

Statistisch signifikante Resultate in Abstracts: wenig glaubhaft!

a -- Gotzsche PC. Believability of relative risks and odds ratios in abstracts: cross sectional study. BMJ 2006 (29. Juli); 333: 231-4

[\[LINK\]](#)

Zusammenfassung: Renato L. Galeazzi

Es wird allgemein angenommen, dass vor allem Studien mit positiven Resultaten, d.h. mit einem p-Wert unter 0,05, publiziert werden. Da aber in den meisten Studien nicht nur statistisch signifikante Resultate erhoben werden und das Abstract der am meisten gelesene Teil einer Publikation ist, hat sich der Autor hier zum Ziel gesetzt, die Glaubwürdigkeit der im Abstract publizierten p-Werte zu untersuchen.

Er untersuchte 260 Abstracts von randomisiert-kontrollierten Studien und je 130 von Kohorten-Studien und Fall-Kontroll-Studien. Er fand signifikante Resultate in 70%, bzw. 84% und 84%. Das sind viel mehr als bei einer ausgeglichenen Hypothese erwartet würde. Besonders interessant war, dass in 5 Studien ein p-Wert unter 0,06 aber über 0,05 gefunden wurde und in 46 ein p-Wert unter 0,05 aber über 0,04. Von den letzteren konnten 23 nachkontrolliert werden, wobei nur 8 sicher unter 0,05 zu liegen schienen. Bei den Kohorten- und Fall-Kontroll-Studien war die Situation noch einseitiger: In allen 17 Studien, in denen sich ein p-Wert über 0,04, aber unter 0,06 fand, betrug der p-Wert weniger als 0,05 und war damit signifikant. In drei Fällen wurden die Vertrauensintervalle mit der Grenze 1,0 angegeben, was als signifikant angesehen wurde, obwohl zum Beispiel in einem Fall dieses Resultat das einzig «signifikante» war von sechs Vergleichen.

Es ist gefährlich, wenn man sich beim «Scannen» von Journals auf die Abstracts verlässt. Es scheint, dass hier noch mehr beschönigt und irgendwelche signifikanten Resultate als Blickfang verwendet werden als im nachfolgenden Volltext. Der häufigste Fehler dürfte wahrscheinlich die Unzahl von möglichen Vergleichen sein, die in Forschungsprotokollen jeweils vorgeschlagen werden. Selten liest man später von durchgeführten Korrekturen, obwohl diese absolut nötig sind, denn die offizielle Signifikanzgrenze ist nur für eine einmalige Untersuchung definiert. Sobald mehr als ein Vergleich gemacht wird, muss die Grenze viel tiefer gesetzt werden. Bei nicht Befolgen dieser Regel wird viel zu häufig etwas als evidenzbasiert angegeben, was diesen Namen in Wirklichkeit gar nicht verdient.

Zusammengefasst von Renato L. Galeazzi