



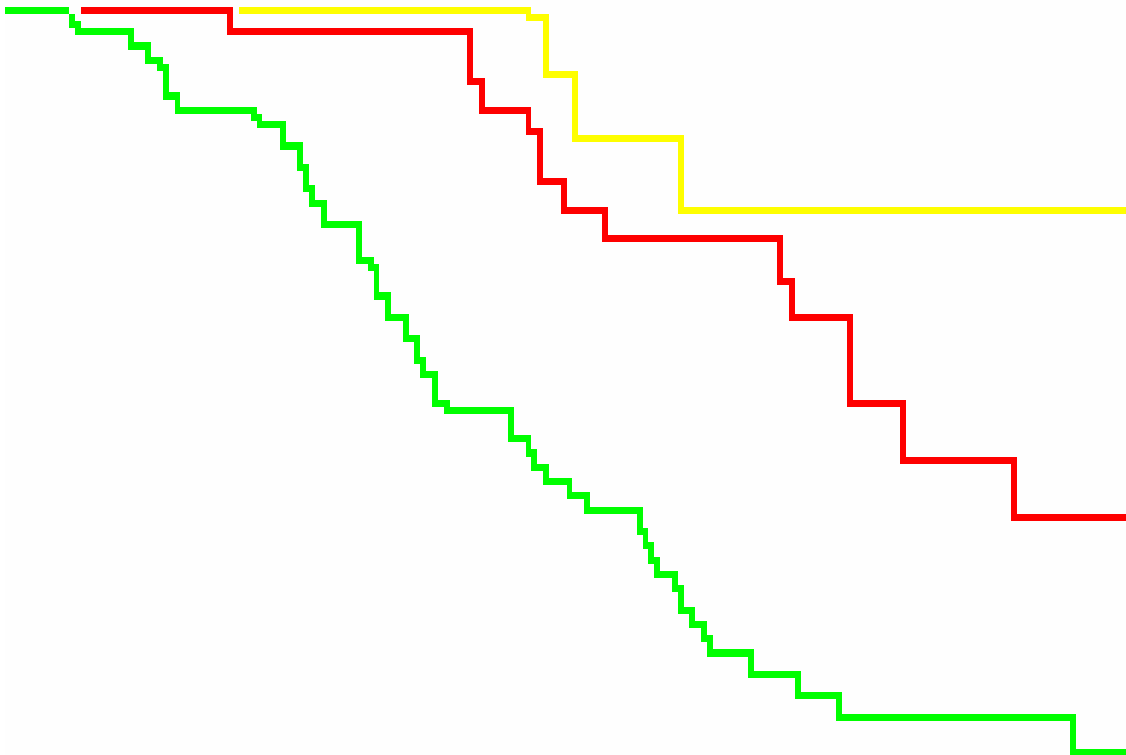
Universitätsklinikum  
Hamburg-Eppendorf

Institut für Allgemeinmedizin

Hans-Hermann Dubben

Hans-Peter Beck-Bornholdt

# Unausgewogene Berichterstattung in der medizinischen Wissenschaft *- publication bias -*



ISBN 3-00-014238-X

Den Patienten gewidmet,  
die an klinischen Studien teilnahmen,  
die niemals veröffentlicht wurden.

*„Positive, aber auch negative Ergebnisse  
müssen veröffentlicht oder der Öffentlichkeit  
anderweitig zugänglich gemacht werden.“*  
Deklaration von Helsinki, Weltärztebund

Gefördert durch die Unna Stiftung, Düsseldorf

2. Auflage

ISBN 3-00-014238-X

Erstellt am Institut für Allgemeinmedizin des Universitätsklinikums Hamburg-Eppendorf

Druck: Eigendruck

Verlag: Institut für Allgemeinmedizin des Universitätsklinikums Hamburg-Eppendorf

Copyright: Hans-Hermann Dubben und Hans-Peter Beck-Bornholdt, 2004.

Vertrieb:

Per Email: [dubben@uke.uni-hamburg.de](mailto:dubben@uke.uni-hamburg.de)

Postalisch: Institut für Allgemeinmedizin, UKE, Martinistrasse 52, 20246 Hamburg

Schutzgebühr: € 5,-

HSH Nordbank, Kontonr. 104 364 000, BLZ 210 500 00, Verwendungszweck: I 549 - 2

## Zusammenfassung

*Publication bias* bezeichnet unausgewogene Berichterstattung in der Wissenschaft, die zu einer Fehleinschätzung der wissenschaftlichen Realität und im Allgemeinen zu einer Überschätzung von Therapieerfolgen führt.

Die Ursache für *publication bias* liegt in der Mentalität der Forscher und an den Erwartungen der Gesellschaft an die Wissenschaft.

*Publication bias* behindert den wissenschaftlichen Fortschritt und verursacht die Fehlbehandlung von Patienten.

Die gegenwärtige Publikationspraxis ist unethisch und mit der Deklaration von Helsinki unvereinbar.

Die einzig wirksame Maßnahme gegen *publication bias* ist die lückenlose Registrierung aller Studien und die Verpflichtung zur Offenlegung aller Ergebnisse.

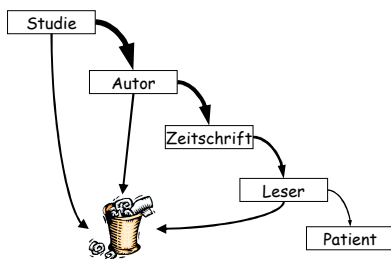
## **Inhaltsverzeichnis**

1. Was ist *publication bias*?
2. **Wie erkennt man unausgewogene Berichterstattung in der Wissenschaft?**
3. **Stand der Forschung zur Frage unausgewogener Berichterstattung (*publication bias*) in der Wissenschaft**
  - 3.1. Erfassung der Studien zur Bestimmung von *publication bias*
  - 3.2. Der Beitrag der Autoren zum *publication bias*
  - 3.3. Der Beitrag der Fachzeitschriften zum *publication bias*
  - 3.4. Der Beitrag der Leser zum *publication bias*
4. **Beispiele für unausgewogene Berichterstattung in der Wissenschaft**
  - 4.1 *Publication bias* bei Blutdruck senkenden Medikamenten
  - 4.2 Modeerscheinung in der Wissenschaft: *split-course* Strahlentherapie
  - 4.3 *Publication bias* bei der Prüfung diagnostischer Tests.
5. **Die Folgen unausgewogener Berichterstattung in der medizinischen Wissenschaft**
6. ***Publication bias* und Evidenzbasierte Medizin**
7. **Gegenmaßnahmen**
8. **Quergedacht**
9. **Nachwort**
10. **Literatur**

# 1. Was ist *publication bias*?

*The human intellect  
... is more moved and excited by affirmatives  
than by negatives.  
Francis Bacon, 1621.*

Vom Beginn einer klinischen Studie über die Lektüre der Publikation bis zur Anwendung ihrer Ergebnisse, sind zahlreiche Hürden zu nehmen. Diese Hürden sind häufig eher psychologischer und publizistischer als wissenschaftlicher Natur. Dies hat zur Folge, dass nicht alle Resultate veröffentlicht werden. Das, was die Endverbraucher wissenschaftlicher Information, also Wissenschaftler, Ärztinnen und Patienten im Rahmen einer Literaturrecherche finden, ist häufig kein getreues Abbild dessen, was tatsächlich in Kliniken und Labors zu dem Thema erforscht worden ist. Man spricht von *publication bias*, wenn die veröffentlichten Forschungsergebnisse nicht repräsentativ für alle erzielten Resultate sind. *Publication bias* ist unausgewogene Berichterstattung in der Wissenschaft.



*Abb. 1: Längst nicht alle wissenschaftlichen Ergebnisse kommen dem Patienten zugute. Auf dem beschwerlichen Weg von der Studie bis zum Patienten bleibt ein großer Teil der Ergebnisse auf der Strecke und landet im Papierkorb.*

Das Ausbleiben einer Publikation kann die unterschiedlichsten Gründe haben. Zunächst sind die finanziellen und logistischen Hürden bis zum Abschluss der Studie zu nehmen. Danach stehen die Studienleiter vor der Entscheidung, ob sie das Ergebnis überhaupt publizieren wollen und, wenn ja, in welcher Zeitschrift und in welcher Sprache. Offenbar gibt es keine Standards an die sich ein Wissenschaftler halten kann und muss, um zu entscheiden, welche Resultate es wert sind, veröffentlicht zu werden und welche nicht (Dickersin 1990). Eine weitere Frage ist, ob und wo das Manuskript zur Veröffentlichung angenommen wird. Manche Studien werden in weit verbreiteten und in Literaturdatenbanken geführten englischsprachigen Zeitschriften publiziert. Andere erscheinen in kleinen nationalen Zeitschriften in der Muttersprache der Autoren. Und wieder andere werden nie gedruckt.

Der wahrscheinlich älteste Hinweis auf dieses Phänomen geht auf den Philosophen und Atheisten Diagoras von der Insel Milos zurück. Diagoras wurde vor 2500 Jahren angesichts der Danktafeln (Abb.2a), die



*Abb. 2a: Publikation positiver Ergebnisse.*

gerettete Schiffbrüchige in einem Tempel aufgestellt hatten, von einem Priester

gefragt, ob dies nicht ausreichend Evidenz für die Existenz der Götter sei. Worauf Diagoras entgegnet haben soll: „Und wo sind die Tafeln der Ertrunkenen?“ (gefunden bei Petticrew 1998). Abb.2b zeigt die auf der Hand liegende Antwort. Dickersin (1990) zufolge war Sterling (1959) der erste, der auf die bevorzugte Veröffentlichung positiver Ergebnisse in der Wissenschaft hingewiesen hat. Der Begriff *publication bias* wurde von Smith (1980) geprägt (Dickersin 1990).



Abb. 2b: Das weniger spektakuläre Schicksal negativer Ergebnisse.

*Publication bias* setzt sich aus den verschiedensten Komponenten zusammen und hat zahlreiche Ursachen und Verursacher. Sie werden in dieser Broschüre benannt und ergründet. Vorher müssen wir uns aber ein paar Gedanken zur Methodik machen.

## 2. Wie erkennt man unausgewogene Berichterstattung in der Wissenschaft?

Wie kann man erkennen, ob die Studien, die beispielsweise in eine Metaanalyse<sup>1</sup> eingegangen sind, repräsentativ für den Stand der Forschung zu einer bestimmten wissenschaftlichen Fragestellung sind? Wie kann man erkennen, ob sie nicht möglicherweise nur eine unausgewogene, viel zu positive Auswahl darstellen? Zu diesem Zweck wurden verschiedene Methoden erdacht. Die Trichtergrafik, die im folgenden Abschnitt erklärt wird, ist eines von mehreren Werkzeugen, mit denen man eine Sammlung von Ergebnissen auf unausgewogene Berichterstattung (*publication bias*) untersuchen kann.

### Was ist eine Trichtergrafik (*funnel plot*)?

Das Prinzip dieses Verfahrens lässt sich mit einem Würfelspiel anschaulich darstellen. Mit einem ganz normalen Würfel können sechs verschiedene Ergebnisse erzielt werden:

1      2      3      4      5      6

Bei einem fairen Würfel sind alle sechs Ergebnisse gleich wahrscheinlich. Die durchschnittliche Punktzahl pro Wurf beträgt folglich  $(1+2+3+4+5+6)/6 = 21/6 = 3,5$ . Wenn man mit sehr vielen Würfeln würfelt, wird die Abweichung des Durch-

<sup>1</sup> Metaanalyse: Synthese der Ergebnisse verschiedener vergleichbaren Studien zum gleichen oder einem ähnlichen Thema. Das Motiv für die Durchführung von Metaanalysen ist, durch Zusammenlegen der Daten hohe Patientenzahlen und damit eine bessere Aussagekraft zu erlangen. Die Ergebnisse hängen allerdings entscheidend von der Auswahl und der Qualität der eingeschlossenen Studien ab.

schnittswertes von diesem theoretischen Wert sehr gering sein. Je weniger Würfel man wirft, um größer werden die zufällig auftretenden Abweichungen sein. Die folgende Grafik zeigt das Ergebnis sehr vieler Spielsimulationen am Computer. Jeder Punkt ist das Resultat eines einzelnen Würfel experiments.

Die Streuung um den theoretischen Wert 3,5 (gestrichelte senkrechte Linie) wird mit zunehmender Anzahl der Würfel geringer. Als Maß für die Messgenauigkeit wurde hier die Anzahl der Würfel verwendet. In klinischen Studien wird dafür oft die Anzahl der Patienten genommen und zur Messgröße der Studie (z.B. der Heilungsrate) in Beziehung gesetzt. Wenn man viele Studien zu derselben Therapie in so einer Grafik darstellt, dann erwartet man ebenfalls eine Trichterform, denn die Ergebnisse von Studien mit wenigen Patienten unterliegen eher dem Zufall als die Resultate großer Studien.

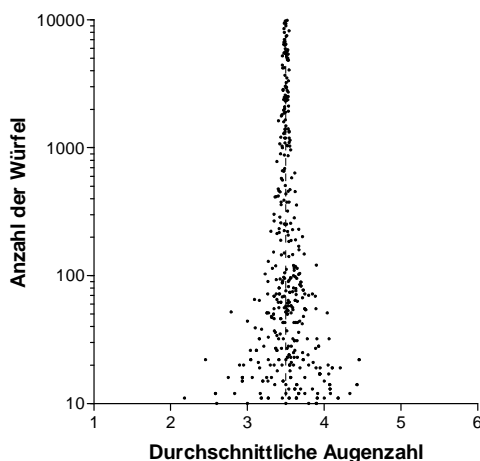


Abb. 3: Durchschnittliche Augenzahl, die beim Wurf verschieden vieler Würfel erzielt wird. Die Schwankungsbreite nimmt mit der Anzahl der Würfel ab. Es ergibt sich die Form eines umgekehrten Trichters (funnel). Die Abbildung zeigt das Ergebnis zahlreicher Computersimulationen.

Hätten wir beim Würfelspiel Ergebnisse mit kleiner Augenzahl verschwiegen, wäre der Trichter unsymmetrisch geworden. Genauso verhält es sich bei unausgewogener Berichterstattung (*publication bias*), wenn Studien mit negativen Ergebnissen mit geringerer Wahrscheinlichkeit veröffentlicht werden. Bei Asymmetrie in der Trichtergrafik besteht somit dringender Verdacht auf unausgewogene Berichterstattung.

### 3. Stand der Forschung zur Frage unausgewogener Berichterstattung (*publication bias*) in der Wissenschaft

Bei Diagoras' Schiffbrüchigen ist die positiv eingefärbte Berichterstattung offensichtlich. Kein einziger Ertrunkener hat nach menschlichem Ermessen die Gelegenheit gehabt, sein Missfallen mit einer Tafel im Tempel kundzutun. Im Tempel werden zu 100 Prozent positive Ergebnisse gezeigt. In einer der führenden medizinischen Fachzeitschriften, dem *New England Journal of Medicine*, beträgt der Anteil positiver Ergebnisse 88 Prozent. In psychologischen Fachzeitschriften beträgt der Anteil sogar über 95 Prozent (Dickersin & Min, 1993, Sterling 1959) und erinnert damit an „Wahlergebnisse“ in totalitären Staaten. Derart hohe Anteile an positiven Ergebnissen lassen deutlich Zweifel aufkommen, ob diese Berichte repräsentativ für das sind, was tatsächlich alles erforscht wurde. Die Diagoras-Frage lautet also: „Und wo sind die Veröffentlichungen der Studien mit negativem Ergebnis?“

Im Werdegang einer Studie und deren (etwaiger) Publikation gibt es vier Stationen, die für die Untersuchung und den Nachweis von *publication bias* besonders interessant sind:

1. Die Erfassung der Studien, auch derer, die später nicht publiziert werden. Unterscheiden sich die publizierten Studien systematisch von den nicht publizierten?
2. Die Autoren der Studien, denn sie entscheiden, ob sie ihre Ergebnisse überhaupt publizieren wollen. Wovon wird deren Entscheidung beeinflusst?
3. Die wissenschaftlichen Fachzeitschriften bzw. deren Herausgeber und Gutachter (*peer reviewer*). Wovon wird deren Entscheidung beeinflusst?
4. Der Verbraucher, also die Ärztin, der Patient, die Wissenschaftlerin, denn sie muss die Studie in der umfangreichen internationalen Fachliteratur finden und möglichst unvoreingenommen lesen, bewerten und verwerten. Wovon hängt es ab, dass eine Publikation auffindbar ist? Welchen Einfluss haben die Erwartungen und Vorlieben der Verbraucherin?

Diese Stationen werden im Folgenden abgearbeitet.

### **3.1. Erfassung der Studien zur Bestimmung von *publication bias***

Für die Untersuchungen des *publication bias* ist zunächst eine Informationsquelle notwendig, die alle klinischen Studien unabhängig von ihrem Ergebnis erfasst. Dies ist dann gewährleistet, wenn ein Forschungsvorhaben bei Initiierung oder zumindest zu einem Zeitpunkt, an dem noch

keine Ergebnisse vorliegen, registriert wird. Zu diesen Quellen zählen:

- Studienprotokolle, die von Ethikkommissionen genehmigt wurden (Easterbrook 1991; Dickersin *et al.* 1992; Stern & Simes 1997),
- Studien, die von der amerikanischen Gesundheitsbehörde NIH (Dickersin & Min 1993) oder von anderen Fördereinrichtungen finanziert wurden (Misakian & Bero 1998)
- und Studien, die von Fachgesellschaften durchgeführt wurden (Ioannidis 1998).

Weitere Hinweise zu *publication bias* können auch aus anderen Quellen gewonnen werden, wie beispielsweise aus Dissertationen (Zimpel & Windeler, 2000) und Zusammenfassungen (*Abstracts*) für wissenschaftliche Fachtagungen (Callaham *et al.* 1998).

#### **Ethikkommissionen**

Easterbrook *et al.* (1991) untersuchten die Forschungsprojekte, die vom *Central Oxford Research Ethics Committee* zwischen 1984 und 1987 genehmigt worden waren. Bei Beobachtungsstudien und Laboruntersuchungen fanden sie eine bevorzugte Veröffentlichung positiver Ergebnisse. Statistisch signifikante Ergebnisse hatten eine mehr als dreifach höhere Chance, publiziert zu werden als nicht-signifikante. Für randomisierte kontrollierte Studien<sup>2</sup> fanden Easterbrook *et al.* aller-

---

<sup>2</sup> Randomisierte kontrollierte Studien werden gegenwärtig als das bestmögliche Studiendesign angesehen, um die Wirkung einer medizinischen Intervention zu untersuchen. Bei diesem Studiendesign wird eine Gruppe von Patienten oder Probanden nach dem Zufallsprinzip in zwei oder mehrere Gruppen aufgeteilt (Randomisierung). Die eine Gruppe erhält das zu untersuchende Medikament bzw. die zu untersuchende Behandlung, während die



dings keinen Hinweis auf *publication bias* (siehe auch den späteren Abschnitt „Quer-gedacht“).

Dickersin *et al.* (1992) untersuchten 737 Studien, die an den *Johns Hopkins Health Institutions* in Baltimore bis 1980 durchgeführt worden waren. Sie fanden, dass dort statistisch signifikante Ergebnisse zwei- bis dreimal so häufig publiziert wurden wie nicht signifikante Resultate.

Stern & Simes (1997) untersuchten das Schicksal von 748 Studien, die zwischen 1979 und 1988 vom *Royal Prince Alfred Hospital Ethics Committee* in Sydney genehmigt worden waren. Es zeigte sich, dass Studien mit positivem Ergebnis etwa doppelt so häufig publiziert wurden wie Studien mit negativem Ergebnis. Außerdem wurden negative Studien später veröffentlicht als positive (*time-lag bias*), denn die mediane Dauer bis zur Publikation betrug etwa 5 Jahre für positive und 8 Jahre für negative Ergebnisse.

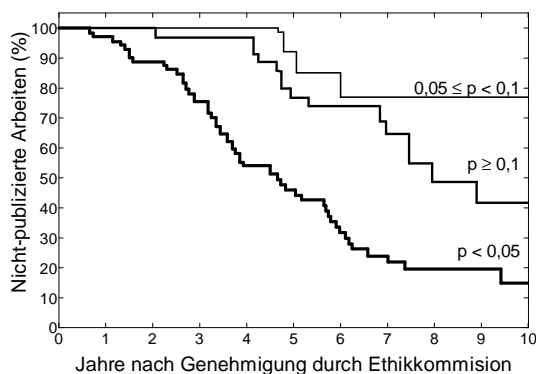


Abb. 4: Anteil der nicht-publizierten Arbeiten im Verlauf der Zeit nach Genehmigung durch eine Ethikkommission. Studien mit statistisch signifikanten Ergebnissen („ $p < 0,05$ “) werden häufiger

andere Gruppe eine konventionelle Therapie, ein Placebo oder nichts erhält.

und früher veröffentlicht als Studien mit nicht-signifikanten Ergebnissen („ $p > 0,05$ “; Stern & Simes, 1997)

Egger & Smith (1998) und Dickersin (1997) führten eine Metaanalyse aller verfügbaren Untersuchungen durch, die auf Anträgen bei Ethik-Kommissionen basieren. Der Anteil nicht veröffentlichter Ergebnisse variierte zwischen 7 und 41 Prozent.

Alle fünf hier zitierten Untersuchungen zeigen eine Bevorzugung von Studien mit positiven Ergebnissen. Die Chance, publiziert zu werden, ist für positive Ergebnisse etwa dreifach größer als für negative.

### Studienregister

Es gibt hunderte fachspezifische Studienregister. Deren Inhalte sind jedoch nicht standardisiert und teilweise überlappen sie. Das Fehlen eines zentralen Studienregisters behindert die Arbeit derjenigen, die Informationen über aktuelle Forschungsvorhaben suchen. Auch Datenbanken der pharmazeutischen Industrie über Medikamente, die sich in der Entwicklung befinden („*pipeline sources*“), könnten als Informationsquellen über gegenwärtig laufende Medikamentenstudien dienen. Manheimer *et al.* (2002) hat drei dieser Quellen verglichen, von denen zwei für sich beanspruchten, ein vollständiges Bild der aktuellen US-Medikamentenforschung zu liefern. Nur drei von 32 Vorhaben waren tatsächlich in allen drei Datenbanken registriert. Die Hälfte war nur in jeweils einem Register zu finden. Von zwölf in der Entwicklung befindlichen Medikamentenstudien zur Behandlung von Prostatakrebs waren drei in anderen Studienregistern nicht auffindbar. Bei Dickdarmkrebs waren es sogar acht von 20 Studien.

Offenbar gewähren derzeit existierende Studienregister keinen vollständigen Überblick über das Forschungsgeschehen.

### **Fachgesellschaften**

Von Fachgesellschaften durchgeführte prospektive Studien<sup>3</sup> sind ebenfalls bereits aktenkundig, bevor die ersten Ergebnisse vorliegen. Inwieweit man nachträglich an diese Studienprotokolle heran kommt, auch an diejenigen unveröffentlicher Untersuchungen, hängt von der Kooperationsbereitschaft der Fachgesellschaft ab. Ioannidis (1998) untersuchte Wirksamkeitsstudien, die im Zeitraum von 1986 bis 1996 von den Fachgesellschaften ACTG (*AIDS Clinical Trials Group*) und CPCRA (*Terry Bein Community Programs for Clinical Research on AIDS*) durchgeführt wurden. Die Ergebnisse seiner Untersuchung werden in Kapitel 3.2 unter *time lag bias* dargestellt.

### **Wissenschaftliche Kongresse**

Wissenschaftliche Kongresse sind häufig das erste Forum, auf dem neue Forschungsergebnisse vorgestellt werden. Daher bieten sie eine interessante Gelegenheit zur Untersuchung von *publication bias*.

Callahan *et al.* (1998) untersuchten die Beiträge für einen Kongress in Notfallmedizin. Hierbei betrachteten sie drei Ebenen: 1. das Einreichen von *Abstracts* zu dem Kongress, 2. die Auswahl dieser *Abstracts* für einen Vortrag, 3. und die spätere Auswahl für eine Publikation. *Abstracts* mit einem positiven Ergebnis hatten doppelt so gute Chancen zum Kongress angenommen

zu werden als mit negativem Ergebnis. Es zeigte sich, dass bereits vor dem Einreichen eine Selektion zugunsten positiver Studien stattgefunden hatte, denn ob aus einem eingereichten *Abstract* später eine Publikation wurde, hing maßgeblich davon ab, ob es zuvor als *Abstract* für einen Vortrag auf dem Kongress akzeptiert wurde und ob ein positives Ergebnis vorlag.

Scherer *et al.* (1994) untersuchten, ob und wann randomisierte kontrollierte Studien, deren Zusammenfassungen auf einem Ophthalmologenkongress im Jahre 1988 und 1989 vorgestellt worden waren, später veröffentlicht wurden. Außerdem führten sie eine Metaanalyse unter Berücksichtigung von zehn weiteren Studien dieses Typs durch. Nur etwa die Hälfte der Studien wurde nachfolgend veröffentlicht, die meisten binnen zwei Jahren. Dabei zeigte sich eine leichte Bevorzugung signifikanter Ergebnisse.

Klassen *et al.* (2002) untersuchten die *Proceedings der Society for Pediatric Research* der Jahre 1992 bis 1995. 41 Prozent der 447 randomisierten kontrollierten Studien, die auf diesen Tagungen vorgestellt worden waren, wurden später nicht veröffentlicht. Von den veröffentlichten Studien waren 64 Prozent zugunsten der experimentellen Therapie ausgefallen, während es bei den nicht veröffentlichten Studien nur 44 Prozent waren ( $p < 0,001$ ). Studien, die nicht zugunsten der experimentellen Therapie ausfielen, wurden folglich seltener publiziert. Darüber hinaus stellten die Autoren anhand einer Trichtergrafik fest, dass bereits bei den auf der Tagung vorgestellten Studien *publication bias* vorliegt.

---

<sup>3</sup> Prospektive Studie: Klinische Studie, bei der vor Studienbeginn festgelegt wird, was untersucht werden soll.

Timmer *et al.* (2002) untersuchten eine Zufallsstichprobe aus den *Abstracts*, die zwischen 1992 und 1995 zur *Digestive Diseases Week*, einem jährlichen Kongress von Gastroenterologen, eingereicht wurden. Knapp die Hälfte davon (47%) wurde später publiziert. *Abstracts*, die für einen Vortrag bei dem Kongress angenommen worden waren, wurden nachfolgend mit größerer Wahrscheinlichkeit veröffentlicht als *Abstracts*, die abgelehnt worden waren (54% vs. 34%). Es wurde kein signifikanter Zusammenhang zwischen der Richtung des Studienergebnisses und der Publikationshäufigkeit gefunden, wohl aber hinsichtlich der Publikationsqualität: Studien mit nicht-signifikanten Ergebnissen erschienen signifikant seltener in Fachzeitschriften mit hohem Verbreitungsgrad und Ansehen.

Krzyzanowska *et al.* (2003) untersuchten die Tagungsbände der Jahrestagung der Amerikanischen Gesellschaft für Onkologie (*American Society of Clinical Oncology ASCO*) von 1989 bis 1998. Hier werden alle eingereichten *Abstracts* abgedruckt, auch wenn sie nicht für eine Präsentation angenommen wurden. Die Autoren berücksichtigten ausschließlich randomisierte Phase III Studien<sup>4</sup> mit mindestens 200 Patienten. 81 Prozent der Studien mit signifikantem Ergebnis und 68 Prozent mit nicht-signifikantem Ergebnis waren

innerhalb von fünf Jahren publiziert worden ( $p < 0,001$ ). Ähnlich sah es bei dem Vergleich von Studien aus, bei denen die experimentelle Behandlung besser abgeschnitten hatte. Hier waren 81 Prozent publiziert worden, im Vergleich zu 70 Prozent, wenn die experimentelle Behandlung nicht besser war ( $p = 0,001$ ). Es gab kaum Studien, die noch nach mehr als fünf Jahren veröffentlicht wurden. Auf die Dauer bis zur Publikation hatten folgende Parameter einen Einfluss: die statistische Signifikanz des Ergebnisses ( $p < 0,001$ ); die Art der Präsentation beim ASCO-Meeting (Vortrag oder „nur publiziert“ im Tagungsband, aber nicht als Vortrag angenommen;  $p = 0,002$ ); sowie der Sponsor (Pharmagesponsorte Studien wurden am schnellsten publiziert,  $p < 0,02$ ).

Hopewell & McDonald (2003) untersuchten 962 Konferenz-*Abstracts* von kontrollierten klinischen Studien, die zwischen 1980 und 2000 im *Australian and New Zealand Journal of Medicine* publiziert wurden. Davon wurde gut die Hälfte (61%) später vollständig veröffentlicht. Von der Präsentation auf der Konferenz bis zur Publikation vergingen im Mittel ein bis zwei Jahre. Die Autoren haben die *Abstracts* nicht nach positiven und negativen Ergebnissen unterschieden.

### Dissertationen

Zimpel & Windeler (2000) untersuchten die 140 Dissertationen der Jahre 1982 – 1992, aus den vier Baden-württembergischen Hochschulen Heidelberg, Freiburg i.Br., Tübingen und Ulm, die zu unkonventionellen medizinischen Verfahren durchgeführt worden waren. Ein positives Ergebnis führte eher zur Publikation, allerdings war der Unterschied nicht statistisch signifikant.

<sup>4</sup> Die Phase III Studie ist eine kontrollierte klinische Studie zum Vergleich der Wirksamkeit zweier unterschiedlicher Therapieverfahren. Die Wirksamkeit eines neuen Medikaments oder einer neuen Behandlungsmethode und das etwaige Auftreten seltener aber schwerwiegender Nebenwirkungen werden untersucht und mit dem Standardverfahren oder gegen ein Placebo verglichen. Die Verträglichkeit, die Sicherheit, sowie mögliche Arzneimittelwechselwirkungen werden erfasst. Erst nach positivem Abschluss dieser Phase kann ein neues Medikament oder eine neue Behandlungsmethode auch außerhalb von Studien eingesetzt werden.

*Zusammenfassend ist festzustellen, dass zahlreiche klinische Studien über Ethikkommissionen, Förderinstitutionen, Dissertationen, Tagungen, Fachgesellschaften, Behörden, etc. erfasst werden können. Die (publizierten) Untersuchungen über publication bias berichten, dass davon später nur etwa die Hälfte publiziert wird. Die Ergebnisse eines großen Teils der Studien scheinen in unbekanntem Schubladen zu verschwinden. Hinzu kommt der Eindruck, dass positive Ergebnisse schneller veröffentlicht werden.*

### **3.2. Der Beitrag der Autoren zum *publication bias***

In diesem Abschnitt wird der Beitrag der Autoren zum *publication bias* erläutert.

Forscher sind darauf aus, etwas Neues zu finden. Das ist auch die Erwartung der Gesellschaft an die Wissenschaft. Den Versuch eines Forscher-Kollegen zu wiederholen und damit seine Ergebnisse zu überprüfen, wird folglich von Wissenschaftlern als langweilig empfunden und ist mit dem Verdacht verbunden, man habe keine eigenen erforschenswerten Ideen und koche aus diesem Grunde lediglich die Studien seiner wesentlich intelligenteren und kreativeren Kollegen nach. So finden kritische Überprüfungen der häufig bejubelten positiven Ergebnisse kaum statt. Aber auch bei der Erforschung neuer und auf eigenen Ideen beruhender Zusammenhänge möchte man lieber herausfinden, dass

dieses oder jenes funktioniert, und nicht, dass es nicht funktioniert. Manche Autoren sind also enttäuscht und entmutigt, wenn sie negative Ergebnisse haben und zögern, diese überhaupt zur Publikation einzureichen.

#### ***Time lag bias: Zeitlich verzögerte Veröffentlichung negativer Ergebnisse***

Ioannidis (1998) untersuchte den Verbleib aller 109 randomisierten Wirksamkeitsstudien, die im Zeitraum von 1986 bis 1996 von der ACTG (*AIDS Clinical Trials Group*) und von der CPCRA (*Terry Bein Community Programs for Clinical Research on AIDS*) durchgeführt wurden. Positive Studien wurden deutlich früher publiziert als negative. (Zur Definition von positiv und negativ in diesem Fall siehe Abschnitt 8.1). Vom Studienbeginn bis zur Veröffentlichung verstrichen bei positiven Studien im Median<sup>5</sup> 4,3 Jahre, bei „negativen“ Studien waren es 6,5 Jahre ( $p < 0,001$ ). Dieser Unterschied war hauptsächlich auf das Zeitintervall zwischen Vollendung der Nachbeobachtung und der Veröffentlichung zurückzuführen. Positive Studien wurden also schneller nach Abschluss der Studie zur Veröffentlichung eingereicht (1,0 vs. 2,4 Jahre;  $p = 0,001$ ), dies ist der Autorenanteil am *time lag bias*. Die positiven Studien wurden auch schneller von den Zeitschriften publiziert (0,8 vs. 1,1 Jahre;  $p < 0,04$ ). Nur Studien mit negativen Ergebnissen wurden von mehr als einer Zeitschrift abgelehnt. Die beiden letzten Punkte sind der Herausgeber- und Gutachteranteil am *time lag bias*.

---

<sup>5</sup> Median: Zur Ermittlung des Medians sortiert man die Messwerte der Größe nach. Der Median ist dann der Wert, der in der Mitte der Reihe steht.

Ioannidis (1998) zeigte auch, dass die Studien mit negativem Ergebnis etwa doppelt so viele Patienten wie die positiven eingeschlossen hatten. Damit haben sie nicht nur eine höhere *Power*<sup>6</sup> und sind aussagekräftiger als die kleineren Studien, sondern die Vorbereitung dieser Studien durch beispielsweise das Anwerben von kooperierenden Kliniken, benötigt mehr Zeit. Damit beginnen die negativen Studien bereits mit Verspätung. Insgesamt erscheinen daher Studien mit negativem Ergebnis meist einige Jahre nach den positiven Studien zu derselben Fragestellung. Während dieser Zeit gibt es übertriebene Vorstellungen über den Nutzen eines Medikamentes oder einer Therapie. Dieser initialen Welle der Euphorie folgt dann vielleicht eine Phase der Ernüchterung. Diese wird aber oft von der nächsten Welle einer neuen Euphorie ersäuft.

### **Sprachen bias (*language bias*)**

Englisch ist die derzeitige Wissenschaftssprache. Ein auf Englisch publizierter Artikel kann und wird von mehr Anwendern gelesen als ein nicht-englischer Artikel. Es gibt Hinweise in der Literatur, dass Studien mit statistisch signifikantem Ergebnis öfter die Hürde zur internationalen englischsprachigen Zeitschrift nehmen. Egger *et al.* (1997) machten sich in diesem Zusammenhang auf die Suche nach randomisierten kontrollierten Studien, die in fünf führenden deutschsprachigen Fachzeitschriften veröffentlicht worden waren. Von diesen Autoren suchten Egger und Kollegen randomisiert kontrollierte Studien, die in etwa demselben Zeitraum in englischsprachigen Journalen veröffentlicht worden

waren. Insgesamt fand Egger 62 Studienpaare, von denen er 40 analysieren konnte. Die Studien, bei denen der primäre Endpunkt<sup>7</sup> statistisch signifikant war, waren von den jeweiligen Autoren eher auf Englisch und diejenigen mit nicht signifikantem Ergebnis eher auf Deutsch publiziert worden (Tabelle 1).

<b>p-Wert des primären Endpunktes</b>	<b>Veröffentlichung auf</b>	
	<b>Deutsch</b>	<b>Englisch</b>
Nicht signifikant	65% (26)	37% (15)
Signifikant	35% (14)	63% (25)

*Tabelle 1: Statistische Signifikanz und Sprache der Publikation. Daten aus Egger et al. (1997)*

*English-language bias* kann zum Tragen kommen, wenn Übersichtsartikel auf englischsprachige Publikationen beschränkt sind. Eine diesbezügliche Auswertung (Gregoire *et al.* 1995) von acht Allgemeinmedizin-Journalen zeigte, dass von 36 Metaanalysen in führenden englischsprachigen Fachzeitschriften 26 die Literatursuche auf englischsprachige Studien begrenzt hatten. Dies hatte zur Folge, dass 19 Publikationen aufgrund der Sprache ausgeschlossen wurden, obwohl sie wissenschaftlich akzeptabel waren.

Offenbar neigen Autoren dazu, positive Ergebnisse eher international zu publizieren und die negativen Ergebnisse in der Muttersprache zu veröffentlichen. Diese Art von *bias* scheint zumindest für die deutsche Sprache am Aussterben zu sein. Der Anteil der in Deutscher Sprache erschienenen Studien nahm mit der Zeit ab (Egger *et al.* 1997). Gleichzeitig gibt es immer mehr

<sup>6</sup> *Power*: Wahrscheinlichkeit, dass ein tatsächlich vorhandener Effekt in einer Studie auch gefunden wird.

<sup>7</sup> Primärer Endpunkt: Dies ist das alles entscheidende Zielkriterium der Studie.

deutsche Fachzeitschriften, die englischsprachige Artikel enthalten oder gar nur noch auf Englisch erscheinen.

Nun kann man argwöhnen, dass es ohnehin einen qualitativen Unterschied zwischen internationalen, und damit in der Regel englischsprachigen und nationalen Fachzeitschriften gibt. Schließlich sind die internationalen angesehenere und können sich daher eher die Rosinen aus dem Angebot herauspicken. Moher *et al.* (1996) verglichen die Qualität englisch- und nicht-englischsprachig publizierter kontrollierter randomisierter Studien und fanden keinen nennenswerten Unterschied. Im Gegensatz dazu kamen Juni *et al.* (2002) zu dem Ergebnis, dass die nicht-englischen Publikationen im Allgemeinen methodisch schlechter seien als diejenigen, die in englischer Sprache erscheinen. Sie kamen ferner zu dem Schluss, dass der *English language bias* die Resultate von Metaanalysen nicht entscheidend veränderte.

### **Weshalb werden Ergebnisse nicht veröffentlicht?**

Bei Befragung werden von Forschern unterschiedliche Gründe angegeben, weshalb sie ihre Ergebnisse nicht veröffentlicht haben: das Ergebnis sei nicht signifikant; der Sponsor halte die Daten zurück; Veröffentlichung sei nicht Ziel der Studie gewesen (Beispiel: Zulassungsstudien); keine Zeit oder kein Interesse mehr; „unwichtige“ Ergebnisse; Ablehnung des Manuskriptes befürchtet; andere Publikation mit ähnlichen Ergebnissen gefunden; Wechsel der Arbeitstelle usw. Methodische Probleme oder eine tatsächliche Ablehnung des Manuskripts betreffen nur jeweils ca. 10 Prozent der nicht publizierten Studien (Dickersin *et al.* 1992, Dickersin & Min

1993, Easterbrook *et al.* 1991, Horton 1997, Misakian & Bero 1998, Weber *et al.* 1998). In einer Untersuchung von Timmer *et al.* (2002) gaben die befragten Autoren Zeitmangel als Hauptgrund an, nicht zu publizieren.

Grundsätzlich sind Wissenschaftler an der Veröffentlichung kleiner negativer Studien wenig interessiert (Easterbrook *et al.* 1991; Dickersin *et al.*, 1992; Dickersin 1993; Hetherington *et al.* 1989). Einige Forscher geben auf Befragung an, dass involvierte pharmazeutische Firmen die Daten verwalten und somit diese Firmen die Publikation initiieren sollten. Häufig wurden derartige Studien nur für die Produktzulassung durchgeführt, ohne dass eine Veröffentlichung beabsichtigt war (Easterbrook *et al.* 1991). Im Jahre 1987 verzeichnete die *Finnish National Agency for Medicines* 274 Medikamentenstudien. Über 183 dieser Studien wurde bis 1993 nichts berichtet (Bardy 1996).

Lexchin *et al.* (2003) prüften, ob die Ergebnisse von Medikamentenstudien eher positiv ausfallen, wenn sie von der Pharmaindustrie finanziert werden und ob sich die methodische Qualität dieser Studien von derjenigen mit anderer Finanzierungsquelle unterscheidet. Zu dieser Fragestellung fanden sie 30 Untersuchungen. Von der Pharmaindustrie finanzierte Studien gelangten seltener zur Publikation als Studien mit anderen Finanzierungsquellen. Ferner gelangten industriefinanzierte Studien etwa viermal so häufig zu einem positiven Ergebnis als die anderen Studien. Es gab keinen offensichtlichen Hinweis darauf, dass die pharmagesponserten Studien eine schlechtere methodische Qualität hatten. Lexchin und Mitarbeiter

vermuten daher als Ursache *publication bias*, d.h. Studien mit negativen Ergebnissen bleiben unter Verschluss. Als weitere mögliche Ursache führen sie inadäquate Vergleichsbehandlungen an, die beispielsweise darin bestehen können, dass man die Kontrollgruppe mit dem Konkurrenzpräparat behandelt – allerdings hoffnungslos unterdosiert. Oder man lässt ein mittelmäßiges Pferd im Rennen gegen eine Horde Maulesel antreten und feiert es dann als Champion.

Letzteres und die Subjektivität der meisten Begründungen lassen viel Platz für Spekulation. So kann es durchaus sein, dass ein misslungener Wirksamkeitsnachweis im Rahmen einer Medikamentenprüfung aus marktpolitischen Erwägungen nicht publiziert wird oder dass ein Autor Ergebnisse zurückhält, weil sie seine bisherigen Arbeiten bzw. die Arbeiten anderer in Frage stellen könnten oder weil die praktischen Folgerungen dem Autor nicht erwünscht erscheinen.

### **Mehrfache Veröffentlichung von Studien**

Das Gegenstück zum Verschweigen unliebsamer Daten ist das umso werbewirksamere Herausposaunen „passender“ Ergebnisse durch die selektive oder mehrfache Veröffentlichung von positiven Studien. Eine Form davon ist die Aufspaltung großer multizentrischer Studien in viele kleinere Berichte, bei denen so getan wird, als ob es sich jeweils um unterschiedliche Studien handeln würde (Salamitaktik). Da die Autoren es versäumen, auf diesen Umstand hinzuweisen und sich in der Autorenschaft abwechseln, ist es sehr schwierig, derartige Mehrfachpublikationen zu erkennen (Gilbody & Song 2000).

Tramèr *et al.* (1997) untersuchten 84 Publikationen zur Wirkung von Ondansetron auf die postoperative Übelkeit. Diese Studien umfassten scheinbar die Daten von insgesamt 11.980 Patienten. Tatsächlich waren es aber nur 8.645 Patienten in 70 Studien. 38 Prozent der Patientendaten waren durch Mehrfachpublikation entstanden. Dabei wurden Studien mit großen Effekten bevorzugt mehrfach publiziert, mit der Folge, dass die Effizienz des Medikaments um 23 Prozent überschätzt wurde. Dies war in zahlreichen Artikeln, Editorials, Übersichtsarbeiten und Buchkapiteln übersehen worden (Rennie 1999). Das ist auch nicht verwunderlich, denn es gab keine klaren Querverweise zwischen den Studien, es wurde in verschiedenen Sprachen und mit völlig unterschiedlichen Autorennamen publiziert, oder es wurden Teile der Daten von Studien zu neuen vermischt. In zwei Fällen wurden „ähnliche Berichte“ zitiert – ohne Hinweis darauf, dass es sich um exakt dieselben Patienten handelte. Die Zusammenhänge kamen nur dadurch ans Licht, weil Tramèr *et al.* den Daten und Autoren mit kriminalistischer Hartnäckigkeit nachforschten, um Existenz und Wirkung verdeckter Mehrfachpublikationen aufzudecken.

Durch die verdeckte mehrfache Veröffentlichung positiver Studien entsteht der Eindruck von sehr viel mehr Studien mit obendrein ähnlichen Ergebnissen, als überhaupt durchgeführt wurden. Dies führt zu einer Überschätzung der Evidenz für diese Ergebnisse.

Eine weitere Vermehrung der Publikationen findet durch zahlreiche Metaanalysen und Übersichtsartikel statt, die sich mehr oder

weniger überschneiden. Auch hier kann inkorrektes Zitieren es schwierig bis unmöglich machen, die tatsächliche Evidenz richtig einzuschätzen.

Melander *et al.* (2003) verglichen 42 placebo-kontrollierte Studien über fünf verschiedene Antidepressiva, die zwischen 1983 und 1999 bei der *Swedish Drug Regulatory Authority* zur Genehmigung eingereicht worden waren, mit den tatsächlich daraus hervorgegangenen Publikationen. Das Ergebnis ihrer Untersuchung für eines dieser Antidepressiva zeigt Abbildung 5. Zu dem Medikament gab es

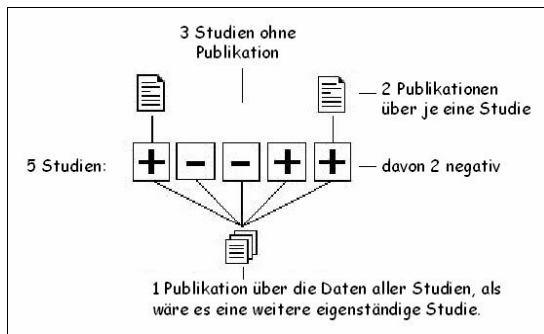


Abb. 5: Durchgeführte Studien und daraus hervorgehende Veröffentlichungen zu einem Antidepressivum (nach Melander *et al.* 2003).

insgesamt fünf Studien, drei positive und zwei negative. Zwei der drei positiven Studien waren veröffentlicht worden (☞), die beiden negativen Studien nicht. Es gab noch eine dritte Publikation, die vorgab eine unabhängige Studie gewesen zu sein. Es war aber lediglich die Zusammenfassung der Daten aller fünf Studien (☞). Die Daten der ganz linken und der ganz rechten Studie (beide positiv) sind somit doppelt veröffentlicht worden.

Das Gesamtergebnis der Untersuchung von Melander *et al.* ist in Abbildung 6 darge-

stellt. Das in Abbildung 5 dargestellte Medikament finden Sie unten rechts. Von den 42 durchgeführten Studien fielen 21 positiv aus. Und 19 der 21 positiven Studien wurden veröffentlicht. Die 21 Studien mit negativen Ergebnissen führten

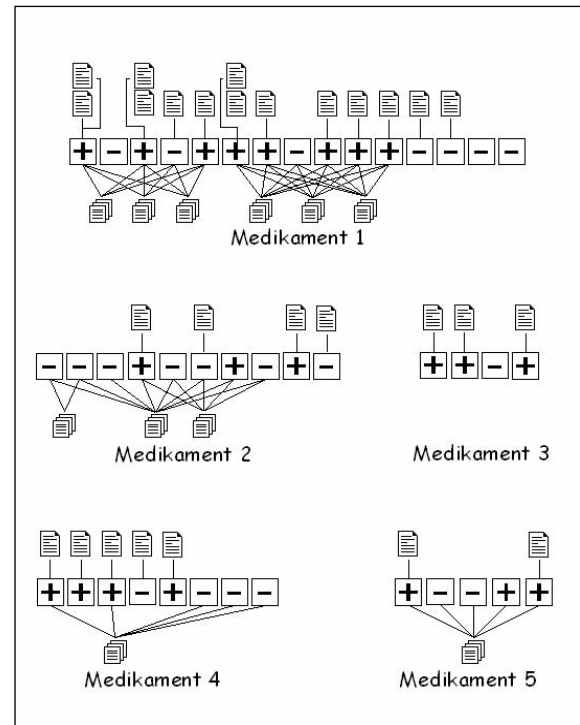


Abb. 6: Publikations-Stammbaum für fünf Antidepressiva, die in Schweden zwischen 1989 und 1994 zur Behandlung zugelassen wurden (Melander *et al.* 2003).

zu lediglich 6 Publikationen. Studien mit signifikantem Ergebnis zum Vorteil des getesteten Medikamentes wurden also mehr als dreimal so häufig publiziert wie Studien mit nicht-signifikantem Ergebnis.

Gøtzsche (1989) untersuchte 244 Studien über nichtsteroidale entzündungshemmende Medikamente zur Behandlung der rheumatischen Arthritis. Fast ein Fünftel (18%) der Studien war mehrfach publiziert worden. Aber nur in einem Viertel dieser mehrfach publizierten Arbeiten wurde darauf hingewiesen. Bei der Hälfte waren der Erstautor



und die Anzahl der Autoren verschieden und ebenfalls bei der Hälfte gab es wichtige Diskrepanzen zwischen den verschiedenen Versionen derselben Studie. So gab es Widersprüche bezüglich der durchgeführten Verblindung, des Studiendesigns, der Anzahl der ausgeschlossenen Patienten, der Anzahl der untersuchten Variablen, der beobachteten Nebenwirkungen und nicht zuletzt auch bezüglich der statistischen Signifikanz der Ergebnisse. Da sich die Suche auf die Datenbank MEDLINE beschränkte, dürfte es sich bei den ermittelten 18 Prozent Mehrfachpublikationen um eine Unterschätzung handeln, da MEDLINE nicht alle Zeitschriften abdeckt. Vor allem die nicht-englischsprachigen Zeitschriften sind unterrepräsentiert.

Huston & Moher (1996) fanden ähnliches, als sie versuchten eine Metaanalyse über das Psychopharmakon Risperidon durchzuführen. Zunächst fanden sie 20 Artikel und verschiedene nicht publizierte Berichte über randomisierte doppelblinde Studien. Tatsächlich waren es jedoch nur sieben kleine und zwei große Studien. Von den großen Studien war eine insgesamt sechsmal veröffentlicht worden – immer mit unterschiedlichen Autoren. Dadurch war in der Fachwelt der falsche Eindruck einer positiven Wirkung des Medikaments entstanden, welche sogar mehrfach reproduziert worden sei. Huston & Moher berichten sogar von einer Studie, die in derselben Zeitschrift innerhalb von fünf Jahren zweimal publiziert wurde (McMillan 1977, 1982). Sie stellen die Frage, ob man derart skrupellosen Autoren überhaupt trauen und die sachgemäße Durchführung ihrer Untersuchungen glauben könne. Sie fordern mehr Transparenz, beispielsweise

durch Veröffentlichung von Hinweisen auf Mehrfachpublikation in den Arbeiten.

Das verdeckte Mehrfachpublizieren derselben Daten führt zu einer Überschätzung der Wirksamkeit beispielsweise eines neuen Medikaments. Es liegt auf der Hand, dass dies zum Schaden der Patienten ist, aber kommerziellen Interessen entgegenkommt – kurzfristig betrachtet. Langfristig jedoch verlieren solche Unternehmen ihr Image und ihre Glaubwürdigkeit. Im Jahre 1996 entschied sich *Schering Health Care* dafür, die Information über alle laufenden kontrollierten Studien im *Cochrane Controlled Trials Register* anzumelden. Im Jahre 1998 zog *Glaxo Wellcome* nach. Es gibt kaum eine bessere Maßnahme für ein Pharmaunternehmen, um seine guten ethischen, klinischen und wissenschaftlichen Absichten zu dokumentieren (Rennie 1999).

### ***Selective reporting***

Eine andere Art von Unausgewogenheit entsteht durch das selektive Berichten über bestimmte Arten der Auswertung. 41 von 42 der *Swedish Drug Regulatory Authority* vorgelegten Studienberichte enthielten zwei oder mehr alternative Datenanalysen („*intention to treat*“<sup>8</sup> und „*per protocol*“<sup>9</sup>).

<sup>8</sup> *intention to treat*: Alle randomisierten Patienten werden in der Gruppe ausgewertet, der sie ursprünglich zugeordnet wurden. Bei diesem Verfahren kann kein einziger Patient zwischen Randomisierung und Auswertung verloren gehen. Daher ist es frei von nachträglicher Selektion.

<sup>9</sup> *per protocol*: Nur die Patienten, die exakt nach Protokoll behandelt wurden, werden ausgewertet. Patienten, die wegen ausbleibender Wirkung, eintretenden Nebenwirkungen oder schwerer Krankheit die Behandlung abbrechen, gelangen nicht in die Auswertung. Daher sind die Ergebnisse aggressiver Therapien, vor allem bei placebokontrol-

Diese Berichte führten zu 28 Publikationen, in denen in nur zwei Fällen beide Analysen präsentiert wurden. Ansonsten wurde nur über das meist industriefreundlichere Ergebnis einer „per protocol“ Analyse berichtet. Dies führt zu einer erheblichen Überschätzung des Nutzens der getesteten Medikamente (Melander *et al.* 2003).

Nicht alle Autoren haben das Glück und die Unverschämtheit, mit ein und denselben Daten mehrfach Lorbeer zu ernten. Es kann auch schwierig sein, eine Arbeit überhaupt zu publizieren. Wenn ein Manuskript aber erst einmal geschrieben ist, wird es meist solange bei verschiedenen Zeitschriften eingereicht, bis es jemand annimmt. Wissenschaftler stehen unter Druck, möglichst viel und in möglichst hochrangigen Zeitschriften zu veröffentlichen. Bei der ersten Einreichung zielt der Autor im Allgemeinen auf das Journal mit dem höchsten Ansehen, das sein Paper nach eigener Einschätzung gerade noch nehmen müsste. In der Regel wird das ein internationales englischsprachiges Blatt sein. Wird die Arbeit vom Traumblatt abgelehnt, gibt ein erfahrener Autor nicht auf. Das Manuskript hat ja viel Mühe und Zeit gekostet und der Autor kann seine Existenz nicht mit abgelehnten Arbeiten rechtfertigen. Das Manuskript landet also auf dem Schreibtisch des Herausgebers der Zeitschrift mit dem nächst niedrigen Rang. Mit der Anzahl der Versuche nimmt das Ansehen der gewählten Zeitschrift ab, damit meist auch der Verbreitungsgrad und am Ende lauert ein nationales Blatt, das den Artikel in Landessprache herausgibt. Wenn nun negative Ergebnisse mit höherer Wahrscheinlichkeit abgelehnt werden als

positive, dann liegt es auf der Hand, dass erstere im Mittel eine längere Strecke auf dem Leidensweg vom internationalen Top-Journal zum Bezirksblatt zurücklegen. Da am Anfang die Eigeneinschätzung der Autoren eine Rolle spielt und negative Ergebnisse auch von den Autoren gering geschätzt werden, beginnt der Weg negativer Studien obendrein bereits bei Zeitschriften mit geringerem Ansehen.

*Insgesamt bleibt festzuhalten, dass es für Forscher zahlreiche Beweggründe gibt, Ergebnisse zurückzuhalten. Dies betrifft in erster Linie negative Ergebnisse. Insbesondere bei positiven Ergebnissen besteht die Tendenz, diese mehrfach zu veröffentlichen. Die (veröffentlichte) Literatur über publication bias zeigt also, dass die Autoren in erheblichem Ausmaß zur Unausgewogenheit der Berichterstattung in der Wissenschaft beitragen.*

---

lierten Studien, bei dieser Art der Auswertung besonders optimistisch.

### 3.3. Der Beitrag der Fachzeitschriften zum *publication bias*

Wir haben oben bereits gesehen, dass Manuskripte mit positivem Ergebnis den Parcours vom Einreichen bis zum Druck schneller zurücklegen als Manuskripte mit negativen Resultaten. In diesem Abschnitt geht es um die Hindernisse, auf die ein Manuskript bei Herausgebern und Gutachtern (*peer reviewer*) treffen kann.

#### Eingereichte Manuskripte mit negativen Ergebnissen werden seltener zur Veröffentlichung angenommen

Das selektive Einreichen von Manuskripten scheint der Hauptgrund für *publication bias* zu sein (Egger & Smith 1998). Aber es tritt auch eine signifikante Verzögerung nach Einreichen des Manuskriptes auf. Ioannidis (1998) und Stern & Simes (1997) stellten in ihren Untersuchungen fest, dass lediglich Studien mit negativen Ergebnissen von mehr als einer Zeitschrift abgelehnt wurden. Es gibt also auch eine selektive Akzeptanz von Arbeiten mit positiven Ergebnissen durch die Fachzeitschriften. Ein eklatantes Beispiel hierfür sind die mittlerweile revidierten Autoreninstruktionen aus *Diabetologia* (Manuscript guideline. *Diabetologia* 1984): *“mere confirmation of known facts will be accepted only in exceptional cases; the same applies to reports of experiments and observations having no positive outcome.”*

In einer prospektiven Kohortenstudie<sup>10</sup> (Olson *et al.*, 2002) wurden die von Februar

<sup>10</sup> Kohortenstudie: Eine definierte Patientengruppe (eine Kohorte) wird über einen bestimmten Zeitraum beobachtet, um zu sehen, wie viele einen gewissen Endpunkt entwickeln.

Dabei wird die Studienpopulation entweder dadurch gewonnen, dass vergleichbare Personen die entweder

1996 bis August 1999 bei *JAMA* (*Journal of the American Medical Association*) eingereichten Manuskripte ausgewertet. Berücksichtigt wurden alle 745 prospektiven, kontrollierten Studien, die einen statistischen Test zum Vergleich der Ergebnisse zwischen den Gruppen verwandten. Die Autoren kommen zu dem Ergebnis, dass bei *JAMA* eingereichte Arbeiten mit negativen Ergebnissen nicht seltener zur Veröffentlichung angenommen werden, als Manuskripte mit positiven Ergebnissen. Die Ergebnisse dieser Untersuchung sind jedoch fragwürdig. Denn Manuskripte, bei denen kein primärer Endpunkt angegeben worden war, wurden als „negativ“ eingestuft, wenn die *Mehrzahl* der untersuchten Endpunkte nicht signifikant war. Häufig werden jedoch diverse Endpunkte untersucht und die Studie wird von den Autoren und den Lesern als „positiv“ bewertet, wenn auch nur *einer* dieser Endpunkte statistisch signifikant ist (vgl. Ioannidis 1998). Trotz der fragwürdigen Einteilung ergab die Untersuchung eine geringe Bevorzugung positiver Manuskripte (20 Prozent vs. 15 Prozent)<sup>11</sup>. Der Schluss liegt nahe, dass Manuskripte mit negativen Ergebnissen bei *JAMA* sehr wohl benachteiligt werden.

---

exponiert sind oder nicht. (z.B. Raucher und Nichtraucher) eingeschlossen werden und ein definierter Endpunkt wird beobachtet. Oder eine definierte Population, z.B. die gesamte Bevölkerung einer Kleinstadt wird ausgewählt, diese dann untersucht auf das Vorliegen von Risikofaktoren (z.B. Hypercholesterinämie, Hypertonie) für einen bestimmten Endpunkt.

<sup>11</sup> Dieser Unterschied ist nach Aussage von Olson *et al.* nicht statistisch signifikant. Dies gilt jedoch nur für einen zweiseitigen Signifikanztest, der hier nicht angebracht ist. Man müsste sonst davon ausgehen, dass auch Manuskripte mit *positiven* Resultaten hätten *benachteiligt* sein können. Dies ist aufgrund der Vorkenntnisse aber eine ziemlich abwegige Annahme. Bei einem einseitigen Signifikanztest wäre die Bevorzugung statistisch signifikant gewesen. Die Studie zeigt also eher das Gegenteil dessen, was Olson *et al.* (2002) am Ende behaupten.

### **Konfirmatorischer bias (*confirmatory bias*)**

*Confirmatory bias* ist die Tendenz der Gutachter von Fachzeitschriften, Ergebnisse, die mit allgemein akzeptierten Theorien übereinstimmen, eher zu akzeptieren als solche, die ihnen widersprechen. Mahoney (1977) ließ 67 Gutachter verschiedene Versionen ein und desselben Berichtes über ein fiktives Experiment begutachten. Wenn die Ergebnisse der allgemein akzeptierten theoretischen Erwartung widersprachen, waren die Gutachter eher dazu geneigt, kritische Fragen zur Methodik zu stellen und die Arbeit abzulehnen.<sup>12</sup>

### **Länder bias (*nationality bias*)**

Link (1998) untersuchte, ob Gutachter aus den USA Manuskripte anders beurteilen als Gutachter aus anderen Ländern (*reviewer bias*). Dabei wurde auch die Herkunft der Manuskripte berücksichtigt und nach US- und Nicht-US-Manuskripten unterschieden. Grundlage der Untersuchung waren alle Manuskripte, die in den Jahren 1995 und 1996 bei der Zeitschrift *Gastroenterology* eingereicht worden waren. Die Gutachter unterteilten diese in vier Kategorien: 1.) annehmen, 2.) vorläufig annehmen, 3.) vorläufig ablehnen mit der Möglichkeit zur Wiedereinsendung nach vollständiger Überarbeitung und 4.) endgültig ablehnen.

<sup>12</sup> Dieses Verhalten der Gutachter ist im Bayesschen Sinne insofern völlig vernünftig, als die *a priori* Wahrscheinlichkeit, dass völlig unerwartete Ergebnisse auf korrektem Wege gewonnen wurden, deutlich geringer ist als bei Resultaten, die unsere Vorurteile bestätigen (vgl. hierzu Beck-Bornholdt & Dubben, 2003). Daher ist eine besonders kritische Beurteilung erstaunlicher Studienergebnisse durchaus angebracht. Im Popperschen Sinne ist das allerdings grober Unfug, da das Falsifizieren dadurch behindert wird.

Die Ergebnisse sind in Tabelle 2 wiedergegeben.

Die Manuskripte, die nicht aus den USA kamen, wurden von den Gutachtern beider Gruppen sehr ähnlich beurteilt ( $p=0,31$ ). Im Gegensatz dazu waren US-Manuskripte für Gutachter aus den USA „überzeugender“ als für ihre nicht US-amerikanischen Kollegen ( $p=0,001$ ). US-Gutachter zeigten eine signifikante Vorliebe für Manuskripte aus den USA ( $p=0,001$ ). Diese Vorliebe findet sich weniger ausgeprägt auch bei den nicht US-amerikanischen Gutachtern ( $p=0,09$ ). Die Arbeit weist somit auf einen deutlichen Länder bias bei den US-Amerikanern hin. Es ist ihnen allerdings zugute zu halten, dass sie dies überhaupt untersucht und veröffentlicht haben. Unseres Wissens sind ähnliche Untersuchungen in anderen Ländern noch nicht durchgeführt worden.

Herkunft der Manuskripte	Nicht aus USA		Aus USA	
	USA	Nicht-USA	USA	Nicht-USA
Annehmen	3,6 %	3,2 %	7,0 %	3,6 %
Vorläufig annehmen	24,9 %	24,6 %	31,3 %	30,5 %
Vorläufig ablehnen	26,8 %	24,6 %	26,1 %	21,6 %
Endgültig ablehnen	44,7 %	47,6 %	35,6 %	44,3 %

*Tabelle 2: Urteil der US-amerikanischen und der nicht-US-amerikanischen Gutachter in Abhängigkeit von der Herkunft der Manuskripte.*

*Insgesamt entsteht der Eindruck, dass bei den Zeitschriften positive und „passende“ Ergebnisse erwünschter sind als deren Gegenstücke (s.a. Abschnitt 4.2).*

### 3.4. Der Beitrag der Leser zum *publication bias*

Auch der Endverbraucher wissenschaftlicher Veröffentlichungen ist nicht objektiv. Er sucht nicht objektiv, wenn er z.B. bei der Suche nach Literatur nicht-englischsprachige Studien ausschließt. Er liest lieber etwas über positive Ergebnisse, als einen Text, der ihm verrät, dass irgendetwas nicht funktioniert. Er liest und zitiert am liebsten das, was seine eigene Auffassung bestätigt. Da ist der Forscher ganz Mensch, die beiden Autoren dieser Broschüre natürlich eingeschlossen. Beim Aufarbeiten der Literatur zu diesem Thema ist uns aufgefallen, dass auch wir selbst Untersuchungen, die die Existenz von *publication bias* feststellten, viel ernster nahmen, als Untersuchungen, die keinen *publication bias* gefunden haben. Als uns das auffiel, stellten wir uns umgehend die Frage, ob es *publication bias* nicht auch in der Literatur über *publication bias* geben könnte. Die irritierenden Ergebnisse der sich daraus ergebenden Recherche sind in Kapitel 8 „Quergedacht“ dargestellt.

#### Selektives Lesen

Daniels *et al.* (2002) hatten festgestellt, dass in ihrer hausinternen Internisten-Konferenz häufiger Studien mit positivem als mit negativem Ergebnis vorgestellt wurden. Sie gingen der Frage nach, ob hierbei die Vorliebe der Referenten eine Rolle spielen könnte. Dazu fertigten sie *Abstracts* an, die eine Hälfte mit positivem und die andere mit negativem Ergebnis, und ließen die Referenten dann auswählen. In 59% der 180

Fälle wurde ein positiver Abstract gewählt, in 41% eine negativer.<sup>13</sup>

#### *Inclusion bias* und *Selection bias*

Wissenschaftler, die eine Metaanalyse durchführen, haben im Allgemeinen Vorkenntnisse über die Ergebnisse in Frage kommender Studien. Diese Vorkenntnisse können die Einschlusskriterien für Studien, die in die Metaanalyse eingehen, beeinflussen (Egger & Smith 1998). In einer Metaanalyse von sieben Studien über Cholesterin-Senkung nach Myokardinfarkt wurde ein (nicht-signifikanter) Vorteil durch die Intervention festgestellt. Eine Studie, die einen Nachteil durch Cholesterin-Senkung zeigte, wurde nicht berücksichtigt, obwohl sie sämtliche Einschlusskriterien erfüllte. Die eingeschlossenen Studien wurden in den ersten fünf Jahren nach ihrem Erscheinen im Mittel 20mal pro Jahr zitiert, die ausgeschlossene nur einmal (*citation bias*). Elf weitere Studien, die zu der Zeit vorlagen, aber die etwas willkürlichen Einschlusskriterien nicht erfüllten, zeigten alle einen signifikanten Nachteil in der Gesamtsterblichkeit durch Cholesterin-Senkung (Egger & Smith 1998).

#### Anzeigen für Medizinprodukte in Fachzeitschriften

Mit zunehmendem Einfluss der Evidenzbasierten Medizin enthalten Anzeigen für Arzneimittel in Fachzeitschriften immer häufiger bibliographische Hinweise. Villanueva *et al.* (2003) untersuchten anhand der Anzeigen für Antihypertensiva und für Lipidsenker in sechs spanischen Fachzeitschriften des Jahrgangs 1997,

<sup>13</sup> Im Gegensatz zur Angabe der Autoren ist diese Diskrepanz nach unserer Berechnung statistisch signifikant ( $p = 0,02$ , Binomialtest).

inwieweit die angeführten Studien tatsächlich die Aussagen der Werbung stützten. Bei der Bewertung war man sehr großzügig: So wurde ein bibliographischer Hinweis selbst dann als die Werbeaussage stützend gewertet, wenn es sich lediglich um ein Zitat aus der Einleitung oder der Diskussion handelte und auch wenn die Qualität der zitierten Arbeit sehr schlecht war. Die bibliographischen Hinweise in den Annoncen stammten zu etwa zwei Dritteln aus den 20 weltweit meistzitierten Fachzeitschriften. Trotz großzügiger Bewertung zeigte sich, dass fast die Hälfte (45 von 102) der Werbeaussagen nicht durch die zitierte Studie belegt wurde. In 20 Fällen wurde in der Werbung ein Medikament einer Patientengruppe empfohlen, die sich von der in der zitierten Studie untersuchten Gruppe unterschied: in 15 Fällen wurde aus einer Studie an Hochrisikogruppen auf alle Patienten geschlossen, in vier Fällen wurde das Medikament spezifisch Patienten empfohlen, die aus der Studie ausgeschlossen worden waren und in einem Fall wurde von *in vitro* und tierexperimentellen Untersuchungen auf den Menschen geschlossen. In zehn Fällen wurde die Wirksamkeit übertrieben dargestellt, in neun Fällen wurden Risikoreduktionen behauptet, die nicht nachgewiesen waren und in sechs Fällen hatten die Studien überhaupt nichts mit dem beworbenen Produkt zu tun. Die Autoren der Untersuchung schließen aus ihren Ergebnissen, dass bei den bibliographischen Hinweisen in der Werbung erhebliche Skepsis geboten ist.

*Insgesamt zeigt sich, dass Leser und sonstige Verbraucher publizierter Ergebnisse die ohnehin schon bestehende Unausgewogenheit der Berichterstattung durch selektives Lesen, selektierende Ein- und Ausschlusskriterien für Metaanalysen und Übersichtsartikel sowie selektierte und falsche Literaturzitate noch weiter verstärken.*

## 4. Beispiele für unausgewogene Berichterstattung in der Wissenschaft

Gegenstand dieses Kapitels sind ausgesuchte Beispiele für unausgewogene Berichterstattung (*publication bias*) in der medizinischen Forschung und die sich daraus für Patienten ergebenden Konsequenzen.

### 4.1 *Publication bias* bei blutdrucksenkenden Medikamenten

Im Jahre 1990 erschien im *Lancet* eine später viel zitierte Metaanalyse aller bis 1994 vorliegenden randomisierten Studien zur medikamentösen Behandlung des Hypertonus (Collins *et al.* 1990, mit einer Aktualisierung von Collins & MacMahon 1994). Insgesamt sind 17 Studien in die Analyse eingeflossen. Eine dieser Studien besteht aus drei unabhängigen Untersuchungen, so dass insgesamt die Daten von 19 Untersuchungen vorliegen. Sie umfassen die Daten von 48.000 Patienten mit einer durchschnittlichen Nachbeobachtungszeit von 5 Jahren.

Die Autoren geben an, dass es ihres Wissens keine bis dahin publizierten oder nicht publizierten Studien gäbe, die nicht in die Analyse eingeflossen wären. Mit Hilfe der Trichtergrafik kann diese Aussage getestet werden. Wie weiter oben in „Methoden zur Erkennung von *publication bias*“ gezeigt, streuen die Ergebnisse kleiner Studien stärker um den tatsächlichen Wert als die Ergebnisse großer Studien. Dies trifft auch auf die Wirksamkeit von blutdrucksenkenden Medikamenten zu, wie die

folgende Grafik zeigt (Abbildung 7). Als Maß für die Messgenauigkeit wurde hier die Anzahl der beobachteten Schlaganfälle gewählt. Auf der linken Seite der senkrechten Trennlinie liegen die Studien, bei denen durch das Medikament die Anzahl der Schlaganfälle reduziert wird. Rechts ist der Bereich, bei dem die Kontrollbehandlung besser war als das Verum. Der durch die Daten eingenommene Trichter ist unsymmetrisch. Es scheinen einige kleine Studien zu fehlen, deren Ergebnisse rechts von der gestrichelten Linie liegen müssten. Die Asymmetrie der Punkte ist ein deutlicher Hinweis auf *publication bias*. Die Wirksamkeit ist folglich geringer, als die Übersichtsarbeit suggeriert.

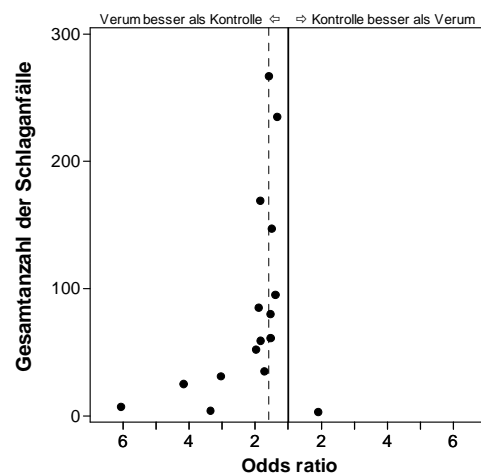


Abbildung 7: Trichtergrafik (funnel plot) aller Studien aus dem Übersichtsartikel von Collins *et al.* (1990, 1994) zur Abnahme des Schlaganfallrisikos bei Blutdrucksenkung.

## 4.2 Modeerscheinungen in der Wissenschaft: Durch Pausen unterbrochene Strahlentherapie von Tumoren.

In der Wissenschaft gibt es Modeerscheinungen wie in der Haute Couture. Dass sich die bearbeiteten Themen mit dem Wind und dem Geldgeber drehen, ist bekannt und zu erwarten. Erstaunlich ist jedoch, dass auch die erzielten Resultate saisonbedingt sind. Ein ausgeprägtes Beispiel zeigt eine systematische Übersichtsarbeit aus der Radio-Onkologie (Dubben *et al.* 2001). Im Allgemeinen erstreckt sich die strahlentherapeutische Behandlung von Tumoren über mehrere Wochen. Die Autoren untersuchten, welche Evidenz zu der Frage vorliegt, ob eine Bestrahlungspause von mehreren Tagen oder Wochen negative Folgen für den Behandlungserfolg hat. Hierbei stellten sie fest, dass Untersuchungen der 70er Jahre fast einhellig zu dem Ergebnis führten, dass sich Bestrahlungspausen positiv auswirken. Seit Mitte der 80er Jahre ist jedoch zu beobachten, dass die Untersuchungen genau zum gegenteiligen Resultat gelangen (siehe Abbildung 8).

Es ist unwahrscheinlich, dass diese Daten eine Kehrtwende in den Naturgesetzen signalisieren, spiegelt die Grafik sehr wahrscheinlich einen Wechsel der Mode wider. Die veröffentlichte Meinung reflektiert daher nicht unbedingt die tatsächliche Sachlage, sondern zum überwiegenden Teil den so genannten *main stream*, die herrschende Meinung. Während früher durch Pausen unterbrochene Strahlentherapien als modern und vorteilhaft galten, sind sie heute verpönt. Früher glaubte man, dass sich das mitbestrahlte Normalgewebe in einer Pause schneller

erholt als der Tumor. Heutzutage glaubt man das Gegenteil. Auslöser für den Umschwung ist eine strahlenbiologische Arbeit aus dem Jahre 1973, in der die Reaktion von Mäusehaut auf Strahlung untersucht wurde (Denekamp, 1973).

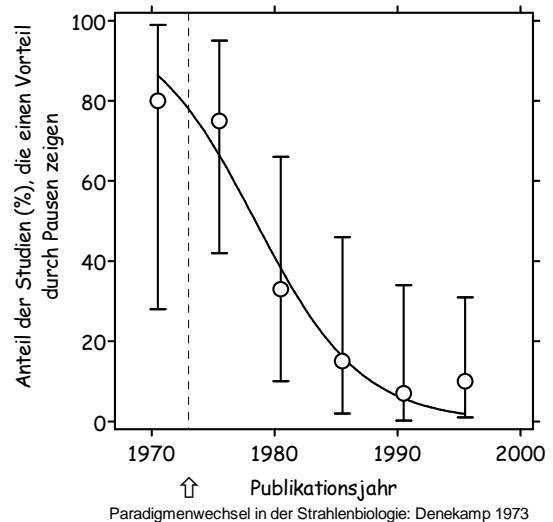


Abb. 8: Anteil „positiver“ Publikationen über Pausen während einer Strahlentherapie in Abhängigkeit vom Erscheinungsjahr. Eine Studie wurde als „positiv“ bezeichnet, wenn sie eine von Pausen unterbrochene Strahlentherapie als vorteilhaft gegenüber einer kontinuierlichen Behandlung darstellte. Gruppen von je 5 Jahren, erste Gruppe von 1966 bis Mitte 1972. 95%-Vertrauensbereiche nach Binomialstatistik. Die Verringerung des Anteils mit der Zeit ist hoch signifikant ( $p < 0,000002$ ;  $\leq 1982$  vs.  $> 1982$ ; Chi-Quadrat-Test).

Es ist denkbar, dass Patienten durch die gegenwärtige Ächtung von Pausen eine nützliche Therapieoption vorenthalten wird. Ebenso ist es möglich, dass Patienten suboptimal behandelt werden, wenn Behandlungspausen auftreten, wie z.B. an Wochenenden, an Feiertagen und bei Geräteausfall. Wegen der unausgewogenen Berichterstattung in der Wissenschaft und



der unzureichenden Trennschärfe der durchgeführten Studien, ist auch nach mehr als 100 Jahren Strahlentherapie der Einfluss von Behandlungsunterbrechungen nicht wissenschaftlich geklärt.

### **4.3 *Publication bias* bei der Prüfung diagnostischer Tests**

Die unausgewogene Berichterstattung treibt ihr Unwesen nicht nur in Metaanalysen, die über Behandlungseffekte berichten, sondern auch in Arbeiten über diagnostische Tests. Song *et al.* (2002) untersuchten 28 Metaanalysen. In fast allen scheint es *publication bias* zu geben, davon in 12 Fällen mit statistischer Signifikanz. Dabei berichten kleinere Studien jeweils über größere Testgenauigkeit. Die Überschätzung der Qualität diagnostischer Testergebnisse dürfte ein immenses Schädigungspotential für Gesundheit und Ressourcen des Gesundheitswesens besitzen.

## **5. Die Folgen unausgewogener Berichterstattung in der Wissenschaft**

Das Zurückhalten von Studienergebnissen hat nach Horton (1997) drei mögliche Folgen für die Patienten:

1. Da Studien mit viel versprechenden Ergebnissen eher veröffentlicht werden, entsteht der Eindruck einer höheren Wirksamkeit von untersuchten Behandlungen. Daher könnten Patienten nutzlosen oder sogar schädlichen Behandlungen ausgesetzt werden.

2. Der Nachweis kleiner, aber klinisch bedeutsamer Effekte wird erschwert, wenn Studien mit moderaten Ergebnissen nicht publiziert werden, denn damit sinkt die *Power* systematischer Übersichtsarbeiten. Daher könnten Patienten effektive Behandlungen vorenthalten werden.
3. Patienten könnten um die Teilnahme an Studien gebeten werden, deren Fragestellung in Wirklichkeit längst erforscht ist.

Die Veröffentlichungen in den medizinischen Fachzeitschriften haben erheblichen Einfluss auf die Behandlungskonzepte für Patienten. Die besprochenen Verzerrungen bewirken, dass Studien mit signifikanten Ergebnissen bevorzugt, schneller und prominenter publiziert werden, und dass diese Arbeiten in Literaturrecherchen entsprechend häufiger, früher und einfacher wieder aufgefunden werden. Daher muss davon ausgegangen werden, dass in Übersichtsarbeiten und Metaanalysen die Behandlungseffekte überschätzt werden und zum Teil Effekte gesehen werden, wo gar keine sind.

Obwohl *publication bias* seit Jahrzehnten bekannt ist, werden seine Folgen in der wissenschaftlichen Literatur kaum berücksichtigt. So zeigte eine Untersuchung über die Qualität systematischer Übersichtsarbeiten, dass nur bei 3 bis 7 Prozent der Metaanalysen der Versuch unternommen wurde, mit Hilfe einer Trichtergrafik zu prüfen, ob *publication bias* vorliegt (Linde *et al.* 1997).

*Publication bias* hat entscheidenden Einfluss auf die Leitlinien, die von den verschiedenen Fachgesellschaften zur

Behandlung von Erkrankungen erstellt werden. Dies gilt selbstverständlich auch für evidenzbasierte Leitlinien. *Publication bias* ist eine der zentralen Schwierigkeiten auf dem Weg zu einer Evidenzbasierten Medizin, denn diese setzt zwingend voraus, dass alle Studien umfassend bzw. in gleicher Zugänglichkeit der Öffentlichkeit zur Verfügung stehen. Sutton und Mitarbeiter (2000) untersuchten alle systematischen Übersichtsarbeiten in der *Cochrane Database of Systematic Reviews* (1998, issue 3). Berücksichtigt wurden alle 48 Metaanalysen mit mindestens 10 Studien bei denen mindestens ein dichotomer Endpunkt<sup>14</sup> untersucht wurde. Etwa die Hälfte der Metaanalysen war von *publication bias* betroffen. Bei vier der 48 Metaanalysen ergaben sich nach Korrektur für *publication bias* erhebliche Änderungen bei den Schlussfolgerungen. Drei Metaanalysen mit zunächst signifikanten Ergebnissen waren nach der Berücksichtigung von *publication bias* nicht mehr signifikant. Eine Metaanalyse die als nicht-signifikant und indifferent eingestuft worden war, stellte sich nach der Analyse als signifikant negativ heraus. Die Autoren schließen aus ihren Ergebnissen, dass die Ergebnisse der Metaanalysen in etwa einem Zehntel der Fälle durch *publication bias* erheblich beeinflusst werden.

LeLorier *et al.* (1997) verglichen die Ergebnisse von großen randomisierten kontrollierten Studien mit mehr als 1000 Patienten, die im *New England Journal of Medicine*, im *Lancet*, in den *Annals of Internal Medicine* oder im *Journal of the American Medical Association* zwischen 1991 und 1994 publiziert wurden, mit den

Resultaten vorausgegangener Metaanalysen. Sie fanden 12 Studien und 19 Metaanalysen, die dieselben Fragestellungen bearbeiteten. Insgesamt wurden 40 Vergleiche mit primären und sekundären Endpunkten angestellt. Die Übereinstimmung bezeichnen die Autoren als mittelmäßig. Der positive und der negative prädiktive Wert der Metaanalysen betrug jeweils 67 Prozent. Das bedeutet, wenn es keine nachfolgende Studie gegeben hätte, dass in 33 Prozent der Fälle eine ineffektive Behandlung etabliert worden wäre, bzw. ebenfalls in 33 Prozent der Fälle eine effektive Behandlung nicht übernommen worden wäre. Etwa die Hälfte (46%) der divergierenden Ergebnisse kamen dadurch zustande, dass auf eine positive Metaanalyse (Behandlung signifikant besser) eine negative (Behandlung nicht besser) randomisierte Studie folgte. Diese Abfolge kann durch *publication bias* erklärt werden, weil positive Ergebnisse schneller veröffentlicht und zu Metaanalysen verarbeitet werden als negative. In 54 Prozent der divergierenden Ergebnisse folgte auf eine negative Metaanalyse eine positive randomisierte Studie. Dies kann an der Heterogenität der in die Metaanalyse aufgenommenen Studien liegen. Heterogenität kann entstehen durch unterschiedliche Einschlusskriterien hinsichtlich Patientencharakteristika und Erkrankung, unterschiedliche primäre Behandlung, Unterschiede bei etwaigen die Therapie unterstützenden Maßnahmen usw.

---

<sup>14</sup> Dichotomer Endpunkt: Das Ergebnis lautet „ja“ oder „nein“.

## 6. *Publication bias* und Evidenzbasierte Medizin

Eine wichtige Säule der Evidenzbasierten Medizin (EBM) ist das Zusammentragen klinischer Ergebnisse in systematischen Übersichtsarbeiten und Metaanalysen. Man kann vernünftigerweise jedoch nicht erwarten, dass eine solche Meta-Arbeit besser ist als die Studien, die darin eingegangen sind. Mitunter reicht ein einziges faules Ei, um ein bis dahin aussichtsreiches Omelett komplett zu verderben. Es gibt zahlreiche Möglichkeiten zu positiven, aber nicht reproduzierbaren Ergebnissen zu gelangen (Beck-Bornholdt & Dubben, 2001). Die wichtigsten und folgenschwersten seien hier genannt:

1. Nachträgliches Ausschließen von Patienten von der Auswertung. Dies erhöht die Chance auf ein falsch positives Ergebnis.
2. Studien mit zu wenigen Patienten und folglich zu geringer statistischer *Power*. Solche Studien produzieren nicht nur eine Vielzahl falsch negativer Befunde, sondern paradoxerweise auch noch einen hohen Anteil falsch positiver Ergebnisse (Beck-Bornholdt & Dubben, 2003).
3. Multiples Testen, unklare Studienzielgrößen und ein unzureichendes Signifikanzniveau. Durch multiples Testen, z.B. durch ein unklar definiertes Studienende und durch Subgruppenanalysen wird die Wahrscheinlichkeit für falsch positive Ergebnisse enorm erhöht.

Durch *publication bias*, und natürlich auch durch andere Arten wissenschaftlicher Unredlichkeit, wird der Anteil der positiven Falschmeldungen noch weiter erhöht. Da es nicht möglich ist, Falschmeldungen

zuverlässig als solche zu erkennen, ist unausgewogene Berichterstattung auch für die Evidenzbasierte Medizin ein Problem. Es besteht die Gefahr, dass unwissenschaftlich erhobene positive Ergebnisse, durch *publication bias* angereichert, zu einem dann selbstverständlich positiven Ergebnis verbacken werden und durch das Etikett „evidenzbasiert“ bzw. „EBM“ das Mäntelchen der Wissenschaftlichkeit erhalten.

Evidenzbasierte Medizin kann den Patienten nur dann zugute kommen, wenn zahlreiche wissenschaftliche und methodische Voraussetzungen erfüllt sind – und zwar bei jeder einzelnen Studie.

In den letzten Jahren häufen sich die Berichte in den öffentlichen Medien über wissenschaftlichen Betrug. Dass gefälschte Daten erheblichen Schaden anrichten können, liegt auf der Hand. Nach Auffassung von Chalmers (1990) hat allerdings die unzureichende Veröffentlichung klinischer Studien wahrscheinlich mehr Schaden angerichtet als die absichtliche Fälschung von Daten. Dies sei ein Unrecht gegenüber den teilnehmenden Patienten, den kooperierenden Wissenschaftlern, den Geldgebern und der Gesellschaft.

Einen unentbehrlichen Schritt hin zu einer nutzbringenden Evidenzbasierten Medizin besprechen wir im nächsten Abschnitt.

## 7. Gegenmaßnahmen

In der Deklaration des Weltärztebundes von Helsinki (52. Generalversammlung des

Weltärztebundes, Edinburgh, Schottland, Oktober 2000) wurde festgehalten:

„Die Pläne aller Studien sind der Öffentlichkeit zugänglich zu machen.“

„Positive, aber auch negative Ergebnisse müssen veröffentlicht oder der Öffentlichkeit anderweitig zugänglich gemacht werden.“

Das bisher Dargestellte zeigt, dass wir hiervon noch sehr weit entfernt sind. Nach Auffassung von Chalmers liegt es in der Verantwortung der Wissenschaftler, der Ethikkommissionen, der Einrichtungen zur Forschungsförderung und der Herausgeber von Fachzeitschriften auf ausgewogene Berichterstattung in der Wissenschaft hin zu arbeiten.

Am besten wäre es, *publication bias* von vornherein zu vermeiden. Chalmers (1990) und Dickersin (1990) schlugen vor, klinische Studien prospektiv zu registrieren, da die vorherige Aufnahme von Studien in ein Register ganz sicher unabhängig von deren zukünftigen Ergebnissen ist. Die Registrierung würde darüber hinaus das Auffinden publizierter Studien erleichtern (Jull *et al.* 2002). Ein Register würde nicht nur *submission bias* und *publication bias* minimieren, sondern auch:

- Patienten und Kliniker über Studien informieren, an denen sie teilnehmen könnten
- Doppelforschung und Doppelförderung minimieren
- Lücken in gegenwärtigen Forschungsaktivitäten deutlich machen
- Zusammenarbeit von Forschern fördern und zu multizentrischen Studien anregen

Eine denkbare Strategie wäre es, Ethikkommissionen mit einzubeziehen. Diese sind ohnehin über alle Studien an Patienten unterrichtet und keine kann ohne deren Genehmigung von statten gehen. Es wäre für diese Kommissionen relativ einfach, auf eine Registrierung der Studien gemäß der *International Standardised RCT Numbers* zu bestehen (Jull *et al.* 2002). Die erhaltene Nummer gilt als Beweis für die Registrierung. Dass so etwas machbar ist, sieht man daran, dass in allen Ländern dieser Welt jedes noch so kleine Auto mit einem eindeutigen und registrierten Kennzeichen versehen ist (Abb. 9) und selbst Bücher eine *International Standard Book Number* (ISBN) haben. Mit der Registrierung sollte natürlich eine Publikationspflicht verbunden sein und beim Publizieren müsste das Kennzeichen genannt werden, um Mehrfachpublikationen zu vermeiden bzw. kenntlich zu machen.



Abb. 9: Fideles Beispiel für die verbindliche Vergabe eindeutiger Kennzeichen.

Die Kosten für ein umfassendes Studienregister würden mehrfach kompensiert werden durch Einsparungen dessen, was zurzeit dadurch verschwendet wird, dass Daten unauffindbar sind und Studien

mehrfach durchgeführt werden (Dickersin 2003).

Dass die *a priori* Registrierung der Studien einen entscheidenden Unterschied machen kann, hat Simes (1986, 1987) gezeigt. Die Analyse aller veröffentlichten Studien zur Behandlung fortgeschrittener Ovarialkarzinome kommt zu dem Ergebnis, dass eine Kombinationschemotherapie der Behandlung mit einem einzelnen Wirkstoff überlegen sei. Eine zweite Analyse jedoch, die ausschließlich auf den Ergebnissen von Studien beruht, die bei ihrer Auflage registriert wurden, führte nicht zu diesem Ergebnis. Möglicherweise war hier die im ersten Fall bestehende Unausgewogenheit der Berichterstattung zugunsten positiver Ergebnisse durch die Registrierung reduziert worden.

Es war sicherlich gut gemeint: in einem Editorial (Shields 2000) kündigte die Zeitschrift *Cancer Epidemiology, Biomarkers & Prevention* an, dass sie eine Rubrik für negative Ergebnisse einrichten werde. Dies sei ein Beitrag zur Vermeidung von *publication bias*. Die Arbeiten müssten sich jedoch auf eine einzige Druckseite beschränken und es werde eine vollständige und nachvollziehbare Darstellung der geprüften Hypothese, der Methodik – einschließlich Biometrie – sowie der erzielten Ergebnisse erwartet. Dieses Feigenblatt erscheint so wirkungsvoll wie eine Gleichstellungsmaßnahme in einem patriarchalischen Land, der zufolge in Zukunft auch weibliche Abgeordnete gewählt werden dürften, sofern sie besonders gut aussähen und sich mit einem Stehplatz in der hintersten Ecke des Plenarsaals begnügten.

## 8. Quergedacht

Die bisherigen Darstellungen gaben in erster Linie die Sicht der jeweiligen Autoren wieder, die in der Literatur zum Thema „unausgewogene Berichterstattung in der Wissenschaft“ berichtet haben. Unsere eigenen Interpretationen haben wir dabei zunächst weitgehend zurückgehalten. Nicht immer konnten wir uns den Argumenten der Autoren anschließen. Hier und da kamen wir zu gänzlich widersprechenden Überzeugungen. Im Folgenden wollen wir einige eigene Gedanken zum Thema *publication bias* vorstellen.

### 8.1. Willkürliche und inkonsistente Definition des Begriffs *publication bias*

Die Begriffe positives Ergebnis, negatives Ergebnis und Null-Ergebnis werden in der Literatur nicht einheitlich gehandhabt. Drei häufige Definitionen einer positiven Studie sind: 1.) das Ergebnis der experimentellen Behandlung ist signifikant besser als die Kontrollbehandlung; 2.) das Ergebnis der Studie ist signifikant, egal welche Behandlung besser ist; 3.) Das Ergebnis der Untersuchung ist interessant oder wird vom Autor als bemerkenswert eingeschätzt. Zwei häufige Definitionen einer negativen Studie sind: 1.) das Ergebnis der experimentellen Behandlung ist nicht signifikant besser oder 2.) die Ergebnisse der Behandlungsarme unterscheiden sich nicht signifikant. Je nach dem, welche Definition man

benutzt, werden Studien als positiv oder als negativ angesehen. Dementsprechend kann man zu unterschiedlicher Einschätzung kommen, ob in einer Sammlung von Studien *publication bias* besteht oder nicht.

Ein Beispiel ist die weiter oben erwähnte Studie von Easterbrook *et al.* (1991), die 720 Forschungsprojekte, die vom *Central Oxford Research Ethics Committee* zwischen 1984 und 1987 genehmigt worden waren, untersucht haben. Sie finden im Hinblick auf die Wahrscheinlichkeit, publiziert zu werden, eine Bevorzugung von Studien mit statistisch signifikanten Ergebnissen. Diese Aussage wird häufig zitiert, ist aber mit äußerster Vorsicht zu betrachten, da die Arbeit in einigen Punkten sehr fragwürdig ist. Erstens: Die Einteilung der Studien macht keinen Sinn: „*statistically significant*“ ist demzufolge ein Ergebnis mit  $p < 0,05$ ; „*non-significant trend*“ geht einher mit  $p \geq 0,05$ ; und ein „*null*“-Ergebnis wurde definiert als „*no difference was observed*“. Damit nicht genug. Selbst Arbeiten, in denen kein Signifikanztest durchgeführt wurde oder werden konnte, z.B. mangels einer Vergleichsgruppe, wurden auf nicht nachvollziehbare Weise in dieses Schema eingeordnet, indem die Wichtigkeit der Ergebnisse eingeschätzt oder auf „*striking observations*“ geachtet wurde.

Bei Ioannidis (1998) wurde eine Studie als „positiv“ gewertet, wenn das Ergebnis des primären Endpunkts zugunsten des experimentellen Studienarms statistisch signifikant war. Studien ohne signifikantes Ergebnis oder mit signifikantem Ergebnis zugunsten des Kontrollarms wurden als „negativ“ gewertet. Gab es keinen Kontrollarm, beispielsweise beim Vergleich zweier

Dosierungen desselben Medikaments, wurde die Studie als „positiv“ gewertet, wenn sich eine der Behandlungsdosen als statistisch signifikant herausstellte. War kein primärer Endpunkt definiert, so wurde die Studie als „positiv“ gewertet, wenn einer der Hauptendpunkte statistisch signifikant war.

Auch Dickersin *et al.* (1992) verwenden eine recht flexible Definition von „signifikant“. Zu den statistisch signifikanten Studien zählten sie diejenigen hinzu, von denen die jeweiligen Autoren meinten, dass sie wichtige Befunde enthielten.

## **8.2. Kritische Bewertung der Trichtergrafik als Instrument zur Erkennung von *publication bias***

Trichtergrafiken sind nicht eindeutig interpretierbar. Neue Verfahren, Therapien und Medikamente werden mit Vorliebe an besonders kranken Personen ausprobiert. Die zugrunde liegende Idee dabei ist: Wenn die Studien bei diesen Patienten keinen messbaren Effekt zeigen, dann wird sie an weniger kranken erst recht keinen Effekt zeigen können. Wenn die Therapie wirksam ist, werden kleine schnelle Studien große Effekte zeigen. Dadurch ermutigt werden spätere größere Studien an weniger kranker Klientel durchgeführt, die in dieser Klientel entsprechend kleinere Erfolge zeigen. Es entsteht dann ein unsymmetrischer Trichter, auch ganz ohne *publication bias*.

Eine weitere Variante ist ein hohler Trichter. Dieser kann entstehen, wenn Studien mit statistisch signifikanten Ergebnissen,

egal in welcher Richtung, systematisch häufiger publiziert werden als Studien, die keinen oder nur geringen Effekt zeigten.

Diese Beispiele zeigen, dass die Trichtergrafik kein verlässliches Werkzeug zum Auffinden oder Ausschließen von *publication bias* ist. Asymmetrische Trichter können andere Ursachen haben als *publication bias* und wiederum kann ein Trichter symmetrisch sein, obwohl *publication bias* vorliegt.

### 8.3. Unausgewogene Berichterstattung über unausgewogene Berichterstattung?

Bei der Ausarbeitung dieser Broschüre fiel uns auf, dass wir Berichten, die keinen oder nur einen geringen *publication bias* nachwiesen, weniger Beachtung schenkten als solchen, die einen statistisch signifikanten Effekt gefunden hatten. Unsere Aufmerksamkeit war positiv-lastig. So kamen wir auf die Idee zu prüfen, ob die Arbeiten zum Thema *publication bias* nicht ebenfalls dem *publication bias* unterliegen.

Hierzu haben wir alle Arbeiten, die wir zum Thema *publication bias* gemustert haben und die eigene Daten enthielten mit Hilfe einer Trichtergrafik (*funnel plot*) ausgewertet (siehe Abbildung; Details s. Dubben und Beck-Bornholdt 2005). Zu den so gefundenen 23 Datenpunkten (Bardy 1998, Callahan 1998, Daniels 2002, De Bellefeuille 1992, Dickersin 1987, Dickersin 1992, Dickersin 1993, Easterbrook 1991, Egger 1997, Evers 2000, Ioannidis 1998, Klassen 2002, Koren 1989, Krzyzanowska 2003, Lexchin 2003, Melander 2003, Misakian

1998, Olson 2002, Scherer 1994, Stern & Simes 1997, Timmer 2002, Weber 1998, Zimpel & Windeler 2000) kamen während des Review-Prozesses auf Vorschlag einer der Reviewer drei weitere hinzu (Chalmers 1990, Cheng 1998, Petticrew 1999).

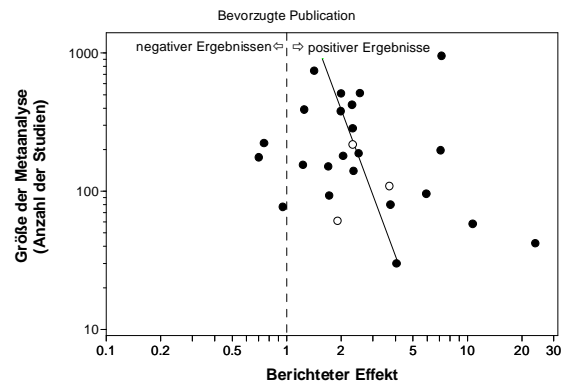


Abb. 10: Trichtergrafik der veröffentlichten Ergebnisse zum Thema „*publication bias*“.

Je nach Untersuchungstyp wurde als Maß für die Größe des *publication bias* entweder die *odds ratio*<sup>15</sup> oder die *hazard ratio*<sup>16</sup> berechnet. Die Trichtergrafik zeigt eine Tendenz dahingehend, dass Arbeiten mit

<sup>15</sup> *odds ratio*: In einer Patienten Gruppe sind 40 von 50 geheilt. 10 sind also nicht geheilt. Die Odd beträgt dann 40:10. Das ist dasselbe wie das Torverhältnis z.B. bei einem Handballspiel, bereitet aber trotzdem vielen Schwierigkeiten. Jetzt kommt noch eine zweite Patientengruppe hinzu, bei der die Odd 30:20 beträgt. Als Maß dafür, in welcher der beiden Gruppen eine bessere Chance auf Heilung besteht, kann man die Odds-ratio (OR) benutzen. Sie ist einfach der Quotient aus den beiden Odds, in diesem Fall ist  $OR = 40:10 / 30:20 = 4 / 1,5 = 2,67$ .

<sup>16</sup> *hazard ratio*: Die Hazard-ratio ist für eine Fußnote schlicht zu kompliziert. Sehr eng verwandt mit ihr ist das relative Risiko, das hier gut her passt. Wir haben dieselben Patientengruppen wie bei der odds-ratio. Das Risiko (das „Risiko“ geheilt zu werden) beträgt  $40/50 = 0,8$  bzw.  $30/50 = 0,6$ . Als Maß dafür, in welcher der beiden Gruppen eine höhere Wahrscheinlichkeit für Heilung besteht, nimmt man hier das relative Risiko (RR, auch *risk-ratio* genannt). Es ist der Quotient aus den beiden Risiken, in diesem Fall ist  $RR = 0,8/0,6 = 1,33$ .

Das relative Risiko ergibt eine kleinere Zahl als die odds-ratio. Die in der Abbildung 10 als hazard ratio markierten Punkte würden also zur odds-ratio umgerechnet noch weiter rechts liegen.

kleinem Stichprobenumfang, d.h. mit wenigen untersuchten Studien, über besonders große positive Effekte berichteten. Diese Tendenz ist nicht statistisch signifikant. Allerdings ist die Power unserer Untersuchung aufgrund der relativ geringen Anzahl der Studien zum Thema *publication bias* eher klein.

## 9. Nachwort

Wir danken Frau Dr. Hanna Kaduszkiewicz für die Durchsicht des Manuskripts. Für verbliebene Fehler und Unverständliches sind wir natürlich selbst verantwortlich. Hinweise, Fragen, Ergänzungen und Vorschläge sind sehr willkommen. Die Adresse dafür lautet: [dubben@uke.uni-hamburg.de](mailto:dubben@uke.uni-hamburg.de).

## 10. Literatur

Die dieser Arbeit zugrunde liegende Literatur wurde in der Medline-Datenbank gesucht mit dem Suchprofil („publication bias“ oder „citation bias“ oder „language bias“ oder „location bias“ oder „reference bias“ oder „multiple publication bias. Die Literaturverzeichnisse der so gefundenen Arbeiten wurden ebenfalls durchsucht.

Bardy AH: Report bias in drug research. *Therapie* 1996, 51: 382-3.

Bardy AH: Bias in reporting clinical trials. *Br-J-Clin-Pharmacol.* 1998, 46(2): 147-50.

Beck-Bornholdt HP, Dubben HH: *Der Schein der Weisen. Irrtümer und Fehlurteile im täglichen Denken.* ISBN 3-499-61450-2, Rowohlt Verlag, 2003.

Beck-Bornholdt HP, Dubben HH: *Der Hund, der Eier legt. Erkennen von Fehlinformationen durch Querdenken.* Buchtitel. Überarbeitete und

aktualisierte Neuauflage. ISBN 3-499-61154 6, Rowohlt Verlag, 2001.

Callaham ML, Wears RL, Weber EJ, Barton C, Young G: Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA* 1998, 280: 254-7.

Chalmers I: Underreporting research is scientific misconduct. *JAMA* 1990, 263: 1405-8.

Chalmers I, Adams M, Dickersin K, Hetherington J, Tarnow-Mordi W, Meinert C, Tonascia S, Chalmers TC. A cohort study of summary reports on controlled trials. *JAMA* 1990; 263: 1401-5.

Cheng K, Preston C, Ashby D, O'Hea U, Smyth RL. Time to publication as full reports of abstracts of randomized controlled trials in cystic fibrosis. *Ped Pulmon* 1998; 26: 101-5.

Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, Godwin J, Qizilbash N, Taylor JO, Hennekens CH: Blood pressure, stroke and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990, 335: 827-838.

Collins R, MacMahon S: Blood pressure, antihypertensive drug treatment and the risks of stroke and of coronary heart disease. *Brit. Med. Bull.* 1994, 50: 272-298.

Daniels CE, Montori VM, Dupras DM: Effect of publication bias on retrieval bias. *Acad-Med.* 2002, 77: 266.

De Bellefeuille C, Morrison CA, Tannock IF: The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. *Annals of Oncology* 1992, 3: 187-191.

Denekamp J: Changes in the rate of repopulation during multifraction irradiation of mouse skin. *Br. J. Radiol.* 1973, 46: 381-387.

Dickersin K, Rennie D: Registering clinical trials. *JAMA* 2003, 290:516-523.

Dickersin K: The existence of publication bias and risk factors for its occurrence. *JAMA* 1990, 263: 1385-9.

Dickersin K: How important is publication bias? A synthesis of available data. *AIDS-Educ-Prev*, 1997, 9: 15-21.

Dickersin K, Min YI: NIH clinical trials and publication bias. *Online-J-Curr-Clin-Trials.* 1993 Apr 28; Doc No 50. (Sekundärzitat gefunden bei Dickersin 1997. Laut Auskunft der Staats- und Universitätsbibliothek Hamburg vom 22.3.2004 ist diese Publikation nicht zu beschaffen. Ärztlichen Zentral....).

Dickersin K, Min YI: Publication bias: The problem that won't go away. *Ann N Y Acad Sci* 1993, 703: 135-48.



- Dickersin K, Min YI, Meinert CL: Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992, 267: 374-8.
- Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H: Publication bias and clinical trials. *Controlled Clinical Trials* 1987, 8: 343-353.
- Dubben HH, Beck-Bornholdt HP: Systematic review of publication bias in studies on publication bias. *BMJ* 2005; 331:433-434.
- Dubben HH, Krüll A, Beck-Bornholdt HP: Split-course radiotherapy: where do we stand? *Strahlenther Oncol* 2001, 177: 227-239.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: Publication bias in clinical research. *Lancet* 1991, 337: 867-872.
- Egger M, Smith GD: Meta-analysis: bias in location and selection of studies. *BMJ* 1998, 316:61-66.
- Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G: Language bias in randomised controlled trials published in English and German. *Lancet* 1997; 350: 326-9.
- Evers JLH: Publication bias in reproductive research. *Hum. Reprod* 2000, 15: 2063-6.
- Gilbody SM, Song F: Publication bias and the integrity of psychiatry research. *Psychological Medicine* 2000, 30: 253-8.
- Gøtzsche PC: Multiple publication of reports of drug trials. *Eur. J. Pharmacol.* 1989, 36: 429-32.
- Gregoire G, Derderian F, Le-Lorier J: Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J-Clin-Epidemiol.* 1995, 48: 159-63.
- Hetherington J, Dickersin K, Chalmers I, Meinert C: Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989, 84: 374-380.
- Hojat M, Gonnella JS, Caellegh AS: Impartial judgment by the "gatekeepers" of science: fallibility and accountability in the peer review process. *Adv-Health-Sci-Educ-Theory-Pract* 2003, 8: 75-96.
- Hopewell S, McDonald S: Full publication of trials initially reported as abstracts in the Australian and New Zealand Journal of Medicine 1980-2000. *Intern-Med-J* 2003, 33: 192-4.
- Horton R: Medical editors trial amnesty. *Lancet* 1997, 350: 756.
- Huston P, Moher D: Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996, 347: 1024-6.
- Ioannidis JPA: Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998, 279: 281-286.
- Jull A, Chalmers I, Rodgers A: Clinical trials in NZ: does anybody know what's going on? *N-Z-Med-J.* 2002, 115: U269
- Juni P, Holenstein F, Sterne J, Bartlett C, Egger M: Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int-J-Epidemiol.* 2002, 31: 115-23.
- Klassen TP, Wiebe N, Russell K, Stevens K, Hartling L, Craig WR, Moher D: Abstracts of randomized controlled trials presented at the Society for Pediatric Research Meeting. *Arch. Pediatr. Adolesc. Med.* 2002, 156: 474-9.
- Koren G, Grahman K, Shear H, Einarson T: Bias against the null hypothesis: the reproductive hazards of cocaine. *Lancet* 1989, 2 (8677):1440-2.
- Krzyzanowska MK, Pintilie M, Tannock IF: Factors Associated With Failure to Publish Large Randomized Trials Presented at an Oncology Meeting. *JAMA* 2003, 290: 495-501.
- LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F: Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997, 337: 536 – 42.
- Lexchin J, Bero LA, Djulbegovic B, Clark O: Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003, 326: 1167-77.
- Link AM: US and non-US submissions: an analysis of reviewer bias. *JAMA* 1998, 280: 246-7.
- Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV, Jonas WB: Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997, 350 (9081): 834-43.
- Mahoney M: Publication prejudices: An experimental study of confirmatory bias in peer review system. *Cognitive Therapy and Research* 1977, 1: 161-175. (Sekundärzitat bei Hojat et al. 2003)
- Manheimer E, Anderson D: Survey of public information about ongoing clinical trials funded by industry: evaluation of completeness and accessibility. *BMJ* 2002;325:528-31.
- Manuscript guideline. *Diabetologia* 1984;25:4A. (Sekundärzitat aus Egger & Smith 1998)
- McMillan JI: Rheumatoid arthritis: a double blind study comparing tolmetin sodium with ibuprofen in patients untreated with either drug previously. *Curr Ther Res* 1982, 31: 813-20.
- McMillan JI: Tolmetin sodium vs ibuprofen in rheumatoid arthritis patients previously untreated with either drug: a double blind cross-over study. *Curr Ther Res* 1977, 22: 266-75.

- Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B: Evidence b(i)ased medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003, 326: 1171-3.
- Misakian AL, Bero LA: Publication bias and research on passive smoking – comparison of published and unpublished results. *JAMA* 1998, 280: 250-3.
- Moher D, Fortin P, Jadad A, Juni P, Klassen T, LeLorier J, Liberati A, Linde K, Penna A: Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996, 347(1998): 363-6
- Olson CM, Rennie D, Cook D, Dickersin K, Flanagan A, Hogan JW, Zhu Q, Reiling J, Pace B: Publication bias in editorial decision making. *JAMA* 2002, 287: 2825-8.
- Petticrew M: Diagoras of Melos (500 BC): an early analyst of publication bias. *Lancet* 1998, 352: 1558.
- Petticrew M, Gilbody S, Song F. Lost information? The fate of papers presented at the 40th Society for Social Medicine Conference. *J Epidemiol Community Health* 1999; 53: 442-3.
- Rennie D: Fair conduct and fair reporting of clinical trials. *JAMA* 1999, 282: 1766-8.
- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. *Münch. med. Wschr.* 139: 644-645, 1997.
- Scherer R, Dickersin K, Langenberg P: Full publication of results initially presented in abstracts – a meta-analysis. *JAMA* 1994, 272: 158-62.
- Shields PG: Publication bias is a scientific problem with adverse ethical outcomes: the case for a section for null results. *Cancer Epidemiology, Biomarkers & Prevention* 2000, 9: 771-2.
- Simes RJ: Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986, 4: 1529-1541. (Sekundärzitat aus Chalmers 1990)
- Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987; 6:11-29. (Sekundärzitat aus Egger & Smith 1998 und Chalmers 1990)
- Smith ML: Publication bias and meta-analysis. *Evaluation Educ.* 1980, 4: 22-24. (Sekundärzitat aus Dickersin 1990)
- Song F, Khan KS, Dinnes J, Sutton AJ: Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *Int-J-Epidemiol.* 2002, 31: 88-95.
- Sterling TD: Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *J Am Stat Assoc* 1959, 54: 30-34.
- Stern JM, Simes RJ: Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997, 315: 640-5.
- Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR: Empirical assessment of effect of publication bias on meta-analysis. *Brit. Med. J.* 2000, 320: 1574-7.
- Timmer A, Hilsden RJ, Cole J, Hailey D, Sutherland LR: Publication bias in gastroenterological research – a retrospective cohort study based on abstracts submitted to a scientific meeting. *BMC Medical Research Methodology* 2002, 2:7 and <http://www.biomedcentral.com/1471-2288/2/7>
- Tramèr MR, Reynolds DJM, Moore RA, McQuay HJ: Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997, 315:635-640.
- Villanueva P, Peiró S, Libero J, Pereiró I: Accuracy of pharmaceutical advertisements in medical journals. *Lancet* 2003, 361: 27-32.
- Weber EJ, Callahan ML, Wears RL: Unpublished research from a medical specialty meeting: Why investigators fail to publish. *JAMA* 1998, 280: 257-259.
- Winkmann G., Schlutius S, Schwelm HG: Publikationssprachen der Impact Faktor-zeitschriften und medizinischer Literaturdatenbanken. *Dtsch. Med. Wochenschr.* 2002, 127: 131-7.
- Zimpel T, Windeler J: Veröffentlichungen von Dissertationen zu unkonventionellen medizinischen Therapie- und Diagnoseverfahren – ein Beitrag zum “publication bias”. *Forschende Komplementärmedizin – Klassische Naturheilkunde* 2000, 7: 71-4.