



PRECISE DATA STATT BIG DATA

Big Data ist im Gesundheitswesen in aller Munde. Doch leider führen viele Initiativen und Projekte in eine Sackgasse. Es genügt nämlich nicht, nur viele Daten zu sammeln. Man muss sich vorher genau im Klaren sein, was man damit machen will, die Nutzung sorgfältig planen und immer gezielt vorgehen. Ein professionelles Informationsmanagement ist notwendig, um ans Ziel zu gelangen.

TEXT: WERNER HACKL, ELSKE AMMENWERTH

Moderne klinische Informations- und Dokumentationssysteme tragen wesentlich zu einer effizienten und hochwertigen Gesundheitsversorgung bei. Sie unterstützen sowohl Patientenversorgung als auch klinische Forschung und haben wichtige ökonomische Funktionen. Durch zunehmende Spezialisierung in der Medizin und Fortschritte in der Informationstechnologie (IT) werden diese Systeme immer komplexer. Auch die Menge an Daten sowie Informationen, die durch diese Systeme produziert und verarbeitet werden, steigt exponentiell an. Durch elektronische Gesundheitsakten und Plattformen zur institu-

tionsübergreifenden Vernetzung haben die einzelnen Gesundheitseinrichtungen auch zunehmend Zugriff auf Daten aus externen Quellen.

Hinzu kommt, dass nicht nur mehr Gesundheitseinrichtungen und Kostenträger allein Daten zu Patienten und deren Behandlungen produzieren und speichern. Die Patienten selbst schlüpfen zunehmend in die Rolle der Datenproduzenten und -nutzer. Laufend werden neue Apps und mobile Geräte für unterschiedlichste Aufgaben entwickelt. Vom hoch spezialisierten Disease Management bei chronischen Krankheiten bis hin zu Wellness- und Lifestyle Anwendungen reicht das Spektrum an Daten messenden, generierenden, verarbeitenden und speichernden Applikationen. In manchen Fällen, zum Beispiel in der Diabetes-Versorgung, schließt sich bereits der Kreis und patientengenerierte Daten können zurück in klinische Systeme fließen. Die Patienten selbst erhalten durch elektronische Gesundheitsakten und spezielle Zugangsportale auch zunehmend Einsicht in und auch Zugriff auf ihre Daten in klinischen Informationssystemen.

KLINISCHE DATEN ALS GOLDMINE?

Die Gesundheitseinrichtungen auf der anderen Seite haben dadurch eine enorme Menge an Daten zur Verfügung, die auch abseits der primären Verwendungszwecke Patientenversorgung und Abrechnung genutzt werden können. Daten werden generell oft als Gold des 21. Jahrhunderts bezeichnet und neue Berufsbilder, die sich mit der Nutzung großer Datenbestände beschäftigen – wie etwa Data Scientists – entstehen. Auch in der Welt des Gesundheitswesens, insbesondere in Medizin und Pflege gibt es große Erwartungen hinsichtlich der Nutzung integrierter medizinischer Datenbestände. Vom vermeintlichen Wert der vorliegenden Daten gebendet, werden aber viele große Heraus-

forderungen, die zu lösen sind, wenn man Daten aus der klinisch-pflegerischen Routine für sekundäre Zwecke nutzen will, außer Acht gelassen. Viele Kliniker/innen machen sich die Welt zu einfach und denken, die sprichwörtliche Goldmine, auf der man sitze, brauche nur endlich einfach erschlossen zu werden. Man habe die einzelnen Daten ja eh schon strukturiert vorliegen und müsse sie nur zusammenführen, also auf einen Haufen zusammenwerfen und diese Datenquelle dann „richtig anzapfen“, um das neue, darin verborgene Wissen, das „Gold“, zutage fördern zu können.

Um für die Zukunft gerüstet zu sein und den Aufsprung auf den Zug ja nicht zu verpassen, beginnt man vielerorts auch schon mal damit, Daten zu sammeln, um sie für den Fall verfügbar zu haben. Der Begriff BIG DATA ist daher auch im medizinisch-pflegerischen Kontext ein oft verwendetes Schlagwort geworden. Übliche Definitionen beschreiben Big Data als große Datenmengen aus unterschiedlichen Quellen, die mit herkömmlichen Methoden und Mitteln nicht mehr beherrscht und analysiert werden können. Unter Big Data versteht man aber landläufig nicht nur, dass es sich um große und komplexe Datenbestände handelt, sondern dass es auch Möglichkeiten gibt, diese Daten sinnbringend zu analysieren. Daher werden auch oft Methoden und Tools zur Verarbeitung und Analyse dieser Daten unter dem Begriff Big Data subsummiert. Eine simplizistische (und vielleicht auch humoristisch zu wertende) Definition beschreibt Big Data als alles das, was zu groß für oder nicht mehr machbar mit Excel sei.

Charakteristisch für Big Data sind unter anderem aber insbesondere folgende Eigenschaften, die im Englischen gerne mit den „big V's“ bezeichnet werden: Volume – also die Menge an Daten; Variety, also ihre Heterogenität bezüglich Herkunft (Datenquel-

len), Typ, Codierung, Strukturierungsgrad, etc.; Velocity, die hohe Geschwindigkeit ihres Zuwachses und ihrer Verarbeitung; Variability – ihre Veränderlichkeit und damit potenziell hohe Inkonsistenz; und Value – ihr hoher Wert. Zusätzlich muss man gerade für Big Data Anwendungen im Gesundheitsbereich folgende Eigenschaften beachten: Veracity – ihre möglicherweise durch mannigfaltigen Bias, Rauschen oder Artefakte beeinträchtigte Wahrhaftigkeit und daraus resultierende Verzerrungen; und Volatility – ihre Flüchtigkeit, also das naturgegebene Ablaufdatum ihrer Gültigkeit oder Relevanz, das für jeden einzelnen Datensatz abhängig vom Anwendungszweck verschieden sein kann. Ein zwei Wochen alter Blutdruckwert kann zum Beispiel im Zuge der Betrachtung eines akuten Ereignisses bedeutungslos, aber für die Einschätzung einer Langzeitentwicklung durchaus relevant sein.

KLINISCHE DATEN: BIG DATA ODER BIG TROUBLES?

Die beiden letztgenannten Eigenschaften gelten insbesondere für Daten aus medizinischen Anwendungen. Im Gegensatz zu Daten aus anderen Bereichen, z.B. Umsatzzahlen, sind medizinische Daten stark kontextabhängig. Das heißt, dass einzelne Werte ohne nähere Information über ihre Entstehung, also wozu, wann, von wem und wie sie erhoben wurden sowie ohne Information über Einflussfaktoren wie aktuelle Diagnose, Medikation oder Gesundheitsstatus eines Patienten nicht ohne weiteres für patientenübergreifende Analysen zusammenführbar sind. Und dies gilt für alle klinischen Parameter und macht den Umgang mit medizinischen Daten so komplex.

Hinzu kommt, dass medizinische und pflegerische Daten in der Regel für bestimmte Zwecke in bestimmter Form dokumentiert werden. Dies hat Auswirkungen auf Inhalt, Detail- >

» Die Qualität jedes Analyseergebnisses hängt direkt von der Qualität der verwendeten Daten ab.«



lierungsgrad und Grad der Standardisierung der Daten und kann sogar so weit gehen, dass es zu Verschiebungen zwischen der wahrgenommenen Realität und der dokumentierten Repräsentation dieser Realität kommen kann. So kann die Dokumentation von abrechnungsrelevanten Diagnosen zum Beispiel anders aussehen, als sie unter rein medizinischen Gesichtspunkten aussehen würde. Und dies kann wiederum signifikante Auswirkungen auf die Validität und Aussagekraft von Analyseergebnissen haben.

Erschwerend kommt hinzu, dass klinische Daten in Gesundheitseinrichtungen in der Regel verteilt in verschiedenen Abteilungen produziert werden und verteilt auf zahlreiche Anwendungssysteme vorliegen. Dabei gibt es oft einen erheblichen Grad an Redundanz und immer wieder kommt es zu Inkonsistenzen und Widersprüchlichkeit. Gleichzeitig ist der Grad der Strukturierung, Codierung und Detaillierung in den verschiedenen Anwendungssystemen oft verschieden.

Dies alles, also die Kontextabhängigkeit, die Zweckgebundenheit und die Verteilung der Daten, macht es besonders schwierig, eine „single source of truth“ aufzubauen, die Grundvoraussetzung für jede reliable und valide Big Data-Anwendung ist. Wenn man die oben beschriebenen Entwicklungen, insbesondere die Rolle des Patienten als Datenproduzenten und die voranschreitende einrichtungs-

übergreifende Vernetzung mitberdenkt, steigen die Schwierigkeiten noch weiter.

KOMPLEXITÄT UND BIAS REDUZIEREN – PRECISE BEATS BIG

Um mit diesen Schwierigkeiten umzugehen, kann es nur eine Lösung geben: Eine gezielte und wohl überlegte Auswahl und Analyse von Daten gemäß dem Motto: Precise Data statt Big Data. Was bedeutet dies genau? Zunächst einmal macht es wenig Sinn, viel Energie aufzuwenden, um möglichst große Datenbestände aufzubauen, die dann ungenutzt brachliegen. Viel klüger ist es, vorher Ziel und Fragestellung einer jeden Analyse genau zu definieren und dann die bestgeeigneten Datenquellen auszuwählen, die diese Fragestellung möglichst valide und verzerrungsfrei zu beantworten. Dies kann insbesondere bedeuten, sich für eine von mehreren möglichen Datenquellen bewusst zu entscheiden. Durch jede einzelne Analyse baut man Erfahrungen auf, die sich für zukünftige Fragestellungen weiter nutzen lassen. Man beginnt mit einem Kerndatensatz, der präzise auf die konkreten Fragestellungen zugeschnitten ist und erweitert ihn bei Bedarf, sobald neue Fragestellungen oder relevante Daten für die vorhandenen Fragestellungen hinzukommen. Anstatt nach Big Data zu streben, kann es also zielführender sein, nur den jeweils minimal nötigen, bzw. maximal relevanten Datensatz zu verwenden.

Überhaupt ist es angezeigt, von Anfang an geplant vorzugeben. Inzwischen gibt es validierte Ansätze und praxistaugliche Vorgehensmodelle, um patientenübergreifende Auswertungen basierend auf heterogenen, verteilten Datenbeständen systematisch planen und zielgerichtet durchführen zu können. Einer davon ist SPIRIT, ein Framework und Vorgehensmodell für die strategische und systematische Planung der Sekundärnutzung von Daten im Gesundheitswesen. SPIRIT beschreibt detailliert, wie man zielgerichtet, ausgehend von einer Analyse der Fragestellung sowie einer Erhebung der vorhandenen Datenquellen unter Berücksichtigung des Kontexts schrittweise eine Big Data-Plattform erstellen kann. Natürlich kann ein solcher Ansatz auch verwendet werden, um einzelne Projekte oder konkrete Anwendungen zur Sekundärnutzung medizinischer Daten systematisch zu planen und umzusetzen.

Zurzeit beschäftigen wir uns zum Beispiel an der Tiroler Health and Life Science Universität UMIT in Kooperation mit klinischen Partnern, Partnern aus dem Pflegebereich und der atacama Software GmbH als Firmenpartner gerade intensiv in einem vom österreichischen Wissenschaftsfonds FWF geförderten Grundlagenforschungsprojekt mit der Frage, wie klinische Daten genutzt werden können, um die Patientensicherheit zu verbessern. Im Zuge des Projektes, in dem auch der

SPIRIT Ansatz zum Einsatz kommt, wird unter anderem untersucht werden, welche relevanten Fragestellungen es in diesem Bereich gibt, was konkrete Indikatoren für Patientensicherheit sein können, wie ein entsprechendes Patient Safety Minimum Data Set aussehen könnte und wie Ergebnisse aus Sekundäranalysen in der Praxis genutzt werden können, um zu einer Verbesserung der Patientensicherheit beizutragen.

Bei aller sorgfältigen Planung ist aber der alte Grundsatz zu beachten: Garbage in – Garbage out. Die Qualität jedes Analyseergebnisses hängt direkt von der Qualität der verwendeten Daten ab. Und diese wiederum hängt entscheidend von der Qualität des Informationssystems und der informationsverarbeitenden Prozesse ab – also vom Informationsmanagement einer Einrichtung. Hierzu zählt, zum Beispiel: gut organisierte Dokumentationsprozesse; einheitliche Codierungen wesentlicher klinischer Daten; eine gute IT-Infrastruktur; Vermeidung von Medienbrüchen und Doppeldokumentation durch geeignete Integration von Anwendungssystemen; gut geschulte Anwenderinnen und Anwender, die den Sinn einer Dokumentation verstehen und zeit- sowie realitätsnah dokumentieren; ein Qualitätsmanagement, welches auch die Güte der Dokumentation überwacht und kontinuierlich verbessert; sowie entsprechend ausgebildetes Fachpersonal wie Medizinische Informatiker/innen, Pflegeinformatiker/innen und insbesondere Informationsmanager/innen, welche die wichtige Funktion von Brückenbauern zwischen den verschiedenen Welten von Klinik und Informatik übernehmen.

FACHKRÄFTE BENÖTIGT!

Derartige, gut ausgebildete Fachkräfte werden zunehmend gesucht, sowohl in Gesundheitseinrichtungen als auch in der Industrie. Dies zeigt auch die

Initiative INIT-G des Bundesverband Gesundheits-IT (bvitg e.V.). Diese Initiative will den Fachkräftemangel in diesem Bereich adressieren. INIT-G wird von namhaften Institutionen und Fachgesellschaften wie BVMI, DVMD, GI, gmds oder KH-IT unterstützt.

Nicht überraschend entstehen derzeit auch neue Studienangebote in diesem Bereich, welche zunehmend auch berufstätige Personen ansprechen, da gerade diese gut die Rolle eines Brückenbauers übernehmen können. Entsprechende Angebote gibt es an verschiedenen Standorten, die Übersicht über Studiengänge auf der Webseite der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie GMDS e.V. zeigt dies. Ein ganz neues Studienangebot, welches dediziert die Brückenbauerfunktion adressiert, ist der Universitätslehrgang Health Information Management an der UMIT. ■

Mehr Informationen zum SPIRIT Ansatz unter <http://tinyurl.com/data-spirit> und zu Health Information Management unter www.umat.at/him.



■ **UNIV.-PROF. DR. ELSKE AMMENWERTH**

leitet das Institut für Biomedizinische Informatik der UMIT in Hall in Tirol

Kontakt: elske.ammenwerth@umat.at



■ **ASS.-PROF. DI DR. WERNER HACKL**

ist Assistenzprofessor und Experte für Sekundärnutzung klinischer Daten an der UMIT in Hall in Tirol.

Kontakt: werner.hackl@umat.at