

# Zahnmedizinische klinische Studien – Wie vertrauenswürdig sind die Ergebnisse?

*Dental Clinical Trials – How reliable are the results?*

Suchen wir eine Antwort auf eine zahnmedizinische Fragestellung, beginnt die Suche danach meist im Internet, das uns bei der Recherche mit einer Flut von Informationen konfrontiert. Auch eine Fokussierung auf wenige biomedizinische Datenbanken, wie Medline (via PubMed), ergibt häufig eine hohe Trefferzahl an Referenzen zu Studienartikeln. Nicht selten verweist die Recherche auf mehrere Studien, die zur gleichen Fragestellung durchgeführt worden sind und von denen jede zu einem anderen Ergebnis und einer anderen Schlussfolgerung gelangte. Für die Sichtung und Einschätzung der gefundenen Referenzen ist nicht nur sehr viel Zeit, sondern vor allem fachliche Expertise nötig.

Wie geht man nun mit diesen Treffern um? Welches dieser Studienergebnisse ist das richtige? Welchem der Ergebnisse können wir vertrauen?

Grundsätzlich sollte man sich zuerst mit denjenigen Studienartikeln befassen, die den höchsten Evidenzlevel [6] aufweisen: In systematischen Übersichtsarbeiten (*systematic reviews*) werden die Ergebnisse aus möglichst allen zur gleichen Fragestellung durchgeführten Einzelstudien zusammengefasst. Viele sind ergänzt durch eine Metaanalyse, mit der ein Gesamtergebnis über alle Studienartikel hinweg statistisch berechnet wurde. Dieses Gesamtergebnis gründet somit auf einer größeren Studienpopulation und ist damit aussagekräftiger als das Ergebnis einer einzelnen Studie.

Wenn zur Fragestellung eine aktuelle und methodisch gute systematische Übersichtsarbeit vorhanden ist (zur



Dr. Anette Blümle



Prof. Dr. Jens Christoph Türp

Überprüfung der Qualität solcher Arbeiten stehen u.a. die Instrumente AMSTAR <<https://amstar.ca>> [4] und ROBIS [5] zur Verfügung), erspart diese die mühevollen Eigenarbeit hinsichtlich systematischer Literaturrecherche, Studienbeurteilung und Zusammenfassung der Ergebnisse.

Sofern es dagegen zu einer definierten zahnmedizinischen Fragestellung noch keine aktuelle systematische Übersichtsarbeit gibt, muss auf die Originalstudienartikel zurückgegriffen werden. Wenn zu einer Frage mehr als eine Studie durchgeführt wurde, werden sich die Studienberichte inhaltlich in etlichen Punkten unterscheiden und wahrscheinlich auch unterschiedliche Ergebnisse erzielt haben. Verschiedene Studien werden an verschiedenen Orten, von verschiedenen Arbeitsgruppen, in verschiedenen Ländern, mit verschiedenen Teilnehmern und in verschiedenen Settings durchgeführt. Jede der Studien hat womöglich andere Ein- und Ausschlusskriterien für die Studienteilnehmer, sodass sich letzte-



Prof. Dr. Gerd Antes

re hinsichtlich Alter, Geschlecht, Komorbiditäten, sozioökonomischem Status etc. unterscheiden. All diese prognostisch relevanten Faktoren wirken sich auf das Studienergebnis aus. Jedoch sind die vorhandenen Unterschiede in den Ergebnissen keine Fehler, sondern jedes der Ergebnisse kann vertrauenswürdig sein,

und zwar für die jeweilige zugrunde liegende Methodik und Studienpopulation.

Doch es gibt auch Studien, deren Ergebnisse nicht vertrauenswürdig sind und nicht die (unbekannte) wahre Situation widerspiegeln. Grund dafür sind sogenannte Störfaktoren (*confounder*). Es handelt sich dabei um systematische Fehler, die zur Verzerrung (*bias*) eines Studienergebnisses führen und dieses vom wahren (unbekannten) Wert abweichen lassen.

**Zentrale Kriterien**

Systematische Übersichtsarbeiten schließen bevorzugt Ergebnisse aus randomisierten kontrollierten Studien (RCTs) ein. Dies hat einen Grund: Methodisch gute RCTs weisen das geringste Potenzial für Verzerrungen auf und liefern die vertrauenswürdigsten, d.h. validesten, Ergebnisse. Durch (zwei oder mehr) parallele zeitgleiche Behandlungsgruppen können Unterschiede in den Ergebnissen zwischen den Gruppen direkt auf die Behandlung zurückgeführt werden, vorausgesetzt, die Gruppen unterscheiden sich ansonsten nicht. Zeittrends, die bei einem Vorher-Nachher-Vergleich auftreten können, spielen keine Rolle.

**Selection bias**

– Randomisierung

Durch Randomisierung werden die Studienteilnehmer zufällig zu den Behandlungsgruppen zugeteilt (Abb. 1). Störfaktoren, die einen (unerwünschten) Einfluss auf das Studienergebnis haben können, werden dadurch gleichmäßig auf die Gruppen verteilt. Ist die Studienpopulation genügend groß, sollte es nach der Randomisierung im Mittel keinen nennenswerten Unterschied in der Zusammensetzung der Behandlungsgruppen geben. Prognostische Faktoren, wie Alter, Geschlecht, Schwere der Krankheit und Komorbiditäten, sollten in beiden Gruppen daher gleich häufig vertreten sein. Neben den bekannten Faktoren werden aber auch alle unbekannt Faktoren gleichmäßig auf die Gruppen verteilt. Da sich die Behandlungsgruppen somit im Mittel in nichts unterscheiden als in der Behandlungsform, kann das erzielte Studienergebnis direkt auf die The-

rapie zurückgeführt werden. Somit liefern RCTs validere Ergebnisse als nicht-randomisierte Studien oder unkontrollierte, d.h. einarmige Untersuchungen.

Geeignete Methoden für die Durchführung einer Randomisierung ist die Telefonrandomisierung, d.h. eine unabhängige Stelle teilt per Telefon mit, welcher Gruppe der nächste Patient zugeteilt wird, oder die Verwendung einer Zufallszahlen-Tabelle oder eines Computer-Zufallszahlengenerators.

Bei einer geringen Zahl von Studienteilnehmern kann es vorkommen, dass die Randomisierung nicht für die gewünschte Gleichverteilung sorgt. Daher sollte man sich immer auch die Charakteristika der Studienteilnehmer ansehen und sich ein Bild über die Gruppenzusammensetzung machen (meist in Tabelle 1 einer Studienpublikation dargestellt).

– Geheimhaltung der Behandlungsfolge (allocation concealment)

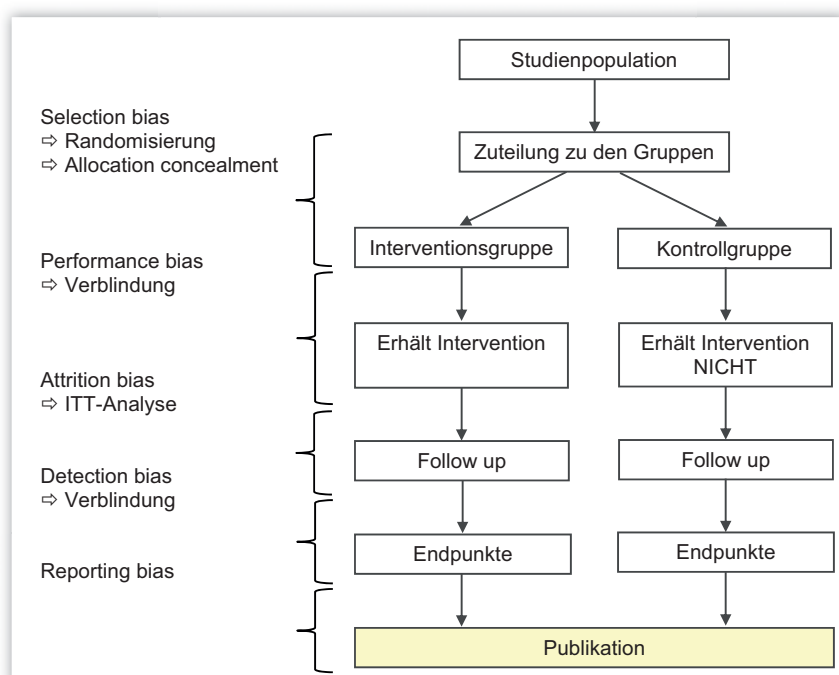
Ein Störfaktor kann bereits vor Studienbeginn auftreten, wenn bei der randomisierten Zuteilung der Studienteilnehmer zur Behandlungsgruppe Einfluss genommen wurde, wodurch die Gleichverteilung der Störgrößen auf die Gruppen verhindert wurde. Weiß beispielsweise das Studienpersonal, das die Patienten

zu den Behandlungsgruppen zuteilt, welche der Behandlungen als nächstes bei der Zuteilung an der Reihe ist, so kann es möglicherweise Einfluss darauf nehmen, welcher Patient als nächster in die Studie eingeschlossen wird. Dies kann im Extremfall dazu führen, dass am Ende z.B. die kränkeren Patienten in der Interventionsgruppe sind und die gesünderen in der Placebo-Gruppe. Damit werden die kränkeren Patienten der neuen Therapie zugeführt und damit ein Bias geschaffen, der sich gegen die neue Behandlung richtet. Das Studienergebnis wäre dann kaum noch auf die Behandlung zurückzuführen, da in diesem Fall die unterschiedliche Schwere der Krankheit höchstwahrscheinlich maßgeblicher am beobachteten Ergebnis beteiligt ist als die Behandlung selbst.

Um diesem Fehler vorzubeugen, ist es wichtig, das Studienpersonal hinsichtlich der Behandlungsfolge zu verblinden. Eine Möglichkeit hierfür wäre die Verwendung von fortlaufend nummerierten, versiegelten und undurchsichtigen Briefumschlägen.

Randomisierung in Kombination mit einer Geheimhaltung der Behandlungsfolge ermöglicht eine nicht vorhersagbare Zuteilungs-Sequenz und ist ausreichend, um einen *selection bias* zu verhindern.

Die Gleichverteilung aller prognostischen Faktoren auf die Gruppen sollte



**Abbildung 1** Formen systematischer Fehler (Bias) und Maßnahmen dagegen

während der gesamten Dauer der Studie, also bis zur Erhebung der Endpunkte, aufrechterhalten worden sein. Jedoch können während der Durchführung einer klinischen Studie Ereignisse auftreten, die diese Gleichheit gefährden, sodass in einem solchen Fall das Studienergebnis nicht mehr nur auf die unterschiedliche Behandlung zurückgeführt werden kann.

**Performance bias**

Teilnehmer einer Gruppe müssen, abgesehen von der eigentlichen Behandlung, gleich versorgt werden wie die der anderen Gruppe. Um bewusste und unbewusste Einflussnahme auf die Behandlung und damit einen sogenannten *performance bias* zu verhindern, soll-

ten auch alle Studien-Ärzte und -Schwestern verblindet sein. Eine Möglichkeit, die Studie zu verblinden, ist die Verwendung eines identisch aussehenden Placebos. Im zahnmedizinischen Bereich sind die Behandlungen jedoch häufig nicht medikamentös. Hier wären methodisch gesehen Schein-Behandlungen eine geeignete Möglichkeit – vorausgesetzt, sie sind ethisch vertretbar.

Dennoch wird es häufig der Fall sein, dass nicht alle an einer Studie beteiligten Personen verblindet werden können. Dies ist beispielsweise der Fall, wenn unterschiedliche Behandlungsverfahren bzw. -techniken angewendet werden. Dann muss der behandelnde (Zahn-)Arzt wissen, was er tun soll. Eine fehlende Verblindung muss jedoch nicht zwingend zu einem verzerrten Er-

gebnis führen, solange der Patient und die Datenerheber und -auswerter die Art der Behandlung nicht kennen.

Auch die Art des Endpunktes spielt bei der Einschätzung, ob ein berichtetes Studienergebnis vertrauenswürdig ist, eine Rolle. Bei subjektiven Endpunkten wie Schmerz oder Lebensqualität kann die Erwartungshaltung der Studienteilnehmer hinsichtlich Wirksamkeit oder Nicht-Wirksamkeit der Behandlung in das Ergebnis mit einfließen und dieses verzerren; bei objektiven Endpunkten, wie Überleben oder Rezidiv, ist das Risiko für Bias geringer. Somit ist es gerade bei subjektiven Endpunkten wichtig, dass neben dem Studienpersonal auch die Studienteilnehmer verblindet sind. Dies bedeutet, dass das Risiko für Bias für jeden einzelnen Endpunkt separat bewertet werden sollte.

**Detection bias**

Um weitere Risiken für Bias zu minimieren, sollten neben den Studienärzten und dem beteiligten klinischen Personal auch alle weiteren Personen, die in die Studie involviert sind, verblindet sein. Dazu gehören die Personen, die die Endpunkte erheben. Nur das „Nicht-Wissen“ der jeweiligen Behandlungsform garantiert, dass Unterschiede in der Endpunkterhebung (*detection bias*) verhindert werden können. Dies ist besonders wichtig bei der Erhebung subjektiver Endpunkte.

Oft wird in Studienberichten der Begriff „doppelt verblindet“ verwendet. Dieser Ausdruck lässt den Leser allerdings im Unklaren, welche an der Studie beteiligten Personen nun tatsächlich verblindet waren und welche nicht. Eine vollständige und transparente Beschreibung der Studienmethodik ist daher Grundvoraussetzung für eine vollständige Risikobewertung (vgl. CONSORT Statement, <[www.consort-statement.org](http://www.consort-statement.org)> [3]).

**Attrition Bias**

Ein weiteres Risiko für Bias besteht darin, dass meist nicht alle in eine Studie eingeschlossenen Teilnehmer diese auch gemäß Protokoll abschließen, sondern sie aus verschiedenen Gründen vorzeitig verlassen haben. Von diesen Teilnehmern konnte der Endpunkt

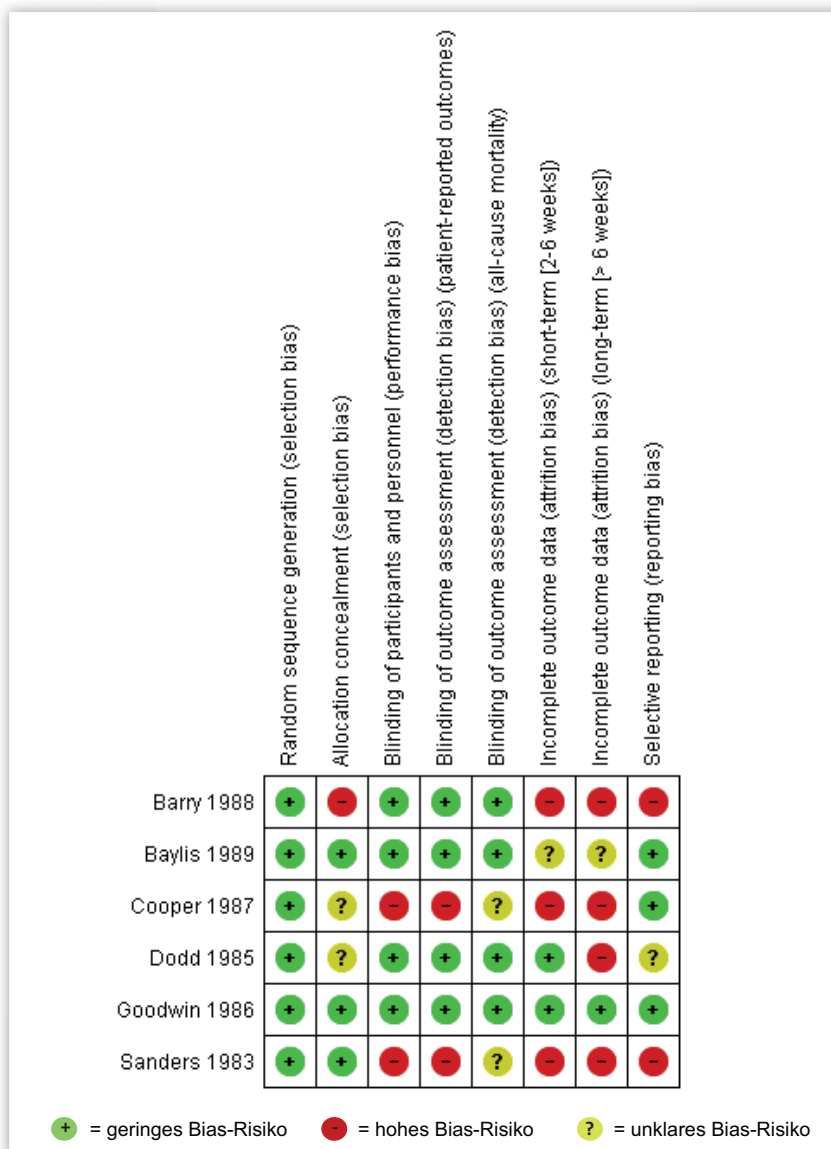


Abbildung 2 Beispiel einer „Risk of bias summary“-Graphik [6]

nicht gemessen werden (der Nachbeobachtung verlorengegangen: *lost to follow-up*). Teilnehmer, die nicht bis zum Ende einer Studie nachverfolgt werden, können andere prognostische Faktoren aufweisen als diejenigen, die bis zum Ende in der Studie verbleiben. Dadurch ist die Gleichverteilung dieser Faktoren zwischen den Gruppen gefährdet (*attrition bias*). Ideal wäre es, wenn von jedem Teilnehmer der Endpunkt bekannt wäre oder zumindest der Grund für ein Ausscheiden aus der Studie.

Gründe für frühzeitiges Verlassen einer laufenden Studie können vielfältiger Natur sein. Steht die Ursache dafür nicht in Bezug zum Endpunkt, wie Umzug in eine andere Stadt, so erhöht dies nicht das Bias-Risiko. Liegen die Ursachen hingegen in einer der Behandlungen oder den Auswirkungen einer der Behandlungen, dann besteht die Gefahr, dass dies zu Unterschieden zwischen den Behandlungsgruppen führt, die die Endanalyse verzerren. Das gleiche gilt für Fälle, in denen sich Studienteilnehmer nicht an die Behandlungsanweisung halten oder die Gruppen wechseln, d.h. wenn in Behandlungsgruppe 1 randomisierte Teilnehmer die Behandlung 2 erhalten oder umgekehrt.

Würden nur diejenigen Personen in die Analyse eingeschlossen, die protokollgemäß die Studie durchlaufen und beendet haben, so würden all diejenigen nicht berücksichtigt werden, die aufgrund von Nicht-Wirksamkeit oder Nebenwirkungen die Studie frühzeitig beendet haben. Dies würde dazu führen, dass das Ergebnis in dieser Gruppe überschätzt würde, da mehr Teilnehmer mit einer positiven Wirkung in dieser Gruppe verbleiben und ausgewertet werden. Mit einer *Intention-to-treat*-Analyse kann dem *attrition bias* begegnet werden. Dabei werden alle Patienten strikt in derjenigen

Gruppe analysiert, zu der sie ursprünglich zugeteilt wurden, und auch diejenigen in der Endanalyse berücksichtigt, die die Studie frühzeitig beendet haben.

### Reporting Bias


Positive und statistisch signifikante Studienergebnisse werden häufiger publiziert als negativ nicht-signifikante [2]. Dies führt dazu, dass wir bei der Recherche nach Studienberichten zu einer definierten Fragestellung vermehrt positive Studienergebnisse finden werden und dadurch den wahren Effekt der Behandlung überschätzen. Dieser sogenannte *reporting bias* ist ein wesentliches Verzerrungs-Risiko. Überprüft werden kann das Vorhandensein dieses Bias z.B. durch Abgleich der Endpunkte, die in der Studienmethodik genannt und derjenigen, die im Ergebnisteil berichtet sind. Auch ein Vergleich von Publikation mit Studienprotokoll und prospektivem Studienregistereintrag kann hilfreich sein, um dieses Risiko einzuschätzen.

### Risk of Bias – Bewertung

Anhand der genannten Bias-Quellen kann die Aussagekraft von Studienergebnissen beurteilt werden. Die internationale gemeinnützige Organisation Cochrane hat dafür ein hilfreiches Instrument entwickelt, das sog. *Risk of Bias Tool* [1]. Mit diesem Instrument kann das Risiko für Bias für jeden einzelnen der oben erwähnten Aspekte bewertet und damit die Validität der Studienergebnisse bestimmt werden. Dabei ist es erforderlich, subjektiv einzuschätzen, ob die jeweilige potenzielle Störgröße das Ergebnis der Studie beeinflussen kann oder nicht. Mit

dem *Risk-of-Bias*-Instrument werden 8 verschiedene Fehlerquellen, die in klinischen Studien auftreten können, in ein geringes, hohes oder unklares Risiko klassifiziert und jeweils in einer Tabelle graphisch dargestellt (Abb. 2). Dadurch lässt sich auf einen Blick feststellen, welche Evidenz zu einer Fragestellung vorliegt und wie die Validität der Ergebnisse einzuschätzen ist.

### Zusammenfassung

Um sich ein Bild über den aktuellen Kenntnisstand zu einer (zahn)medizinischen Fragestellung zu machen, sind umfangreiche Recherchen in mehreren Literaturdatenbanken und anderen Quellen erforderlich. Liegt zur Fragestellung eine aktuelle und methodisch gute systematische Übersichtsarbeit vor, stellt das darin zusammengefasste Gesamtergebnis über mehrere Einzelstudien den höchsten Evidenzlevel und damit das vertrauenswürdigste Ergebnis dar. Gibt es dagegen noch keine systematische Übersichtsarbeit, muss auf die Originalstudienartikel zurückgegriffen werden. Jeder (Zahn-)Mediziner sollte in der Lage sein, relevante Studienartikel kritisch zu bewerten, um abschätzen zu können, wie vertrauenswürdig die berichteten Ergebnisse sind. Dafür kann das Cochrane *Risk-of-Bias*-Instrument hilfreich sein. 

Dr. Anette Blümle,  
Prof. Dr. Gerd Antes,  
Cochrane Deutschland,  
Universitätsklinikum Freiburg,  
Medizinische Fakultät,  
Albert-Ludwigs-Universität Freiburg,  
Deutschland;  
Prof. Dr. Jens Christoph Türp,  
UZB-Universitätszahnkliniken,  
Basel, Schweiz

### Literatur

- Higgins J, Altman D, Sterne J: Chapter 8: Assessing risk of bias in included studies. Higgins J, Green S, editors. In: Higgins JPT, Green S (Hrsg.): Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. <www.handbook.cochrane.org>
- Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K: Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009; (1): MR000006
- Schulz KF, Altman DG, Moher D: CONSORT 2010: Aktualisierte Leitlinie für Berichte randomisierter Studien im Parallelgruppen-Design. Dtsch Med Wochenschr 2011; 136: e20–e23
- Shea BJ, Grimshaw JM, Wells GA et al.: Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007; 7: 10
- Whiting P, Savovic J, Higgins JP et al.: ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016; 69: 225–234
- www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ (letzter Zugriff am 19.03.2017)