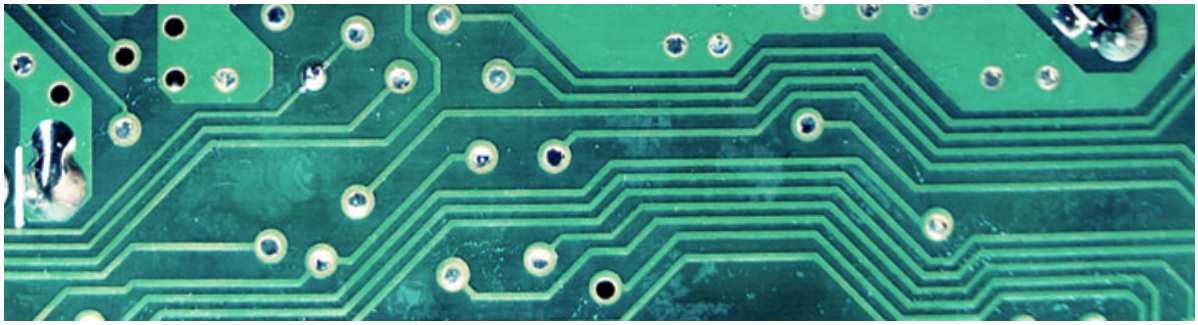
www.vems.ch

Factsheet

Welche Gefahren birgt das Anhäufen grosser Datenmengen für das Gesundheitswesen?

Daten sind nicht gleich Wissen

Stand Juli 2014



Verein Ethik und Medizin Schweiz, Ziegelfeldstrasse 1, 4600 Olten
Tel: 062 212 44 10, www.vems.ch

Kurzantwort:
Dass grosse Datenmengen zwingend zu einem entsprechend grossen Wissens führen, ist ein gefährlicher Irrtum. Wenn mangelhafte Daten falsche Schlüsse nahelegen und scheinbar auch beweisen, dann entsteht nicht nur Nichtwissen, sondern sogar eine falsche Sicht der Dinge, aus welcher falsche Handlungen folgern.

Daten sind nicht gleich Wissen

Die moderne Informationstechnologie hat auch im Gesundheitswesen zur Anhäufung riesiger Datenmengen geführt, die stetig wachsen und unter dem Begriff Big Data gefasst werden. Die Schlussfolgerung, dass sich mit Big Data zwingend auch das Wissen vergrössert, ist allerdings falsch und führt zu gefährlichen Trugschlüssen (siehe auch: http://en.wikipedia.org/wiki/Ecological_fallacy). An sich sagen Daten nichts aus; hier verhält es sich wie mit dem Telefonbuch: auch wenn wir alle Nummern mit zugehörigen Namen kennen, wissen wir noch immer nichts über die Menschen, denen diese Nummern gehören. Die Frage ist also nicht die, wie grosse Datenmengen erfasst sind, sondern wie akkurat diese Daten Wirklichkeit abbilden.

Problem eins: lückenhafte Datengrundlagen

Eine Studie, welche die Ursachen von Lungenkrebs untersuchen will und auf einer Datengrundlage fusst, in der keine Variablen zu den Rauchgewohnheiten erfasst sind, kann noch so sorgfältig ausgearbeitet sein, sie wird aus beobachteten Korrelationen wie beispielsweise Wohnkanton, Alter und Beruf mit der Häufigkeit von Lungenkrebs zwar Hypothesen generieren können, die vielleicht sogar interessant sind, doch keinen Wissensgewinn schaffen. Hier liegt das Problem vieler Studien unseres Gesundheitswesens: sie arbeiten mit Datengrundlagen, die eben gerade billig verfügbar sind, in denen wichtige, die behaupteten Zusammenhänge begründende Variablen aber gar nicht erfasst sind.

Problem zwei: aggregierte Daten

Bei der Aggregation werden Daten einer bestimmten Gruppe zugeordnet, wobei allgemeine Aussagen über die gesamte Gruppe gemacht werden. Im obigen Beispiel etwa könnte man die Gruppe aller Bauarbeiter zusammenfassen und ihnen aufgrund der Beobachtung einer bestimmten Häufigkeit im Mittelwert eine höhere Anfälligkeit attribuieren, an Lungenkrebs zu erkranken, ev. aufgrund hoher Staubemissionen. Dies wäre allerdings eine unzulässige Aggregation, weil sie die Wirklichkeit ungenügend abbildet. Dies tut die Variable Rauchgewohnheit besser, die eine Zusammenfassung aller Raucher erlaubt, um dann zu untersuchen, mit welcher Häufigkeit diese an Lungenkrebs erkranken. Ohne den Einschluss der Variable Rauchgewohnheit in die Studie, welche Staubemissionseffekte auf die Lungenkrebs-Inzidenz untersucht, wird der Effekt der Staubemission wegen der Aggregation der Daten massiv überschätzt (aggregation bias).

Beispiele Aggregation

Typische Beispiele für den Aggregationsbias im Gesundheitswesen sind Mittelwerte und Regressionsanalysen. Das Mass für den Erklärungsgehalt eines Mittelwertes existiert nicht. Je kleiner die Standardabweichung (die Streubreite um den Mittelwert) ist, desto geringer ist der Täuschungseffekt. In der Regressionsanalyse tritt an die Stelle der Standardabweichung der Determinationskoeffizient r2. Je kleiner dieser ist, desto schlechter ist die Beziehung zwischen der abhängigen Variable (z.B. Kosten pro Patient) und der erklärenden, unabhängigen Variable (z.B. Alter des Patienten). Bei den Versicherern erklärt das Alter pro einzelnen Patienten höchstens 4% der Kosten. Aggregiert man diese Patienten zu einem Mittelwert in Altersklassen (z.B. Alter 40-44, 45-49 etc.), so steigt das r2 auf sehr hohe Werte (z.B. 40%). Damit wird der Erklärungsgehalt auf die Kosten pro Patient um einen Faktor 10 überschätzt. Ein klassischer Aggregationstrugschluss kann ferner anhand des Simpsons-Paradoxes erkannt werden: http://de.wikipedia.org/wiki/Simpson-Paradoxon.

Forderungen

Der VEMS hat beim Ethikrat für Statistik bezüglich im obigen Sinne fehlerhafter, unwissenschaftlicher und irreführender Studien interveniert, doch dieser fühlt sich dafür nicht zuständig, ebenso wenig wie die Schweizerische Akademie der Medizinischen Wissenschaften SAMW und die Akademien der Schweiz. Diese Institutionen sind aber per definitionem für die Wahrung wissenschaftlicher Standards verantwortlich und sollen ihre Verantwortung endlich konsequent wahrnehmen und Studien, die mit aggregierten Beobachtungsdaten arbeiten, als das kennzeichnen, was sie häufig sind: reine Hypothesen, Wissenschaft, die kein Wissen schafft.

Weitere VEMS-Factsheets: www.vems.ch/fakten-und-standpunkte

Verein Ethik und Medizin Schweiz, Ziegelfeldstrasse 1, 4600 Olten
Tel: 062 212 44 10, www.vems.ch