https://mail.gen-ethisches-netzwerk.de/node/3885

Gen-ethischer Informationsdienst

Big Data, big Errors

Mehr Daten bedeuten nicht zwangsläufig mehr Evidenz

AutorIn
Gerd Antes

Seit über 20 Jahren setzte sich Gerd Antes für eine evidenzbasierte Medizin und Qualitätssicherung in der biomedizinischen Forschung ein. Am aktuellen Trend der Präzisionsmedizin und Big Data-Gesundheitsforschung kritisiert er die Untergrabung von Standards guter Wissenschaft und Forschung zugunsten leerer Versprechen.

Die Präzisionsmedizin verspricht alle Erkrankungen durch die Einteilung in Subgruppen anhand von Biomarkern präzise behandeln oder präventiv therapieren zu können – ist das erfüllbar?

Die Antwort auf diese Frage erfordert selbst eine hohe Präzision, um das oft plumpe Marketing einerseits von einer realistischen Einschätzung und den realisierbaren Zielen zu trennen. Die rote Linie befindet sich zwischen sicheren Vorhersagen und der Angabe von Wahrscheinlichkeiten, mit der Ereignisse eintreten werden. Der Begriff Präzision vermittelt die Botschaft, dass etwas unter völliger Kontrolle ist und genau gesteuert werden kann. Das würde jedoch bedeuten, dass wir auf eine deterministische Medizin zusteuern. Davon ist gegenwärtig nichts wahrzunehmen und es ist auch längerfristig sicherlich extrem unrealistisch. Bücher und andere Beiträge, die das Ende des Zufalls verkünden, belegen vor allem, dass die Autor*innen die Materie nicht durchdrungen haben. Es kann also tatsächlich nur um eine Verbesserung von Ursachenforschung und Prognosegenauigkeit gehen, das heißt um Verbesserungen auf der Wahrscheinlichkeitsskala. Den disruptiven Sprung ins nächste Zeitalter sehe ich nicht.

Sie stehen Big Data-Gesundheitsforschung kritisch gegenüber, warum?

Big Data ist ein Schlagwort, das mit einem Heiligenschein daherkommt, Dinge verspricht, die nicht so einfach realisierbar sind und dabei konsequent die Nutzen-Risiko-Kosten-Abwägung missachtet und auf Nutzenversprechungen reduziert. Das mag im Marketing der Internetfirmen akzeptabel sein, für die Medizin und die Gesundheitsforschung und -versorgung ist das jedoch hochgradig fahrlässig, kann es doch die Gefährdung von Gesundheit und Leben bedeuten. Es gibt keinen Nutzen ohne potenziellen Schaden, und gerade hochwirksame Maßnahmen unterliegen dieser Gefahr umso mehr.

Was ist aus statistischer Sicht das Problem an Big Data-Gesundheitsforschung, bei der genetische Varianten mit Erkrankungen korreliert werden?

Hinter allem steckt die – aus Sicht des gesunden Menschenverstandes sehr einsichtige – Einschätzung, dass mehr Daten mehr Erkenntnisse liefern. Das mag für manche Zusammenhänge gelten, ist jedoch in vielen Situationen nicht richtig. Letzteres ist für viele in diesem Feld so kontraintuitiv, dass die wissenschaftliche Auseinandersetzung damit vermieden wird und der scheinbar gesunde Menschenverstand als wissenschaftliche Wahrheit zum Ausgangspunkt für die Visionen und Versprechungen gemacht wird. Tatsächlich birgt jedoch gerade das Herumsuchen in unstrukturierten Datenmengen die Gefahr, Zusammenhänge zwischen Einflussgrößen zu „erkennen“, die zwar zahlenmäßig korrelieren, jedoch in einem völlig sinnfreien Zusammenhang stehen. Solche Treffer nennt man Falsch Positive, und diese von den Richtig Positiven zu trennen, ist seit Jahren Thema von mathematisch-statistischer Grundlagenforschung. Schlagworte wie „Big Data = Big Errors“ weisen auf diese lauernden Fallen hin. Oder als anschauliches Bild aus der Informationstheorie gezeichnet: Die falschen Signale werden auch als Rauschen (noise) bezeichnet, und wenn das stärker wächst als die richtigen Signale (signal), so wird der Erkenntnisgewinn stärker behindert. Das ist bekannt, wird jedoch in vielen Narrativen der Werbung unter dem Titel Analytics einfach ignoriert. Dort sieht es so aus, als sei das Problem mit großen Speichern und schnellen Rechnern gelöst. Genau das ist nicht so.

Warum teilen viele Forscher*innen und Mediziner*innen Ihre Skepsis nicht?

Weil sie sich großenteils nicht damit auseinandersetzen und die Nutzen-Risiko-Kosten-Bilanz nicht aufmachen. Das ist einerseits bequemer und damit menschlich verständlich. Deutlich unangenehmer ist der Blick auf die Anreize, die die ganze Entwicklung in die falsche Richtung treiben. Förder- und Investitionsprogramme werfen Milliarden in diese Richtung, ohne klare Zielvorstellungen zu formulieren. Forscher*innen und Mediziner*innen können diesem Trend nicht ausweichen und auf der persönlichen Ebene ist es die Ausnahme, zum Beispiel kurz vor einer Promotion oder Habilitation etwas zu formulieren,

was dem angekündigten Sprung ins Paradies widerspricht. Skepsis und kritische Betrachtung erfordern eine gewisse Auseinandersetzungsfähigkeit und sind unbequemer, als im Mainstream akademischen Fortschritt zu ergattern und die eigene Institution zu vergrößern. Der Mangel an kritischer Auseinandersetzung wird leicht verständlich, wenn man Big Data und Digitalisierung als das anschaut, was sie sind: Eine Ideologie.

Wie blickt die Community der evidenzbasierten Medizin auf Präzisionsmedizin und Personalisierte Medizin?

Die Evidenzbasierung mit ihren hohen Ansprüchen an rigide Methode, Transparenz, Reproduzierbarkeit hat die empirische Forschung in den letzten Jahrzehnten enorm nach vorn gebracht. Viele tausende Artikel zu methodischen Projekten lassen uns heute sehr gut wissen, wie Erkenntnisse aus Daten zu gewinnen sind. Gleichzeitig wurden aus diesen Arbeiten auch tiefe Einsichten in die Fehler und Defizite gewonnen, die zu fundamentalen Fehlschlüssen führen können und dies oft auch tun. Systematische Fehler bei der Datenerhebung und -auswertung (Bias) zu kontrollieren ist das oberste Prinzip dabei. Methodisch ist die Kontrolle und Minimierung dieses Risikos (Risk-of-Bias) zu einer der wichtigsten Leitlinien geworden. Um dem gerecht zu werden, muss in jedem einzelnen Schritt der Qualität größte Priorität gegeben werden. Good-scientific-practice [1] und alle darunter subsumierten Forderungen müssen erfüllt werden, um möglichst unverfälschte Ergebnisse zu erzielen.
Diese Welt befindet sich gegenwärtig auf Kollisionskurs mit einer neuen Welt, in der die bekannten Forderungen scheinbar nicht mehr gelten. Ich bin immer wieder erstaunt, wie wenig das von der Community der evidenzbasierten Medizin wahr- und ernstgenommen wird. Unbeschränkter Zugang zu „riesigen“ Datenmengen machen Theorie überflüssig, wie 2008 in einem Artikel verkündet wurde.[2] Qualitätssicherung, ausführliche Methodenbeschreibungen und Aussagen zur Datenqualität, die zu den Grundelementen der Evidenzkultur gehören, sucht man bei Big Data vergeblich. Ebenso Aussagen zur Genauigkeit und zum Vertrauen, das man in Resultate haben kann.

Freut sich die evidenzbasierte Medizin nicht über immer mehr Daten?

Nein. Die evidenzbasierte Medizin freut sich über mehr relevante und hochwertige Daten, salopp formuliert. Vor allem aber würde sie sich freuen über die Überwindung der unglaublichen Defizite wie zum Beispiel der Nichtpublikationsrate von immer noch nahe 50 Prozent, die eine enorme Schieflage in die Bewertungen von diagnostischen und therapeutischen Verfahren in Richtung chronischem Überoptimismus bringen. Das ist nur ein Beispiel für eine sehr schädliche Schieflage, die seit langem bekannt ist und trotzdem nicht konsequent angegangen wird.

Was wäre ein besserer Ansatz für die Gesundheitsforschung?

Wichtig ist, die seit vielen Jahren bekannten Qualitätsparameter guter Wissenschaft und Forschung weiterhin sehr ernst zu nehmen, sie sogar noch viel ernster zu nehmen, und nicht zu glauben, dass sie in einer Welt unbeschränkter Daten und Rechenleistung nicht mehr nötig seien. Auf dem Weg wird die Entwicklung unter einem Dach weitergehen und damit auch Evidenzbasierung und Big Data, künstliche Intelligenz und Personalisierte Medizin nicht gegeneinander, sondern miteinander die Medizin voranbringen.

Das Interview erfolgte schriftlich, die Fragen stellte Isabelle Bartram.

- 1Good-scientific-practice sind Richtlinien, die der wissenschaftlichen Selbstkontrolle dienen sollen.
- 2Anderson, C. (2008): The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. In: Wired, online: www.wired.com/2008/06/pb-theory [23.06.2008].

Informationen zur Veröffentlichung

Erschienen in:
GID Ausgabe 248 vom Februar 2019
Seite 10 - 11