



Wer sucht, ...

der findet oft mehr
als er zu suchen ging.

Molière

Big Data – die Medizin im Datenrausch

Evidenz aus Studien vs. unstrukturierte Daten?

Die allgegenwärtig fehlende Bewertung von Nutzen – Risiko – Kosten der Digitalisierung ist nicht wissenschaftlich zu erklären, berichtet Gerd Antes. Sie wird jedoch sofort einleuchtend, wenn man erkennt, dass wir es nicht mit der Entwicklung zum Vorteil der Patienten zu tun haben, sondern mit Realisierung einer Ideologie, deren Ziele und Sinn nicht hinterfragt werden können. Ärzteschaft, Gesundheitsprofession und Patienten sind gefordert, das Menschliche in der Medizin zu erhalten oder aber bewusst darauf zu verzichten, sich jedoch nicht von der Technik überrollen zu lassen¹.

Die Welt scheint nur noch ein Ziel zu haben: Digitalisierung! Es gibt keine Entwicklungen mehr, für die nicht phantastische Fortschritte angekündigt werden, wenn die Digitalisierung nur konsequent angegangen wird. Selbst eine oberflächliche Betrachtung lässt jedoch schnell Ernüchterung einkehren. Für eine solch alles erfassende, technikgetriebene Umwälzung sollte man eine auf Hochtouren laufende Technikfolgenabschätzung erwarten. Weit gefehlt, man findet diese überhaupt nicht oder von den dort üblichen Dimensionen:

- Nutzen
- Risiko/Schaden
- Kosten

nur den Nutzen. Mögliches Risiko und Schäden wie zum Beispiel Fehlentwicklungen werden nicht thematisiert, genauso wenig die Kosten. Wo nur Nutzen existiert, scheinen alle Kosten gerechtfertigt.

Big Data: Zukunftsversprechen auf einem brüchigen Fundament

Seit circa 15 Jahren wird die Vision entwickelt und über vielfache Kanäle verbreitet, dass der

kommende Überfluss an Daten uns eine bessere Welt durch unbeschränkte Erkenntnis eröffnet. Für den gesunden Menschenverstand erscheint diese Aussage so selbstverständlich, dass sie nicht hinterfragt werden muss. In einem einflussreichen Artikel wurde diese eingängige, jedoch naive Einschätzung vor circa 10 Jahren wissenschaftlich formuliert, indem das Ende der Theorie verkündet wurde: Die Datensintflut macht die wissenschaftliche Methode überflüssig². Die Aussagen in diesem Artikel führten, da sie mit den Grundlagen des Wissenschaftssystems nicht vereinbar waren, in eine neue Wissenschaftswelt und verfestigten diese durch regelmäßige, unkritische Wiederholung.

Während die empirische Forschung der alten Erkenntniswelt in den letzten Jahrzehnten durch ein stetes Ringen um Qualität charakterisiert war, erscheint das in der neuen, unbeschränkten Datenwelt überflüssig. Aufwendige Qualitätssicherungsmaßnahmen werden ohne theoretisches Fundament durch unbegrenzten Datenreichtum ersetzt. Aussagen zur Ergebnisqualität sowie Forderungen zur guten wissenschaftlichen Praxis und speziell zu Good Clinical Practice sucht man vergebens.

Sehr ähnlich sieht es bei der Nutzung und Auswertung großer Datenbestände mit künstlicher Intelligenz (KI) aus: Statt detaillierter Methodenbeschreibungen in klassischen Auswertungen findet man jetzt oft nur den Satz „... wurde mit künstlicher Intelligenz ausgewertet ...“.

Fundamentale wissenschaftliche Schwachpunkte

Big Data entzieht sich der kritischen Bewertung durch eine Definition, die keine ist. Üblicherweise wird Big Data durch 3 (inzwischen auch 5) „V“ charakterisiert:

- Datenmenge (Volume),
- Geschwindigkeit (Velocity) und
- unterschiedliche Beschaffenheit (Variety).

Die fehlende Quantifizierung erlaubt keine eindeutige Qualifizierung, ob ein großer Datenkörper dazu gehört oder nicht. Entsprechend groß sind die terminologische Verwirrung und der daraus folgende Begriffswirrwarr.

Zentraler Mechanismus von Big Data ist die Interpretation von „entdeckten“ Korrelationen als Kausalzusammenhang. Bedingung ist, dass ausreichend – und damit unbeschränkt – Daten zugänglich sind. Diese Aussage ist falsch, wie in sehr anspruchsvollen mathematischen Arbeiten³ dargelegt wird. Mehr Daten sind nicht äquivalent mit mehr Information, vor allem nicht mit weniger Aufwand, wie im Big-Data-Mainstream allerorten vermittelt wird. Diese kontraintuitive Aussage ist nur schwer zu veranschaulichen. Am besten so: Bei der Suche nach Zusammenhängen sind falsch positive Funde unvermeidlich. Diese können in wachsenden Datenmengen schneller zunehmen als die richtigen Funde, das heißt diese Signale gehen zunehmend in wachsendem Rauschen unter. Anschaulich formuliert heißt das, die Nadel befindet sich in einem größer werdenden Heuhaufen und der Aufwand, sie zu finden, nimmt nicht ab, sondern wächst sogar⁴.

Big Data in der Medizin: Spezielle Herausforderungen

Medizin und Gesundheitsversorgung sind in dieser Entwicklung Getriebene und Treiber. Getrieben, da sie in dem vor allem in dem von Internetriesen ge-

triebenen Hype aufgrund der enormen Budgets ein einladendes Ziel mit neuen, großen Märkten sind. Treiber, da sowohl Patienten und Gesunde wie auch die Gesundheitsprofessionen und Institutionen sehr anfällig für die Verkündigungen von verbesserter Diagnostik und Therapie durch Big Data sind.

Die oben dargestellten Charakteristika von Big Data gelten in jeder Beziehung auch für die Medizin, in mehrfacher Hinsicht hat die Medizin darüber hinaus eine Sonderrolle. Vor allem die Missachtung von Risiken und Kosten hat hier eine Bedeutung, die sehr viel ernster genommen werden muss als in anderen fachlichen Zusammenhängen, wo auftretende Schädigungen als Verschwendung abgetan werden können. In der Medizin kann es vermeidbare Krankheit und Tod bedeuten und ist damit unvereinbar mit dem Prinzip, dass der Schutz des Patienten über allem steht.

Spezielle Ausprägungen aufgrund der Datenflut sind die Hoffnung auf immer individualisiertere Behandlungsmöglichkeiten mit automatisierten Entscheidungsprozessen. Diese Entwicklungen machen es weitgehend unmöglich, einzelne Begriffe aus Digitalisierung, künstlicher Intelligenz, Deep Learning, Big Data, personalisierter oder individualisierter Medizin isoliert zu betrachten, da sie untrennbar miteinander verbunden sind. Eine Folge davon sind verwirrte Diskussionen aufgrund nicht eindeutiger Definitionen.

Deutlich wird das beim Thema „Vertrauen in Daten“ und den unvermeidlichen Entscheidungen unter Unsicherheit (uncertainty). Ebenso wie der Qualitätsbegriff ist die Betrachtung und Quantifizierung von Unsicherheit in der KI- und Big-Data-Welt weitgehend verschwunden. Dank unbegrenzter Daten wird das Bild vermittelt, dass damit Qualitätsprobleme und Unsicherheit nicht mehr thematisiert werden müssen. Das ist fachlich nicht haltbar und birgt die Gefahr von Patientengefährdung und systematischer Fehlentwicklung (mehr Daten erlauben, Fehler mit größerer Präzision zu machen).

Die Medizinsysteme sind gefordert, alle Anstrengung darauf zu verwenden, in den unter dem Schlagwort Digitalisierung laufenden Entwicklungen die Spreu vom Weizen zu trennen und wieder das in den Mittelpunkt zu stellen, was an vielen Brennpunkten nicht mehr in der ersten Reihe steht: der Patientennutzen.

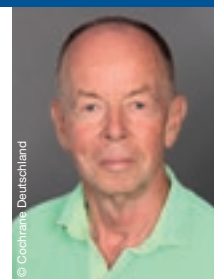
Das bedeutet die konsequente Nutzung der über Jahrzehnte entwickelten Bewertungsinstrumente für diagnostische und therapeutische Verfahren. Besondere Aufmerksamkeit gebührt dabei nicht offengelegten, kommerziellen Gesundheits-Apps sowie den sogenannten Prozeduren der künstlichen Intelligenz, die weitgehend als Black Box beschrieben werden und ohne Regulierung in die Gesundheitsversorgung eindringen⁵. Ebenso gilt es, dem wissenschaftlichen Grundprinzip, an vorhandenes Wissen anzuknüpfen, wieder höchste Priorität zu geben, anstatt es unter dem Schlagwort „disruptiv“ zu ignorieren und zu Sprunginnovation aufzurufen. ■

Literatur:

- ¹ Thielscher C, Antes G. Der Arzt behält die Deutungshoheit trotz KI. Dtsch Arztebl 2019; 116: A-18/B-18/C-18. Im Internet: www.aerzteblatt.de/archiv/204288/Der-Arzt-behaelt-die-Deutungshoheit-trotz-KI
- ² Anderson C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Im Internet: www.wired.com/2008/06/pb-theory
- ³ Meng XL, Xie X. Forthcoming. I Got More Data, My Model Is More Refined, but My Estimator Is Getting Worse! Am I Just Dumb? *Econometric Reviews*. Im Internet: <https://dash.harvard.edu/handle/1/10886849>
- ⁴ Taleb N. The Big Errors of Big Data. Im Internet: <https://fs.blog/2013/02/the-big-errors-of-big-data/>
- ⁵ Leetaru K. How Data Scientists Turned Against Statistics. Im Internet: www.forbes.com/sites/kalevleetaru/2019/03/07/how-data-scientists-turned-against-statistics/

Aktuell erschienen im Kongressband des 125. Kongresses der Deutschen Gesellschaft für Innere Medizin e.V., 4. – 7. Mai 2019. Nachdruck mit freundlicher Genehmigung des Karl Demeter Verlags im Georg Thieme Verlag, Stuttgart.

PROF. DR. RER. NAT. GERD ANTES



Institut für Didaktik und
Ausbildungsforschung in der Medizin,
Klinikum der Universität München,
LMU München