

Lars Erik Mihan

Dr. med.

Die Bewertung der methodischen Qualität von kontrollierten Studien mittels Indices - Untersuchung anhand von Publikationen der „Medizinischen Klinik“

Geboren am 01.08.1971

Reifeprüfung am 14.05.1990 in Osterburken

Studiengang der Fachrichtung Medizin von SS 1992 bis SS 1999

Physikum am 25.03.1994 an der Universität Heidelberg

Klinisches Studium in Heidelberg

Praktisches Jahr in Heidelberg

Staatsexamen am 23.04.1999 an der Universität Heidelberg

Promotionsfach: Medizinische Biometrie

Doktorvater: Prof. Dr. med. J. Windeler

Über die Anforderungen, die eine kontrollierte Studie erfüllen sollte, ist in den letzten Jahrzehnten viel publiziert worden. Es gibt einige Checklisten, die es Medizinern erleichtern sollen, methodisch einwandfreie Studien zu planen bzw. die Qualität von Studien zu bewerten.

Die vorliegende Arbeit hatte zwei Fragestellungen:

1. Inwieweit wurden diese Qualitätskriterien in den Publikationen einer anerkannten deutschen Fachzeitschrift erfüllt ?
2. Eignen sich Indices für die Qualitätsbeurteilung klinischer Studien ?

Als Material dienten kontrollierte Studien, die zwischen 1979 und 1996 in der Zeitschrift „Medizinische Klinik“ veröffentlicht wurden. Durch eine Handsuche wurden 132 Studien gefunden, die den Einschlußkriterien entsprachen. Mit Hilfe eines Erhebungsbogen wurde untersucht, ob ausgewählte Kriterien bezüglich Planung, Durchführung und statistischer Analyse in den Studien erfüllt wurden.

Es ließ sich feststellen, daß viele Qualitätskriterien nur unzureichend erfüllt wurden. Fallzahlplanungen wurden nur für 6 der 132 Studien (5%) angestellt, Randomisierung war in 107 Publikationen (81%) erwähnt, nähere Angaben zur Art der Randomisierung in 20 Publikationen (15%). Ein Vergleich der Gruppen vor Behandlungsbeginn wurde für 52

Studien (39%) vorgenommen, in 57 Publikationen (43%) die Verblindung von Patient und Untersucher erwähnt. Angaben über Studienabbrüche fanden sich in 87 Publikationen (66%). Die Angabe der Untersuchungsergebnisse war in 96 Publikationen (73%) ausreichend, in den übrigen fehlten entweder die Standardabweichungen oder die Daten waren nur ungenau aus Diagrammen abzulesen. Für 91 Studien (69%) wurden Signifikanztests bezüglich des Gruppenvergleiches durchgeführt, in 24 dieser Publikationen (26%) waren diese Tests nachvollziehbar, in 15 Publikationen (16%) teilweise. Die Daten der Studienabbrecher wurden bei der statistischen Analyse meist weggelassen, eine Intention-to-treat-Analyse wurde für 7 Studien (5%) angewendet.

Die Ergebnisse zeigen, daß Anforderungen, die an eine methodisch hochwertige kontrollierte Studie gestellt werden, von den Studien, die in den letzten 18 Jahren in der „Medizinischen Klinik“ veröffentlicht wurden, nicht oder nicht befriedigend erfüllt wurden. Sie sind jedenfalls den Publikationen nicht zu entnehmen. Über den gesamten Zeitraum hat sich die methodische Qualität verbessert, wobei für den letzten Zeitabschnitt aber eher eine Tendenz zur erneuten Verschlechterung auszumachen ist.

Dieses Problem betrifft nicht nur die Medizinische Klinik alleine. Auch in den renommierten englischsprachigen Zeitschriften finden sich ähnliche Ergebnisse, dort ist allerdings in den letzten Jahren eine deutlichere Verbesserungstendenz auszumachen.

Eine quantitative Abschätzung der Qualität einzelner Studien soll mit Hilfe von Indices, einer Weiterentwicklung der Check-Listen, ermöglicht werden. Bisher konnte aber noch nicht geklärt werden, wie gut die damit erzielten Ergebnisse tatsächlich die methodische Qualität einer kontrollierten Studie widerspiegeln.

In der vorliegenden Arbeit sollte die Übereinstimmung verschiedener Indices untersucht werden, außerdem ihre Fähigkeit, zwischen qualitativ besseren und schlechteren Studien zu differenzieren. Von besonderem Interesse war, inwieweit sich kürzere von umfangreicheren Indices unterscheiden.

Der oben erwähnte Erhebungsbogen diente auch als Grundlage für die Bewertung mittels der Indices. Als Ausgangspunkt erfolgte eine eigene Einteilung der Publikationen nach vorher festgelegten Qualitätskriterien in 4 verschiedene Kategorien. Verwendet wurden die Indices nach Jadad et al., Kleijnen et al., Linde et al. und Chalmers et al. In keinem der 4 Indices erreichte eine Publikation die Höchstpunktzahl, der Mittelwert lag für alle Indices unter der Hälfte aller möglichen Punkte. Die Indices erbrachten ein uneinheitliches Ergebnis, es ergaben sich große Schwankungen zwischen den Punktwerten bei Verwendung der verschiedenen Indices.

Eine zuverlässige Trennung zwischen den Publikationen der verschiedenen Kategorien erbrachte keiner der verwendeten Indices, allenfalls zwischen Kategorie II und III fand eine gewisse Differenzierung statt, am deutlichsten für die Indices nach Linde und Chalmers. Die Punktwerte waren in den verschiedenen Kategorien zwar signifikant verschieden, die Überlappungen zeigen aber, wie wenig aussagekräftig dieser Unterschied bezüglich der Trennschärfe ist.

Die Reliabilität (Übereinstimmung zwischen den Indices) und Validität (Trennschärfe zwischen Publikationen der verschiedenen Kategorien) der verwendeten Indices konnte in der vorliegenden Untersuchung nicht untermauert werden. Auch zur Frage, ob ein kürzerer Index zur Beurteilung der Qualität ausreicht, kann daher keine abschließende Bewertung erfolgen.

Auch in der Literatur führten Untersuchungen über die Reliabilität zu gegensätzlichen Ergebnissen. Bezüglich der Validität dagegen konnte in einer Arbeit für den Index nach Jadad et al. ein positives Fazit gezogen werden.

Abschließend muß festgestellt werden, daß die Indices weiterer Evaluation bedürfen und eine strukturiertere Entwicklung wünschenswert wäre.