




---

 AUS DEM DEUTSCHEN NETZWERK EVIDENZBASIERTE MEDIZIN
 

---

## »Big data« in aller Munde

Warum die Auswertung riesiger Datenmengen künftig eine größere Rolle spielen könnte – aber nicht geeignet ist, klinische Studien zu ersetzen.

VON GABRIELE MEYER IM AUFTRAG DES DEUTSCHEN NETZWERKS EVIDENZBASIERTE MEDIZIN  
(DNEBM – [WWW.EBM-NETZWERK.DE](http://WWW.EBM-NETZWERK.DE))

# D

Durch die Mess- und Speichermöglichkeiten entstehen auch in der Medizin und Gesundheitsversorgung große Datenmengen. „Big data“ sind zu groß, zu komplex oder ändern sich in zu großer Geschwindigkeit, um mit den üblichen Methoden der Datenanalyse ausgewertet werden zu können. Von „Big data“ wird gesprochen, wenn Datenmengen im Terabyte-Bereich (1 Terabyte = 1024 Gigabyte) untersucht werden (1). „Big data“ können auch unstrukturiert, heterogen und unvollständig sein (2). Solch große Mengen an heterogenen Daten werden von Krankenkassen über die Versicherten erhoben und im stationären Bereich durch Untersuchungen wie Röntgen, CT, MRT oder Blutuntersuchungen erzeugt und in elektronischen Akten der Krankenhausinformationssysteme mit ärztlichen Berichten und Behandlungsverläufen archiviert. Prominent geworden ist das Projekt „Google Flu Trends“ (3) zur Vorhersage der Influenzaaktivität auf Regionalebene durch Bestimmung der Häufigkeit von Anfragen in der Suchmaschine Google. Üblicherweise liegen solchen Vorhersagen aufwändige

Datenerhebungen zugrunde, zum Beispiel durch die US Centers for Disease Control and Prevention (1). Das Google-Projekt galt als Vorzeigebeispiel für „Big data“. Spätere Untersuchungen haben allerdings gezeigt (4), dass es deutliche Abweichungen zwischen Vorhersage und Influenzageschehen gab. Verantwortlich sind vermutlich Modifikationen des Suchmaschinenalgorithmus zur Optimierung der primären geschäftlichen Nutzung, die dann die Influenzavorhersage als Sekundärnutzung in Mitleidenschaft gezogen hat (1).

Ist die „Data-based Medicine“ das Ende der Evidenzbasierten Medizin (EbM), wie manche Apologeten meinen? Was können „Big data“ zum wissenschaftlichen Erkenntnisgewinn in der Medizin und Gesundheitsversorgung beitragen?

Grundsätzlich repräsentieren EbM und „Big data“ zwei vollkommen unterschiedliche Ansätze der Informationsgewinnung. Im Verständnis von EbM haben Daten einen hohen Wert, wenn ihre Erhebung einem vorab festgelegten methodischen Studienprotokoll folgt, eine Hypothese zum Gegenstand des Untersuchungsziels prädefiniert wurde, qualitativ abgesicherte Studienmethoden benutzt wurden, das heißt eine sorgfältige Fallzahlplanung getätigt wurde, geeignete Maßnahmen der Vermeidung methodischer Verzerrungen (Bias) ergriffen wurden (bei Interventionsstudien sind dies insbesondere: externe, randomisierte Zuteilung zu den Studiengruppen, Verblindung der Intervention und/oder der Erhebung der Erfolgsparameter, Analyse nach Intention-to-treat-Prinzip), angemessene biostatistische Methoden zur Überprüfung der Hypothese gewählt wurden, die klinische Relevanz

und Übertragbarkeit (externe Validität) sorgsam erwogen wurde.

„Big data“ entspringt der Tradition der Datenverarbeitung und ist datengetrieben anstatt hypothesengesteuert. „Big data“-Methoden arbeiten sich an rohen Beobachtungen ab und berücksichtigen kein kontextuelles Wissen in der Datengenerierung. Mit Algorithmen können Muster in den analysierten „Big data“ identifiziert werden und Zusammenhänge zwischen Merkmalen dargestellt werden. Ähnlich wie beim Zusammenhang zwischen Storchenflug und der Geburtenhäufigkeit (5) kann es sich jedoch um die Koexistenz von nicht voneinander abhängigen Merkmalen handeln - statt um einen echten Zusammenhang.

EbM hingegen schreibt der Kontrolle von Störgrößen und methodisch bedingten Verzerrungen im Bemühen um die bestmögliche interne Validität höchste Priorität zu. Das Ziel von klinischen Studien zum Nachweis von Wirksamkeit und Sicherheit von Interventionen ist es, Ursache-Wirkungs-Zusammenhänge zu belegen, die reproduzierbar sind.

„Big data“-Ansätze sind selten protokollgestützt und folgen dem Diktum: „Mehr Daten sind besser als bessere Daten“ (6). „Big data“ teilen nicht das Problem klinischer Studien, eine angemessene Fallzahl zu generieren, denn es gibt keine Datenknappheit. Statistische

Signifikanz ist nicht das Problem von „Big data“, denn die großen Datenmengen führen mit Leichtigkeit zu statistisch signifikanten Ergebnissen zwischen identifizierten Zusammenhängen (7).

Zwar sind „Big data“ nicht geeignet, sorgfältig geplante klinische Studien und Beobachtungsstudien abzulösen, doch haben „Big data“ möglicherweise in Zukunft einen wichtigeren Stellenwert in der wissenschaftlichen Medizin. So zum Beispiel bei der Generierung von Computeralgorithmen zur Vorhersage kritischer Ereignisse wie einer Sepsis (8). Selbstverständlich müssen solche Verfahren zur Vorhersage auch einer Validierung unterzogen werden, mit dem üblichen Standardverfahren verglichen werden und vor allem im Hinblick auf patientenrelevante Ergebnisparameter untersucht werden. Für die Nutzer von wissenschaftlicher Literatur sind EbM-Kompetenzen unerlässlich, um Ergebnisse aus den „Big data“-Studien auf ihre Aussagekraft und Gültigkeit hin hinterfragen zu können. ■

**Prof. Dr. Gabriele Meyer**  
**Martin-Luther-Universität Halle-Wittenberg**  
**Medizinische Fakultät**  
**Institut für Gesundheits- und Pflegewissenschaft**  
**E-Mail: Gabriele.Meyer@medizin.uni-halle.de**

(1) Binder H, Blettner M. Big data in medical science - a biostatistical view. Part 21 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2015; 112: 137-142

(2) Zukunftsrat der Bayerischen Wirtschaft. Zukunft digital – Big Data Analyse und Handlungsempfehlungen. Stand Juli 2016. <http://www.vbw-zukunftsrat.de>, Zugriff am 18.12.2016

(3) Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009; 457: 1012-1014

(4) Lazer D, Kennedy R, King G, Vespignani A. The parable of google flu: Traps in big data analysis. Science 2014; 343: 1203-1205

(5) Höfer T, Przyrembel H, Verleger S. New evidence for the theory of the stork. Paediatr Perinat Epidemiol 2004; 18: 88-92

(6) Sim I. Two Ways of Knowing: Big Data and Evidence-Based Medicine. Ann Intern Med 2016; 164: 562-563

(7) Antes G: Big data und personalisierte Medizin. Goldene Zukunft oder leere Versprechungen? Es erscheint überfällig, die Versprechungen für die Zukunft auf den Prüfstand der Qualitätsdiskussion (Good Scientific Practice) zu stellen. Deutsch Arztebl 2016; 113: A-712-713

(8) Lindner HA, Balaban Ü, Sturm T, Weiß C, Thiel M, Schneider-Lindner V. An algorithm for systemic inflammatory response syndrome criteria-based prediction of sepsis in a polytrauma cohort. Critical Care Medicine 2016; 44: 2199-2207

