines lokalen Tumorrezidivs. Die Größe des Sicherheitsabstands beim invasiven Mammakarzinom muß nicht in Abhängigkeit von der Histologie modifiziert werden. Beim Bronchialkarzinom ist das gleichzeitige Vorkommen eines Diabetes kein offenkundiger zusätzlicher Risikofaktor, wohl aber der Raucherstatus.

3.3 Ergebnisse mit wissenschaftlichem Wert

Bei der Riesenzellarteriitis scheint die Menge des GR eine Aussage über das Ansprechen der Cortikoidtherapie zu geben. Die Fälle einer Riesenzellarteriitis mit reichlicher GR Expression sprechen in einer ersten Pilotstudie besser auf Kortikosteroidrezeptur an als Fälle mit niedriger Expression. Diese Ergebnisse fallen in die Rubrik der Hypothesengeneration aus retrospektiv verfügbaren Datenbanken. Eine hohe Zahl von CD 38 positiven Monocyten und eine hohe Expression von Glukokortikoidrezeptoren erwiesen sich als Variablen, die eine komplette Remission der Erkrankung unter Therapie mit Kortikosteroiden voraussagte. Die Bestätigung dieser Arbeitshypothese setzt gut gepflegte Daten- und Gewebe-banken und statistische Auswerteprogramme voraus.

Beim Krankheitsbild der sklerosierenden encapsulierenden Peritonitis erlauben Patienten und Gewebebanken eine kontinuierliche Überwachung der Fragestellung, inwieweit Komorbiditäten oder Therapiemodalitäten (Perfusionslösungen) Einfluß auf die Entwicklung dieser gefürchteten Komplikation der Peritonealdialyse haben. Wir konnten zusätzlich an einer ESP Daten- und Gewebebank nachweisen, dass es zu unerwarteter Expression von Östrogen-, Progesteron-, und Androgen-rezeptoren in der Peritonealmembran kommt. Ob der Expression dieser nukleären Rezeptoren eine Bedeutung zukommt oder nicht, kann nur mit den Mitteln von Langzeituntersuchungen erfolgen. Auf Grund der Seltenheit des Krankheitsbilds der ESP sind prospektive Studien kaum durchführbar und nicht finanzierbar.

3.4 Problem bei der Nutzung auf Datenebene

Ungelöste Probleme, die in die Alltagsnutzung von Datenbanken hineinspielen, sind die Fragen nach Eigentumsverhältnissen solcher Daten- und Gewebebanken. Welche Person und welche Institution hat Zugriff zu den Datenbanken? Ein weiteres Spannungsfeld ist die Anonymisierung solcher Datenbanken, welches im Einzelfall dazu führen kann, dass therapeutisch relevante Forschungsergebnisse dem Merkmalsträger nicht mehr zuteil werden können.

3.5 Problem bei der Nutzung auf IT Ebene

Bei den von uns geführten Datenbanken lagen die Hauptschwierigkeiten in zahllosen Schnittstellenproblemen. Als praktisches Beispiel für die enormen Schwierigkeiten bei der Zusammenführung von Datenbanken sei die Verschlüsselung von Datumsfunktionen und ihre Umwandlung in numerische Variablen angeführt. XML Technologien in der Schnittstelle zwischen einzelnen Datenbanken sind im wissenschaftlichen Bereich derzeit die Ausnahme. Allein in dem von uns genutzten Arbeitsbereich kamen 14 verschiedene Datenbanksysteme und Auswertungssysteme zum Einsatz

4 Diskussion

Die ausserordentliche Leistungsfähigkeit von Gewebe- und Datenbanken im medizinischen Alltag und in der medizinischen Forschung wurde beispielhaft in 5 ganz verschiedenen Subdisziplinen der Medizin dargestellt. Alle 5 Beispiele zeichnen sich dadurch aus, dass die Untersuchungsergebnisse praktische Konsequenzen im klinischen Alltag nach sich ziehen oder in die Rubrik von Hypothesen generierenden wissenschaftliche Ergebnisse fallen. Alle 5 Beispielen gemeinsam ist auch die Verwendung nicht-kommerzieller Datenbanken und der damit verbundenen Einschränkungen. Den unbestreitbaren Vorteilen stehen auf drei Ebenen Probleme gegenüber: (1) Ethische Beschränkungen bei der Nutzung dieser Gewebe- und Datenbanken, (2)Schnittstellenproblem, (3) Probleme der Datenpflege

Die ethischen Aspekte der Nutzung von Gewebe- und Datenbanken sind im Klinikalltag auch dann nicht vollständig gelöst, wenn für alle Anträge Ethikvoten und Einverständniserklärungen vorliegen. Die dann immer noch auftretenden Problematik wie etwa Eigentumsverhältnisse an Datenbanken sind am besten im Zusammenarbeit mit Selbsthilfegruppen lösbar (siehe Beitrag Alsmeier). Hier kann am besten eine Risikoabschätzung zwischen möglichem Schaden solcher Datenbanken und einem individuellen Nutzen für den Erkrankten erfolgen unter Einbeziehung des Umstandes, dass bei schweren tödlich verlaufenden Erkrankungen ein einzugehendes Risiko des Datenmissbrauchs geringer wiegt als bei weniger schweren Krankheitsverläufen.

Träger, die wissenschaftliche Ansätze mittels Gewebe- und Datenbanken über Tumorerkrankungen und/oder Patientenselbsthilfegruppen verfolgen, verfügen in der Regel über geringe Geldmittel, die eine kommerzielle Lösung von Schnittstellenproblemen nahezu unmöglich erscheinen lassen. Die Bereitstellung von IT Software Lösungen für Schnittstellenprobleme auf „open source“ Ebene wäre ein substantieller Beitrag zu einer wissenschaftlicheren Medizin.

Ähnlich wie die Schnittstellenproblematik ist die Pflege von Daten- und noch mehr von Gewebebanken auch ein finanzielles Problem, das derzeit auf institutioneller Ebene nicht gelöst ist. Neben zu verallgemeinernden und damit internationalen Problemen wie dem Datenaustausch existieren nationale Schwierigkeiten, die auf die komplexe Struktur des deutschen Gesundheitssystems zurückzuführen sind [KP09, HB09]. Diese Schwierigkeiten reichen tief bis hin zu den anstehender Entscheidung, welche Austauschformate bei Neuentwicklungen zu bevorzugen sind (HL7CDA,D2D).: Lediglich im kommerziellen Bereich existieren Ansätze von Lösungen wie etwa die Möglichkeit Tumorgewebe langfristig gegen Gebühren einzufrieren.

Zusammenfassend stehen wir ganz offenkundig am Beginn einer IT Durchdringung der Medizin. Ohne eine weit engere Zusammenarbeit von Informatikern/Mathematikern und Ärzten/Biologen wird wie in der Vergangenheit die kommerzielle Nutzung der IT Technologie in der Medizin weiter rasch voranschreiten ohne dass die ärztliche und wissenschaftliche Nutzung damit Schritt hält

Literaturverzeichnis

[Adj] Adjuvant t-online

[Als08] MD (2008). Chance für eine hochwertige Versorgung. Computereinsatz in der Medizin. Deutsches Ärzteblatt 105: B1632-1634

[Bla03] Black N (2003) Using clinical databases in practice (BMJ 326(7379):2-3, URL <http://www.bmj.com>

[DWK05] Delen D, Walker G, Kadam A (2005). Predicting breast cancer survivability: a comparison of three data mining methods. Artificial Intelligence in medicine 34:113-127

[BFR09] Braun N, Fritz P, Rieth A, Schroth W, Kimmel M, Biegger D, Zakim D, Alscher MD (2009). Predictors for treatment success and the expression of the glucocorticoid receptor in giant cell arteritis and polymyalgia rheumatica. J Rheumatology

- [DFK02] Dippon J, Fritz P, Kohler M (2002). A statistical approach to case based reasoning, with application to breast cancer data. *Computational Statistics & Data Analysis*. 40:579-602
- [HB09] Heidenreich g, Blobel b (2009). IT Standards für telemedizinische Anwendungen. Der Weg zum effizienten Datenaustausch in der Medizin . *Bundesgesundheitsbl-Gesundheitsforsch-Gesundheitschutz* 52:316-323
- [Fin] Finprog Studie <http://www.finprog.org/>
- [KDF09] Klenk S, Dippon J, Fritz P, Heidemann G (2009). Interactive survival analysis with the OCDM system. Interactive survival analysis from development to application . *Information Systems Frontiers*
- [KP09] Klar R, Pelikan E (2009).Stand, Möglichkeiten und Grenzen der Telemedizin in Deutschland. *Bundesgesundheitsbl-Gesundheitsforsch-Gesundheitschutz* 52:263-267
- [Lu03] Lunding J, Lundin M, Isola J, Joensuu H (2003). Onfopoints: A web-based system for individual survival estimation in breast cancer. *BMJ* 326:29-, URL, <http://www.bmj.com>
- [RRS03] Radespiel-Tröger M, Rabenstein T, Schneider HT, Lausen B (2003). Comparison of tree-based methods for prognostic stratification of survival data. *Artificial Intelligence in Medicine*. 28:323-341
- [RSA08] Rokavec M, Schroth W, Amaral SM, Fritz P, Antoniadou L, Glavac D, Simon w, Schwab M, Eichelbaum M, Brauch H, (2008). A polymorphism in TC21 promoter associates with an unfavorable tamoxifen treatment outcome in breast cancer. *Cancer Res* 68:9799-9808
- [Sc09] Schmidt s, Grimm G (2009). Versorgungsforschung zu telemedizinischen Anwendungen. . *Bundesgesundheitsbl-Gesundheitsforsch-Gesundheitschutz* 52:270-277
- [ZBF08] Zakim D, Braun N, Fritz P, Alscher DM. Underutilization of information and knowledge in everyday medical practice: Evaluation of a computer-based solution. *BMC Medical Informatics and Decision Making* 2008, 8:50

Die vorliegenden Arbeiten wurden unterstützt aus Fördermitteln der Robert Bosch Stiftung und der Sabine Döriges Stiftung Ludwigsburg