

## Wissen

# «Vieles ist blankes Marketing»

**Big Data** Der profilierte Medizinstatistiker Gerd Antes kritisiert die leeren Versprechungen von Big Data in der Medizin und warnt vor falschen Datenanalysen mit gefährlichen Folgen für Patienten.

Felix Straumann, Freiburg i. B.

**Alle wollen unsere Daten und versprechen sich oder uns grossen Nutzen davon. Sie kritisieren die überzogenen Erwartungen an Big Data. Fühlen Sie sich nicht manchmal wie ein einsamer Spielverderber?**

Manchmal vielleicht. Aber ich sehe auch, dass ich etwas bewege. Die Leute haben ein Bauchgefühl, dass etwas nicht stimmt. Zu Big Data hört man nur das Dauerfeuer aus falschen Versprechungen und überhöhten Erwartungen, gerade was die Diagnose und die neuen Therapien von Krankheiten angeht. Risiken werden praktisch gar nie angesprochen. Und die Kosten auch nicht. In Deutschland geben wir bereits Milliarden für die Erfassung von persönlichen medizinischen Daten aus, ohne zu wissen, welchen Nutzen wir davon haben.

**Spätestens seit dem Datenskandal bei Facebook wird viel über Datenschutz und Ethik debattiert. Ihnen geht es aber in erster Linie um Big Data im Gesundheitsbereich. Wo liegt das Problem? Wer mehr weiss, kann besser therapieren.** Ich will die Datenschutzprobleme nicht kleinreden. Sie sind real. Doch die Schwierigkeiten reichen weit darüber hinaus. Mehr Daten bedeuten eben gerade nicht mehr Wissen. Im Gegenteil. Das geht gegen die Intuition und ist nicht ganz einfach zu erklären.

**Versuchen Sie es.**

Ein Beispiel, das ich jeweils an meinen Vorträgen bringe: Gemäss einer Statistik folgt der Pro-Kopf-Käsekonsum in den USA praktisch der gleichen Kurve wie die Zahl der Todesfälle durch Verheddern im eigenen Bettlaken. Zwischen den zwei parallel laufenden Entwicklungen besteht rein zahlenmässig ein Zusammenhang. Es wäre aber völliger Unsinn, daraus zu schliessen, dass man durch weniger Käse das Todesrisiko verringern könnte. Solche falschen, zufälligen Korrelationen finden sich viele, wenn Sie grosse Datenmengen durchsuchen. Das Big-Data-Problem dabei ist: Deren Anteil steigt schneller als derjenige der wahren Zusammenhänge.

**Wieso?**

Dahinter liegt eine komplizierte Mathematik, aber veranschaulichen lässt es sich. Wenn ich den Einfluss von hundert Faktoren untersucht habe und dann einen einzigen weiteren hinzunehme, so kann dieser mit jedem der vorhandenen zusammenhängen. Als einfachste Variante bekomme ich also hundert zusätzliche Zusammenhänge.

**Durch einen einzigen zusätzlichen Faktor?**

Genau. Das bedeutet, die Wahrscheinlichkeit für zufällige, aber falsche Korrelationen steigt schneller, als die Zahl der Daten selber zunimmt. Bei grossen Datenmengen wird dieser Effekt dramatisch. Man spricht in dem Zusammenhang auch von Rauschen. Der Lärm der falschen



«Dieses Fehlen von Fehlerbewusstsein ist irritierend»: Gerd Antes im Sitzungszimmer der Cochrane-Deutschland-Stiftung. Foto: Lucian Hunziker

Korrelationen überdeckt die tatsächlichen. Ein Problem, von dem übrigens auch die Medien betroffen sind. Beim Datenjournalismus besteht ebenfalls die Gefahr, dass Sie auf einen Zusammenhang stossen und eine Story draus machen. Dabei war die Korrelation zufällig und ohne inhaltliche Bedeutung.

**Die Leute, die grosse Datensätze analysieren, sind doch in der**

**Wegbereiter der evidenzbasierten Medizin**

Gerd Antes (68) leitet die Cochrane-Deutschland-Stiftung und das Institut für Evidenz in der Medizin am Universitätsklinikum Freiburg i. B. Der Mathematiker ist im deutschsprachigen Raum ein Wegbereiter der evidenzbasierten Medizin, die hochwertige Studien als Basis von Therapien fordert.

**Lage, solche Fehlschlüsse zu vermeiden.**

Bei den Aussagen dazu wird es mit der Wahrheit oft nicht sehr genau genommen. Vieles ist blankes Marketing. Oft verstehen die Leute auch die Basics nicht. Darum gibt es insbesondere in der Medizin keine konkreten Anwendungen. Big Data steht für das Versprechen, dass mit dem Zugang zu allen Daten alles besser wird. Es stört keinen, wenn ich an der Kasse mit der Kreditkarte einen neuen Anzug bezahle und dann auf meinem Handy die passende Krawatte erscheint. In der Medizin ist das anders. Wenn ich mich aufgrund einer falschen Datenanalyse für eine Behandlung entscheide, kann das böse Folgen haben. Hinzu kommt der volkswirtschaftliche Schaden. Die überall blind gesammelten Daten müssen nachher aufgeräumt werden, damit sich die echten Zusammenhänge entdecken lassen.

**Schlampige Datenanalysen sind ja aber nichts Neues. Das gab es schon vor Big Data.**

Es gibt einen Unterschied. Die Anhänger von Big Data würden gerne die statistischen Errungenschaften der letzten 100 Jahre über den Haufen werfen. So behaupten sie, dass ab einer gewissen Datenmenge eine Korrelation automatisch zur Kausalität wird. Die Sicherheit einer Vorhersage sei dann so hoch, dass diese zur Realität werde. Ein viel beachteter Artikel von 2008 titelte sogar «Das Ende der Theorie: Die Datensintflut macht die wissenschaftliche Methode überflüssig». Das ist blanker Unfug. Zu dieser Haltung passt, dass bei Big-Data-Analysen in der Regel keine Unsicherheitsbereiche angegeben werden wie sonst bei statistischen Analysen üblich.

**Warum?**

Die Data-Scientists leben in einer anderen Welt und haben eine

eigene Ideologie. Für unsereins ist dieses Fehlen von Fehlerbewusstsein irritierend. Wir betreiben evidenzbasierte Medizin, und da ist Unsicherheit aufgrund von fehlenden oder ungenauen Daten Alltag. Bei den Leuten, die mit grossen Datenmengen herumrechnen, ist dieses Bewusstsein grossenteils nicht vorhanden. Sie sprechen lieber vom Ende des Zufalls. In einem gleichnamigen Buch wird zum Beispiel behauptet, dass dank den vielen Daten alles vorausberechnet werden kann und es deshalb keinen Zufall mehr gibt.

**Ist das völlig abwegig?**

Es ist überhaupt nicht durchdacht. Zum Beispiel die Vorstellung, dass, wenn die Polizei genug Daten hat, sie Einbrüche prognostizieren kann. Die Modelle, die bis jetzt hinter entsprechenden Prognoseprogrammen stecken, sind geradezu peinlich primitiv. Belege für deren Nut-

zen sind nur anekdotisch. So wie generell bei Big Data. Es gibt keine belastbaren Belege für einen umfassenden Nutzen.

**Das Phänomen Big Data ist relativ neu. Gibt es Beispiele, wo es schiefgelaufen ist?**

Das bekannteste sind die Flu-Trends, die Grippevorhersage von Google. Dort war anfangs das Getöse riesig. Normalerweise lässt sich die Ausbreitung der Grippe nur relativ umständlich anhand von Arztbesuchen verfolgen. Vor rund zehn Jahren behauptete Google, dass dies mithilfe der Klicks von Usern möglich sei. Zwei Jahre hat das tatsächlich geklappt. Dann ist das Ganze an die Wand gefahren. Die hatten einfach zwei Jahre Glück.

**«Mehr Daten bedeuten eben gerade nicht mehr Wissen. Im Gegenteil.»**

**Aber es leuchtet doch ein, dass Suchabfragen Hinweise auf den Gesundheitszustand der Nutzer geben.**

Ja, schon. Aber Sie haben keine Kontrolle über systematische Fehler. Wenn die Grippe weit weg ist, die Medien aber darüber berichten, dann klicken Leute die entsprechenden Begriffe, obwohl sie völlig gesund sind. Das Klicken als Abbild der Infektionsraten funktioniert nicht.

**Als Negativbeispiel gilt auch der digitale Assistent Watson von IBM, der mithilfe von Big Data medizinische Diagnosen erstellen soll.**

MD Anderson der Universität Texas ist eines der grössten und renommiertesten Krebskrankenhäuser weltweit. Die haben das Projekt letztes Jahr beendet, nachdem sie 62 Millionen Dollar investiert haben. Laut Medienbericht wurde das Wissen manuell in den Computer eingegeben, statt es durch Maschinenlernen zu erzeugen, wie man es immer wieder in der Werbung liest. Das ist erstens peinlich, zweitens Betrug. Und dann war es drittens so, dass die Resultate schlechter waren als die der Ärzte.

**Wird es denn jemals möglich sein, medizinische Diagnosen mithilfe von grossen Datenmengen zu erstellen?**

Ja, man wird sicher besser werden. Aber noch ist es lange nicht so weit. In Deutschland hat der Ethikrat kürzlich für einen weniger strengen Datenschutz plädiert, um zusätzliche Datenauswertungen zu ermöglichen. Dazu sage ich Nein. Die Vorstellungen vom möglichen Nutzen sind zum grossen Teil Wunschenken. Auf Vorrat Milliarden ausgeben, in der Hoffnung, dass sich die Lösungen von selbst ergeben, ist sinnlos.