

Systematische Fehler in Beobachtungsstudien am Beispiel von Digoxin

Den Einfluss von Digoxin bei Patienten mit Vorhofflimmern und/oder Herzinsuffizienz auf die Häufigkeit von Krankenhauseinweisungen und die Gesamtsterblichkeit untersuchten Oliver Ziff und Kollegen 2015 in einer Metaanalyse (1). Dabei analysierten sie Qualität der eingeschlossenen Studien und deren Fehleranfälligkeit aufgrund systematischer Verzerrungen (Bias) durch Störfaktoren (Confounder). Ausgewertet wurden 52 Studien mit mehr als 620.000 Patienten. In den Beobachtungsstudien unterschieden sich die Patienten, die Digoxin erhalten hatten, von den Kontrollgruppen: Sie waren älter, hatten häufiger Diabetes, eine schlechtere kardiale Auswurfraction und nahmen häufiger Diuretika und Antiarrhythmika ein. Dass ältere und kränkere Patienten Digoxin erhielten, spricht für einen Verordnungsbias (Prescription bias) oder eine Verzerrung durch Indikation (Confounding by indication). In der Metaanalyse zeigte sich, dass diese Verzerrungen nicht durch statistische Methoden gelöst werden konnten. Je größer das Verzerrungspotenzial der Studie war, desto stärker zeigte sich eine höhere Sterblichkeit unter Digoxin. Die Verzerrung war am größten in nicht risikoadjustierten Beobachtungsstudien, fand sich aber auch in risikoadjustierten Beobachtungsstudien sowie ebenfalls in Beobachtungsstudien, bei denen durch ein sogenanntes propensity matching systematische Unterschiede oder Verzerrungen zwischen den Vergleichsgruppen kontrolliert werden sollten. Bei den randomisierten kontrollierten Studien (RCT) fand sich dagegen kein Einfluss von Digoxin auf die Sterblichkeit. Unabhängig vom Studiendesign war Digoxin mit einer kleinen, aber statistisch signifikanten Reduktion von Krankenhauseinweisungen assoziiert.

Die Autoren folgerten, dass Beobachtungsstudien nur zur Generierung von Hypothesen verwendet werden sollten. Ein begleitendes Editorial mit dem provokativen Titel „Trials are best, ignore the rest“ führt weitere Beispiele auf, bei denen sich die Evidenz aus nicht randomisierten Studien und RCT widerspricht, darunter die Behandlung von ventrikulären Extrasystolen nach Herzinfarkt durch Antiarrhythmika (2). Auch für die Hormontherapie bei Frauen in der Postmenopause wurden aus Beobachtungsstudien positive Wirkungen abgeleitet, die in RCT nicht nachweisbar waren (3).

RCT gelten gemeinhin als „Goldstandard“ in der klinischen Forschung (vgl. 4). Für einige Fragestellungen sind epidemiologische Beobachtungsstudien jedoch unverzichtbar, wie beispielsweise die Untersuchung des Einflusses von Zigarettenkonsum auf die Entstehung von Lungenkrebs. Während in RCT durch Randomisierung eine gleiche Verteilung bekannter und unbekannter Störfaktoren in den zu vergleichenden Gruppen erreicht werden soll, ist dies in Beobachtungsstudien selten möglich (5). Das kann zu systematischen Verzerrungen und fehlerhaften Ergebnissen führen. Systematische Fehler werden von englischsprachigen Epidemiologen als

systematic error oder als *bias* bezeichnet (6). Bias kann so stark sein, dass er einen Nutzen vortäuscht, wo in Wirklichkeit sogar ein Schaden vorliegt.

Zu den wichtigsten systematischen Fehlern in Beobachtungsstudien gehören bei der Auswahl von Studienteilnehmern und bei der Erhebung von Informationen der Selektionsbias (Verzerrung durch ungeeignete Auswahl der Teilnehmer) und der Informationsbias (Verzerrung durch unzutreffende Informationen). Ein Selektionsbias tritt auf, wenn sich die für die Studie ausgewählten Menschen systematisch von der Bevölkerung unterscheiden, aus der die Auswahl erfolgte (5). Zu Selektionsfehlern kann es kommen, wenn kontaktierte Personen nicht antworten oder die Teilnahme an einer Studie verweigern (Nonresponse bias) sowie ebenfalls, wenn sich Teilnehmer freiwillig für eine Studie melden (Freiwilligenbias). Auch wenn Studienteilnehmer im Verlauf einer Studie verloren gehen (loss to follow up) und sie sich deutlich von den Teilnehmern unterscheiden, die in der Studie verbleiben, kann es zu starken Verzerrungen der Studienergebnisse kommen.

Der Informationsbias entsteht, wenn Messungen oder Klassifikationen einer Krankheit oder einer Exposition in den verschiedenen Studiengruppen ungenau oder fehlerhaft sind. Eine Form des Informationsbias ist der Erinnerungsbias (Recall bias). Beispielsweise erinnern sich Mütter von Kindern mit Fehlbildungen besser an mögliche Risikofaktoren während der Schwangerschaft als Frauen mit einem gesunden Kind.

Unter einem Störfaktor in einer Studie versteht man einen nicht berücksichtigten Faktor, der das Ergebnis jedoch beeinflussen kann. Ein Beispiel: Es wird ein Zusammenhang zwischen Alkoholkonsum und Lungenkrebs hergestellt und dabei nicht beachtet, dass Alkohol trinkende Personen auch mehr rauchen. Das Rauchen ist in diesem Beispiel der Störfaktor, während der Alkoholkonsum ein Risikoindikator ist. Confounding wird auch als Analysebias bezeichnet. Im Unterschied zum Bias, der nachträglich kaum zu beheben ist, beruht Confounding jedoch auf einem Fehler in der Interpretation der Daten, der auch nachträglich erkannt und bei der Datenauswertung berücksichtigt werden kann, wenn er in der Studie gemessen wurde (6).

Bei Beobachtungsstudien können noch eine Reihe weiterer Fehler auftreten; eine Übersicht findet sich im Deutschen Ärzteblatt (5). Hier wurden auch hilfreiche Fragen zur Beurteilung von Beobachtungsstudien zusammengestellt:

- Ist das Studienkollektiv repräsentativ?
- Sind die betrachteten Teilkollektive vergleichbar?
- Wurden die Informationen in vergleichbarer Weise erhoben?
- Werden potenzielle Messfehler beschrieben?
- Berücksichtigt das Studiendesign mögliche Fehlerquellen?

- Wie gut ist die Qualität der erhobenen Daten?
- Werden Korrekturverfahren angewendet?

Fazit: Beobachtungsstudien können sinnvoll sein, wenn es um lange Zeiträume oder seltene Ereignisse geht bzw. eine Randomisierung unethisch wäre. Weil sie durch nicht quantifizierbare Verzerrungen mit erheblichen Unsicherheiten behaftet sind (3), können sie die Ergebnisse randomisierter kontrollierter Studien jedoch nicht in Frage stellen oder ersetzen. Dies ist aber beim Beispiel Digoxin exemplarisch geschehen.

Literatur

1. Ziff, O., et al.: BMJ 2015, **351**, h4451. [Link zur Quelle](#)
2. Cole, G., et al.: BMJ 2015, **351**, h4662. [Link zur Quelle](#)
3. Mühlhauser, I.: KVH-Journal 1/2016. [Link zur Quelle](#)
4. AMB 2017, **51**, 64DB01. [Link zur Quelle](#)
5. Hammer, G.P., et al.: Dtsch. Arztebl. Int. 2009, **106**, 664. [Link zur Quelle](#)
6. Razum, O., et al.: Epidemiologie für Dummies 2009. Wiley-VCH Verlag GmbH Co. KGaA, Weinheim.