# DER ARZNEIMITTELBRIEF

**Jahrgang 48 · Nr. 7 · Juli 2014** **www.der-arzneimittelbrief.de** **Ausgabe Deutschland**

## Von Subgruppen und statistischen Assoziationen

1988 erschien im Lancet eine der einflussreichsten Studien in der Kardiologie, die ISIS-2-Studie. Bei über 17.000 Patienten wurden verschiedene gerinnungshemmende Strategien beim akuten Myokardinfarkt miteinander verglichen: Streptokinase, Streptokinase plus ASS, nur ASS sowie Plazebo-Kontrolle. Ein Ergebnis war, dass ASS eine hochsignifikant positive Wirkung hatte hinsichtlich der Prävention eines Plötzlichen Herztods. Bei der Analyse der fast 40 Subgruppen fanden die Statistiker u.a. heraus, dass ASS bei Patienten mit dem Sternzeichen Waage oder Zwilling nicht wirksam war (9% höhere Letalität). Dieser Befund war zwar statistisch signifikant, aber wissenschaftlich nicht plausibel. Generell wird die Interpretation der Ergebnisse von Subgruppenanalysen auf Studienebene durch drei Aspekte erschwert (1):

- Subgruppenanalysen werden selten a priori geplant und haben deshalb häufig keinen Beweischarakter.
- Bei der Analyse zahlreicher Subgruppen ("Multiples Testen") besteht mitunter eine recht hohe Wahrscheinlichkeit, dass die Ergebnisse irgendeiner Subgruppe statistische Signifikanz erreichen, obwohl es sich um ein zufälliges Ergebnis handelt. Dies kann zu schwerwiegenden Fehlinterpretationen führen.
- Oft erreichen Subgruppen nicht die Größe von Stichproben, die für das Aufdecken moderater Unterschiede nötig ist, d.h. sie haben eine geringe Power.

Statistiker und Epidemiologen diskutieren deshalb Subgruppenanalysen häufig sehr kritisch und warnen vor falscher Interpretation dieser Ergebnisse (1-3). Die Chance, dass zufällig ein „positiver" Befund dabei ist, steigt natürlich mit jeder zusätzlich berechneten Subgruppe. Es gibt übrigens statistische Tests, um solchen „Multiplizitätsproblemen" auf die Schliche zu kommen. Diese werden jedoch nach wie vor selten angewendet. So findet man in der RE-LY-Studie nicht weniger als 38 Analysen zu Subgruppen, ohne dass im Methodikteil der Publikation ein Korrekturverfahren für multiples Testen genannt wird (4). Bei der Interpretation randomisierter, kontrollierter Studien (RCT) sollte man auch immer bedenken, dass manche Autoren dazu neigen, aus ihrem Datenschatz post hoc viele, nicht vordefinierte Subgruppen zu analysieren. Für die Publikation picken sie sich dann die Rosinen heraus (selektive Darstellung der Ergebnisse). Auf dieses Phänomen stößt man besonders dann, wenn das primäre Studienziel nicht erreicht wurde und die Studie irgendwie doch noch ein positives Ergebnis zeigen soll. Eine solche Kaffeesatzleserei aus den Subgruppen kann man gerade bei der negativ verlaufenen SYMPLICITY-HTN-3-Studie (5, 6) zur renalen Denervation bei refraktärer arterieller Hypertonie beobachten. Generell gilt also: Vorsicht bei Subgruppenanalysen!

In einer retrospektiven, bevölkerungsbasierten Kohortenstudie konnte jüngst nachgewiesen werden, dass Iren, die den Familiennamen Brady (sic!) tragen, ein signifikant höheres Risiko für die Implantation eines Schrittmachers haben als Iren mit anderem Familiennamen (7). Für diese Erkenntnis haben Wissenschaftler aus Dublin Patienten der Stadt untersucht, die in den Jahren 2007 bis Februar 2013 an der Universitätsklinik einen Herzschrittmacher implantiert bekommen hatten (n = 1012). 13 Patienten wurden von den weiteren Analysen ausgeschlossen, weil wichtige klinische Daten fehlten. Das Durchschnittsalter der verbliebenen 999 Patienten war 77 Jahre, 55,8% waren Männer. Acht der 999 Patienten trugen den Familiennamen Brady (0,8%). Ein Abgleich mit dem Dubliner Telefonbuch ergab, dass im Einzugsbereich der Klinik insgesamt 579 Menschen mit dem Namen Brady leben und 161.388 Menschen mit anderen Namen. Hieraus errechnet sich eine Schrittmacher-Prävalenz von 1,38% bei den Bradys und 0,61% bei den Nicht-Bradys. Das Risiko, einen Herzschrittmacher zu erhalten, ist also für einen Dubliner namens Brady mehr als doppelt so hoch wie für einen Nicht-Brady: die nicht-adjustierte Odds-Ratio beträgt 2,27 (95%-Konfidenzintervall: 1,13-4,57; p = 0,03). Die Autoren interpretieren diesen statistisch eindeutigen Befund augenzwinkernd dahingehend, dass der Familienname möglicherweise eine größere Rolle bei der Indikationsstellung zur Schrittmachertherapie spielt als bislang angenommen und schaffen dafür auch einen wissenschaftlichen Terminus: „nominativer Determinismus". Als möglicher Grund für diese Beobachtung wird eine über viele Generationen weitervererbte Mutation im kardialen HCN4-Ionenkanal diskutiert, die zu Bradykardien führt. Ein Leserbriefschreiber diskutiert auch die Möglichkeit einer medizinischen Selektion: allein der Name Brady könnte Ärzte daran denken lassen, dass diese Patienten bei unklaren Symptomen an bedeutsamen Arrhythmien leiden und verstärkt danach suchen. Dies führt möglicherweise zu einer höheren Überlebenswahrscheinlichkeit der Bradys („survival of the Bradys"). Diese Möglichkeit ist nicht ausgeschlosen.

An diesem Beispiel sieht man, dass statistische Assoziationen aus Kohortenstudien oder Registern – so wichtig sie sein mögen – in ihrer klinischen Bedeutung nicht überinterpretiert werden dürfen, auch wenn sich Theorien zu pathophysiologischen Erklärungen oder Plausibilitäten im Nachhinein generieren lassen.

Als gutgläubiger Leser von Studien fühlt man sich oft der Beweiskraft der Statistik ausgeliefert und unterscheidet häufig nicht zwischen signifikant und relevant. Die Beispiele zeigen darüber hinaus, dass statistisch signifikante Assoziationen oder Ergebnisse aus Subgruppen nicht unbedingt der klinischen „Wahrheit" entsprechen müssen. Die medizinische Statistik beruht auf der Abschätzung von Wahrscheinlichkeit, wobei Irrtum bei einer definierten Zahl von „Fällen" bedacht werden muss. Vor allem Ergebnisse aus den definitionsgemäß kleineren Subgruppen sind fehleranfällig und dienen – nicht selten unter dem Einfluss spezieller Interessen – dazu, neue Hypothesen zu generieren. Diese müssen dann in weiteren Studien überprüft werden. Ärzte und Gesundheitsökonomen müssen sich mehr mit der Statistik auseinandersetzen, um nicht (den manchmal gezielten) Fehlinterpretationen aufzusitzen. In diesem Sinne: Bleiben Sie kritisch, auch gegenüber signifikanten p-Werten!

**Literatur**

1. https://www.iqwig.de/download/IQWiG_Methoden_Entwurf-fuer-Version-4-2.pdf
2. Rothwell, P.M.: Lancet 2005, **365**, 176.
3. Schulz, K.F., und Grimes, D.A.: Z. Arztl. Fortbild. Qual. Gesundh.wes. 2007, **101**, 51. http://www.ebm-netzwerk.de/pdf/zefq/schulz-epidemiologie5.pdf
4. Conolly, S.J., et al. (RE-LY = **R**andomized **E**valuation of **L**ong-term Anticoagulation therap**Y**): N. Engl. J. Med. 2009, **361**, 1139. Erratum: N. Engl. J. Med. 2010, **363**, 1877. Vgl. AMB 2010, **44**, 6. AMB 2011, **45**, 7. AMB 2014, **48**, 41.
5. Persu, A., et al.: Curr. Hypertens. Rep.: 2014, **16**, 460.
6. Bhatt, D.L., et al (SYMPLICITY HTN-3): N. Engl. J. Med. 2014, **370**, 1393. Vgl. AMB 2014, **48**, 16.
7. Keaney, J.J., et al.: BMJ 2013, **347**, f6627.