



AUS DEM NETZWERK EVIDENZBASIERTE MEDIZIN

Wie kategorielle Klassifizierung in die Irre führt

Am Beispiel Antidepressiva und Diabetesprävention

VON INGRID MÜHLHAUSER IM AUFTRAG DES NETZWERKS EVIDENZBASIERTE MEDIZIN E. V.
(WWW.EBM-NETZWERK.DE)

IN

Nach Analysen der BARMER-Krankenkasse ist bereits jeder dritte junge Berliner wegen psychischer Probleme in ärztlicher Behandlung. Die Diagnose Depression würde besonders häufig gestellt (1). Die Techniker Krankenkasse verweist in ihrem Gesundheitsreport 2018 auf eine Verdoppelung der Verordnungen an Antidepressiva während der letzten zehn Jahre (2). Die Krankenkassenexperten gehen jedoch davon aus, dass nicht die Krankheit Depression zugenommen hat, sondern die Diagnoseraten. Die Kampagnen der letzten Jahre zur Erhöhung der Aufmerksamkeit auf Depression scheinen Erfolg zu haben – zumindest gemessen an der enormen Zunahme an Diagnosen und Behandlungen. Die Frage nach dem Nutzen für die Betroffenen bleibt dennoch unbeantwortet. Wird den Diagnostizierten wirklich geholfen oder entsteht hier durch Überdiagnostik und Übertherapie ein Schaden an gesunden Menschen?

Die Wirksamkeit von Antidepressiva ist jedenfalls nach wie vor Gegenstand kontroverser Diskussion. Der Nutzen bei leichter und mittelschwerer Depression bleibt fraglich. Da schien die Botschaft einer kürzlich im Fachjournal LANCET publizierten Meta-Analyse zur Wirksamkeit von Antidepressiva bei Akutbehandlung schwerer Depressionen (3) sehr willkommen. Das *Deutsche Ärzteblatt* titelte dazu auf seiner News-Website „Antidepressiva wirken ohne Ausnahme besser als Placebo“ (4). Selbst in der eher kritischen *Süddeutschen Zeitung* wurde der Tenor der Studie unhinterfragt mit der Überschrift „Antidepressiva sind besser als ihr Ruf“ rezitiert (5).

Lediglich das unabhängige *arznei-telegramm* hat sich die Mühe gemacht, die Publikation im LANCET kritisch zu analysieren. In dem vorbildlich recherchierten Artikel „Antidepressiva sind besser als ihr Ruf? Viel Rummel um neue Meta-Analyse“ zeigt die Redaktion die Limitierungen der Meta-Analyse und verdeutlicht die Trugschlüsse, die sich aus einer missverständlichen Datenpräsentation ergeben (6). Die folgenden Ausführungen referieren einen Teil dieser wertvollen Aufarbeitung des *arznei-telegramms*.

Die wichtigste Frage ist: Woran wird der Erfolg der Behandlung beurteilt? Bei akuter schwerer Depression sollen Antidepressiva Suizid verhindern. Das wäre somit der primäre Endpunkt zur Beurteilung des Nutzens von Antidepressiva. Eine Suizid-verhütende Wirkung ist jedoch weiterhin für kein Antidepressivum belegt. Hingegen wird in der LANCET-Publikation die Wirksamkeit der Antidepressiva an einem Surrogat, einem sogenannten

Ersatzparameter, dem „Ansprechen“ der Patienten auf die Behandlung, beurteilt. Das Kriterium „Ansprechen“ gilt üblicherweise als erfüllt, wenn es zu einer mindestens 50-prozentigen Reduktion der Punktzahl auf einer standardisierten Symptomskala kommt. Zumeist wird die Hamilton-Depression-Skala (HAM-D-17) eingesetzt. Die Skala hat eine Spanne von 0 bis 52 Punkten, eine schwere Depression liegt bei einer Punktzahl ab 25 vor.

Der HAM-D-Test ist eine Fremdbeurteilungsskala und umfasst unter anderem depressive Stimmung, Schlafstörungen und Schuldgefühle. Besserungen in wenigen Bereichen (z.B. Schlafstörungen) können ein globales Ansprechen auf die Therapie vortäuschen. Zudem kommt es auch bei „Non Responder“ häufig zu einer Besserung, ebenso wie unter Placebo. Das Ausmaß der Besserung ist dann aber eben oft weniger als das geforderte Ansprechkriterium von 50 Prozent. In der LANCET-Arbeit war der Unterschied in der Ansprechrate zwischen allen untersuchten Antidepressiva und Placebo durchweg statistisch signifikant.

Die entscheidende Frage ist jedoch, ob die Unterschiede auch klinisch relevant sind. Dazu müssen die Punktunterschiede auf der HAM-D-Skala zwischen den Therapiegruppen beurteilt werden. Die Autoren des arznei-telegramm-Artikels betonen, dass sie die nötigen Informationen hierzu erst mühsam aus Grafiken ableiten mussten, die nur in dem 290 Seiten langen Appendix zur Hauptpublikation zu finden wären. Sie kommen zu dem Ergebnis, dass die ohnehin kleine Effektstärke von im Mittel etwa 0,3 (Spannbreite: 0,17 unter Reboxetin bis 0,48 unter Amitriptylin) umgerechnet nur einem Unterschied von ungefähr 1,8 bis 2,5 Punkten auf der HAM-D-Skala (0-52 Punkte) entspricht. [Anmerkung: eine Effektstärke ab 0,2 entspricht einem schwachen, ab 0,5 einem mittleren und ab 0,8 einem starken Effekt.] Der Unterschied zu Placebo von ungefähr 1,8 bis 2,5 Punkten auf der HAM-D-Skala (der mittlere Basiswert war 25,7) – wäre demnach klinisch nicht relevant. Ärzte könnten eine Veränderung um 3 Punkte auf der HAM-D-Skala nicht als klinische Besserung erkennen. Dies wäre üblicherweise erst ab einem Unterschied von 10 Punkten auf der HAM-D möglich (6).

Als ein weiteres Beispiel zu Trugschlüssen, die sich aus kategoriellen Klassifizierungen ergeben, seien hier die Diabetespräventionsstudien angeführt. Schon vor Jahren hat unsere Arbeitsgruppe dazu Untersuchungen mit Diabetesexperten aus Deutschland und anderen europäi-

schen Ländern durchgeführt (7,8). Inzwischen liegen mehrere qualitativ gute randomisiert-kontrollierte Studien und ein aktueller Cochrane-Review zur Auswirkung von Lebensstiländerung auf das Auftreten von Typ-2-Diabetes vor. Die Dauer der Intervention lag zwischen 2 bis 6 Jahren (9). Die ersten Ergebnisse für eine kombinierte Intervention aus Diät und körperlicher Aktivität wurden eindrücklich kommuniziert. Typ-2-Diabetes ließe sich um 30 bis 60 Prozent reduzieren. Sogar das LANCET brachte diese Information schlagzeilenartig auf das Titelblatt einer Ausgabe (Heft 9702, Band 374, Jahrgang 2009). Auch wenn die „60 Prozent“ relative Risikoreduktionen sind, so ist auch die absolute Risikoreduktion beeindruckend. Der Unterschied im Auftreten von Diabetes zwischen Interventions- und Kontrollgruppe kann durchaus mehr als 10 Prozentpunkte betragen (7,9).

Auch hier geht es um kategorielle Daten – Diabetes „ja“ oder „nein“. Genau genommen sind es Diagnosestellungen, es müsste also heißen „Diabetesdiagnose ja“ oder „Diabetesdiagnose nein“.

Entscheidend für die Prognose der Patienten sind jedoch die den Diagnosen zugrundeliegenden metabolischen Veränderungen. Für Diabetesdiagnosen sind es die Blutzuckerwerte. Sehr hohe Blutzuckerwerte über viele Jahre sind ursächlich für die gefürchteten diabetesspezifischen Komplikationen am Auge, den Nieren und Nerven verantwortlich.

In den Diabetespräventionsstudien betrug der Unterschied zwischen den Interventions- und Kontrollgruppen im Mittel jedoch maximal 5 Milligrammprozent, die Differenz im HbA_{1c}-Wert etwa 0,1 Prozent.

Wie kann es sein, dass minimale Unterschiede in den Blutzuckerwerten zu eindrucksvollen Unterschieden in den Diagnoseraten führen?

Mehrere Faktoren sind hier von Bedeutung. In die Studien wurden Personen mit gestörter Glukosetoleranz eingeschlossen, sie hatten gerade noch keinen Diabetes. Durch die intensive Lebensstilintervention betrug der Unterschied im Körpergewicht zwischen den Gruppen bis Studienende durchschnittlich etwa 2 bis 4 kg. Das reicht, um einen Unterschied im Blutzucker um einige Milligrammprozent zu erreichen. Entsprechend hatten in der Interventionsgruppe deutlich mehr Personen ihren Blutzucker gerade noch unter der Schwelle zur Diabetesdiagnose halten können, in der Kontrollgruppe hatten mehr Personen diese Schwelle überschritten. →



→ **DIABETESDIAGNOSEN UND METABOLISCHE PARAMETER**
aus randomisiert-kontrollierten Studien zur Prävention von Typ-2-Diabetes durch Lebensstiländerung

	Interventionsgruppe	Kontrollgruppe
Nüchternblutzucker (mg%)		
Beginn der Studie	109	110
Ende der Studie	107	113
HbA1c Wert (Ende der Studie)	6,0%	6,1%
Personen mit einer Diabetesdiagnose (Ende der Studie)	14%	29%

Anmerkung: Die Daten sind aus unterschiedlichen Studien, es geht um eine beispielhafte Gegenüberstellung von metabolischen Parametern und Diagnosen (Detailangaben siehe Ref. 7 und 9).

Noch bedeutsamer als die Diabetesdiagnosen oder metabolischen Veränderungen ist die Frage, ob die Intervention auch zu einer Reduzierung der kardiovaskulären

ren Komplikationen führt. Die Lebenserwartung dieser Hochrisikopatienten ist vor allem durch vorzeitigen Tod an Herz-Kreislaufkrankungen eingeschränkt. Bisher ist es nicht gelungen, einen solchen Nachweis zu führen. Auch der Cochrane-Review betont, dass ein Beleg für die Verbesserung der Lebensprognose fehlt. Weder eine Verminderung diabetesspezifischer (=Hyperglykämie-abhängiger) noch kardiovaskulärer Komplikationen ließe sich bisher nachweisen (8).

Fazit: Kategorielle Klassifizierungen werden in vielen Bereichen der Medizin benutzt, zum Beispiel auch bei der Diagnostik von Demenz/Alzheimer, der Bewertung von Schmerz oder körperlichen Funktionsstörungen. Es lohnt sich nachzufragen, welche klinische Bedeutung Veränderungen von Skalenwerten haben, selbst wenn Unterschiede statistisch hoch signifikant sein mögen. Dies gilt auch dann, wenn Studien in renommierten Wissenschaftsjournalen mit prominenter Autorenschaft publiziert werden. ■



Univ.-Prof. Dr. med. Ingrid Mühlhauser
Universität Hamburg
MIN Fakultät
Gesundheitswissenschaften
E-Mail: Ingrid_Muehlhauser@uni-hamburg.de
Tel: 040 42838 3988

1. Jeder dritte junge Berliner wegen psychischer Probleme beim Arzt. DÄB Nachrichten, 17. Mai 2018
2. Antidepressiva auf dem Vormarsch. DÄB Nachrichten, 26. Juli 2018
3. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018; 391: 1357–66
4. Antidepressiva wirken ohne Ausnahme besser als Placebo. DÄB Nachrichten, 22. Februar 2018
5. Antidepressiva sind besser als ihr Ruf. Süddeutsche Zeitung, 22. Feb. 2018
6. Antidepressiva sind besser als ihr Ruf? Viel Rummel um neue Meta-Analyse. arznei-telegramm a-t 2018; 49: 41-3
7. Mühlhauser I, Kasper J, Meyer G, Federation of European Nurses in Diabetes. Understanding of diabetes prevention studies: questionnaire survey of professionals in diabetes care. Diabetologia 2006; 49: 1742-46
8. Mühlhauser I. Diabetes experts' reasoning about diabetes prevention studies: a questionnaire survey. BMC Research Notes 2008; 1: 90
9. Hemmingsen B, Gimenez-Perez G, Mauricio D, Roqué i Figuls M, Metzendorf MI, Richter B. Diet, physical activity or both for prevention or delay of type 2 diabetes mellitus and its associated complications in people at increased risk of developing type 2 diabetes mellitus. Cochrane Database of Systematic Reviews 2017, Issue 12. Art. No.: CD003054. DOI: 10.1002/14651858.CD003054.pub4.