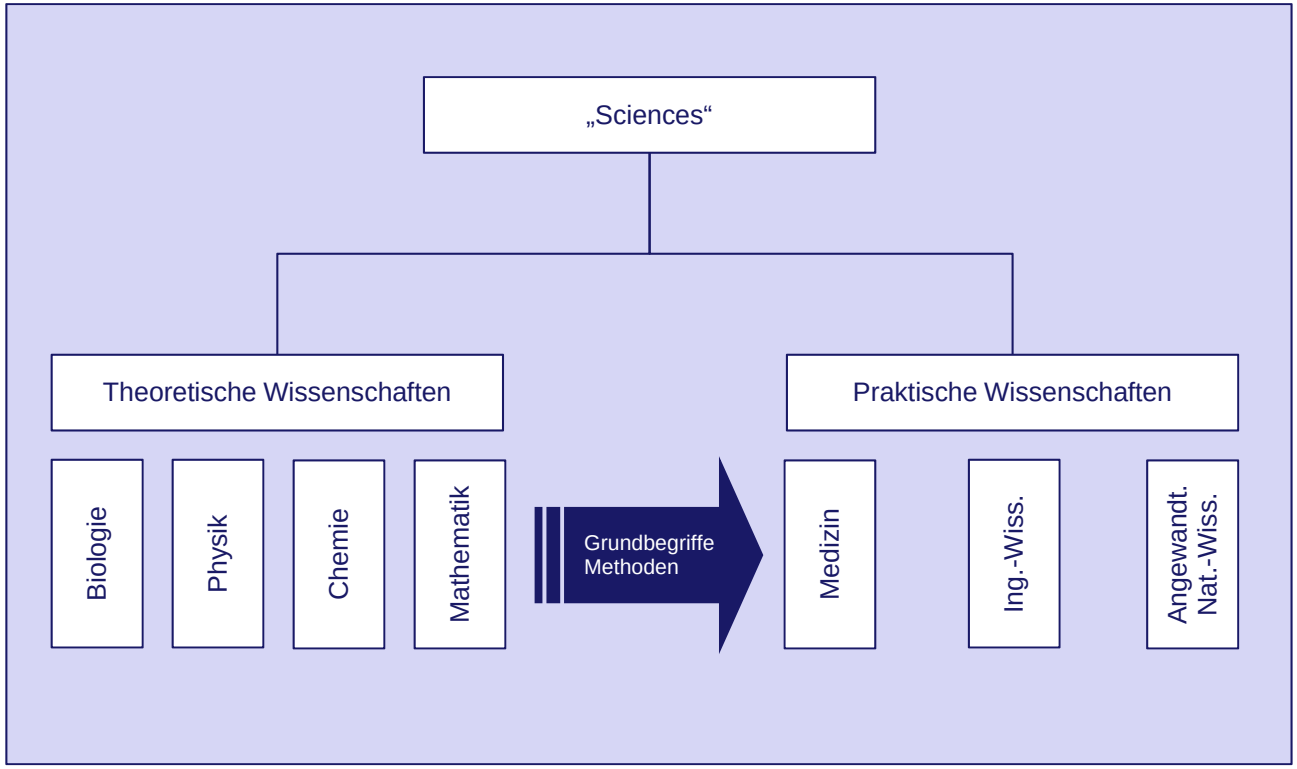
Wissenschaftstheoretische Minimalstandards der Förderwürdigkeit in der biomedizinischen Forschung

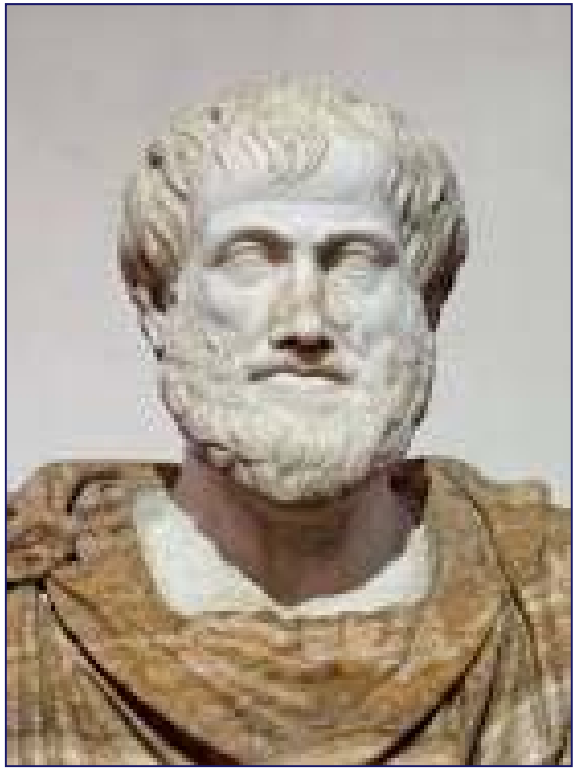
Dr. phil. Martin Langanke, M.A.

Berlin, 24. September 2015

Gliederung

1. Hintergrund – Zur Unterscheidung von praktischem und theoretischem Wissen





Aristoteles
384-322 v. Chr.

1. Hintergrund – Die Medizin als praktische Wissenschaft

Reducing waste …

1 In praktischen Wissenschaften ist Erkenntnis nicht Selbstzweck, sondern Mittel zum Erreichen praktischer Zwecke.

2 In der Medizin ist Heilung von Krankheiten und Linderung von Leiden das Ziel, nicht human-biologische Erkenntnis.

3 Es ist wissenschaftstheoretisch legitim, die Medizin an der Erreichung ihrer praktischen Ziele zu messen.

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2×2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2×2 table, one gets $\text{PPV} = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

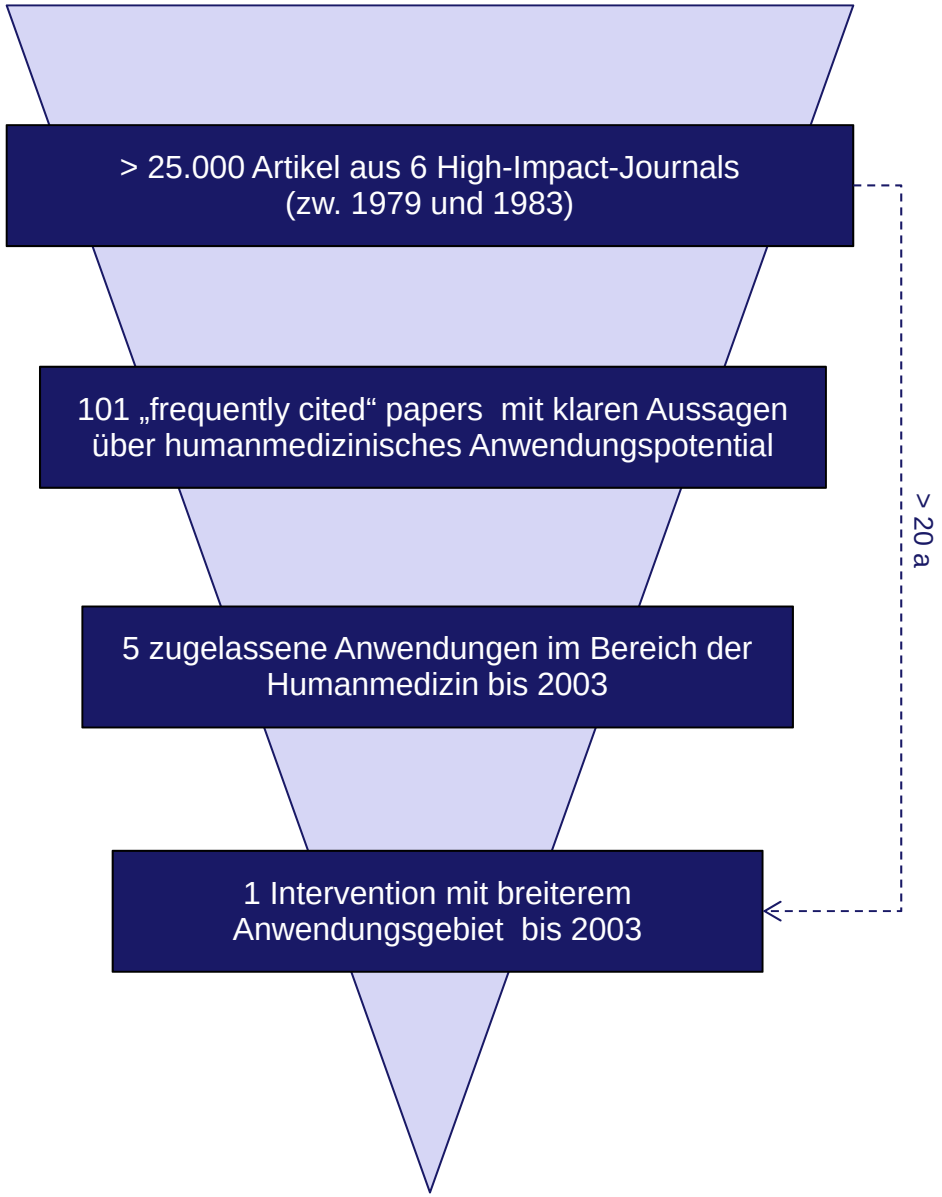
John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

Ioannidis 2005

1. Hintergrund – Translation als Herausforderung in der Medizin



SPECIAL ARTICLES

Translation of Highly Promising Basic Science Research into Clinical Applications

Despina G. Contopoulos-Ioannidis, MD, Evangelia E. Ntzani, MD, John P. A. Ioannidis, MD

PURPOSE: To evaluate the predictors of and time taken for the translation of highly promising basic research into clinical experimentation and use.
METHODS: We identified 101 articles, published between 1979 and 1983 in six major basic science journals, which clearly stated that the technology studied had novel therapeutic or preventive promises. Each case was evaluated for whether the promising finding resulted in relevant randomized controlled trials and clinical use. Main outcomes included the time to published trials, time to published trials with favorable results ("positive" trials), and licensed clinical use.
RESULTS: By October 2002, 27 of the promising technologies had resulted in at least one published randomized trial, 19 of which had led to the publication of at least one positive randomized trial. Five basic science findings are currently licensed for clinical use, but only one has been used extensively for the licensed indications. Promising technologies that did not lead to a published human study within 10 to 12 years were unlikely to be tested in humans subsequently. Some form of industry involvement in the basic science publication was the strongest predictor of clinical experimentation, accelerating the process by about eightfold (95% confidence interval: 3 to 19) when an author had industry affiliations.
CONCLUSION: Even the most promising findings of basic research take a long time to translate into clinical experimentation, and adoption in clinical practice is rare. **Am J Med. 2003; 114:477–484.** ©2003 by Excerpta Medica Inc.

Medical progress is highly dependent on the products of basic research (1), which occasionally lead to discoveries that have clinical promise. However, it is not known how often and how fast original basic research findings translate into clinical development and use, as well as what are the predictors of and obstacles to realization of these findings. To address these questions, we evaluated a sample of basic research publications in highly cited journals that had presented findings showing a clear clinical promise, and studied whether the original expectations materialized over a period of 20 years.

METHODS

Inclusion Criteria

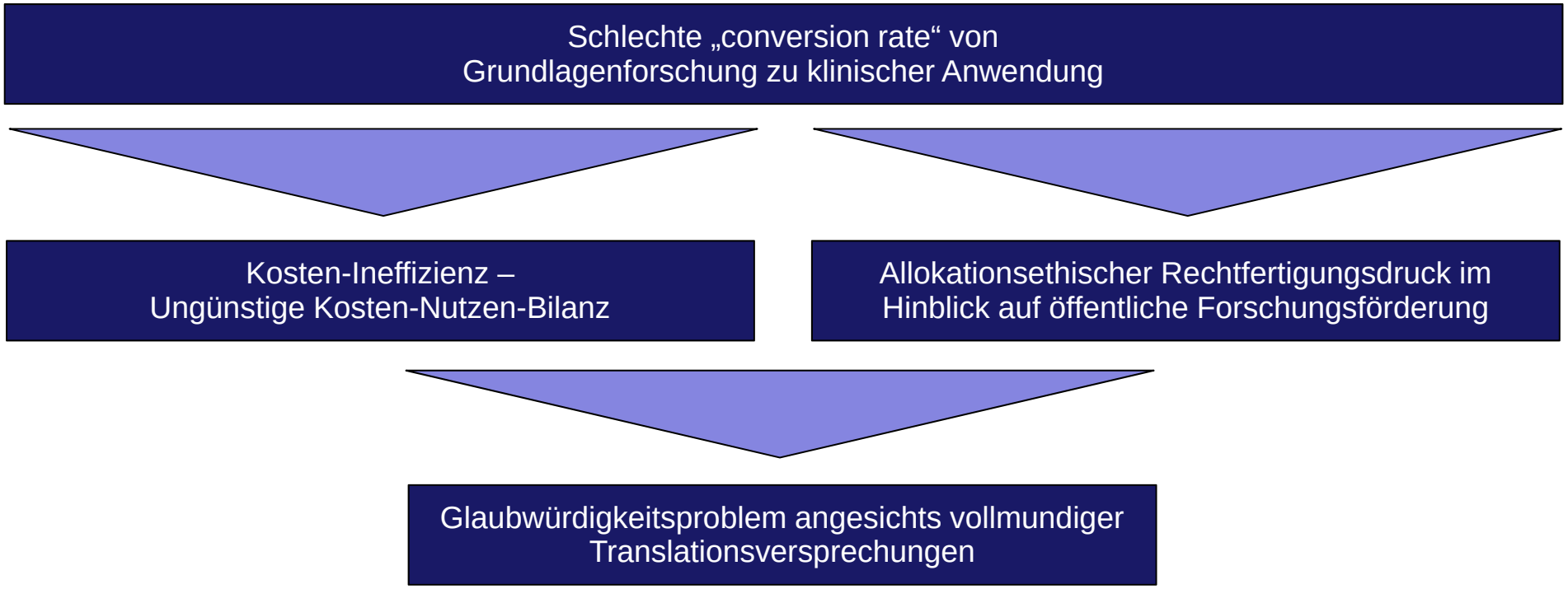
We searched PubMed for articles published from 1979 to 1983 in six highly cited basic science journals: *Science*, *Nature*, *Cell*, the *Journal of Experimental Medicine*, and the *Journal of Clinical Investigation*, which had the highest impact factors in 2000, and the *Journal of Biological Chemistry*, which receives the most citations. We identified all articles that contained the word *therapy*, *therapies*, *therapeutic*, *therapeutical*, *prevention*, *preventive*, *vaccine*, *vaccines*, or *clinical*. From these articles, we retained all original publications that clearly stated that the studied technology might have future clinical therapeutic or preventive application. The 5-year period (1979 to 1983) allowed a meaningful time of approximately 20 years to elapse for examining the translation of basic science research into clinical research and practice. Eligible technologies included substances, antibodies, vaccines, gene therapies, technical devices and other nonpharmacologic interventions, combination therapies, or novel techniques for production of the above technologies. We only considered technologies that were still at an experimental stage (molecular, cellular, animal, and early nonrandomized human studies) that did not have prior application in humans for the specific promise. We also included articles that focused on a novel application (different disease or indication) of a technology already in use in humans or on a novel strategy combining technologies already in use. We excluded articles that did not describe a clear clinical promise in the abstract; editorials; commentaries; reviews; news articles; articles that focused on mechanisms of action, pathophysiology, or diagnosis; and articles on agricultural or veterinary applications. Initial

From the Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology (DCI, EN, JI), Department of Pediatrics (DCI), University of Ioannina School of Medicine, Ioannina, Greece; Biomedical Research Institute (JI), Foundation for Research and Technology-Hellas, Greece; Department of Pediatrics (DCI), George Washington University School of Medicine and Health Sciences, Washington, D.C.; and Division of Clinical Care Research (JI), Tufts–New England Medical Center, Boston, Massachusetts.
Requests for reprints should be addressed to John P. A. Ioannidis, MD, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece, or jioannid@cc.uoi.gr.
Manuscript submitted August 31, 2002, and accepted in revised form November 12, 2002.

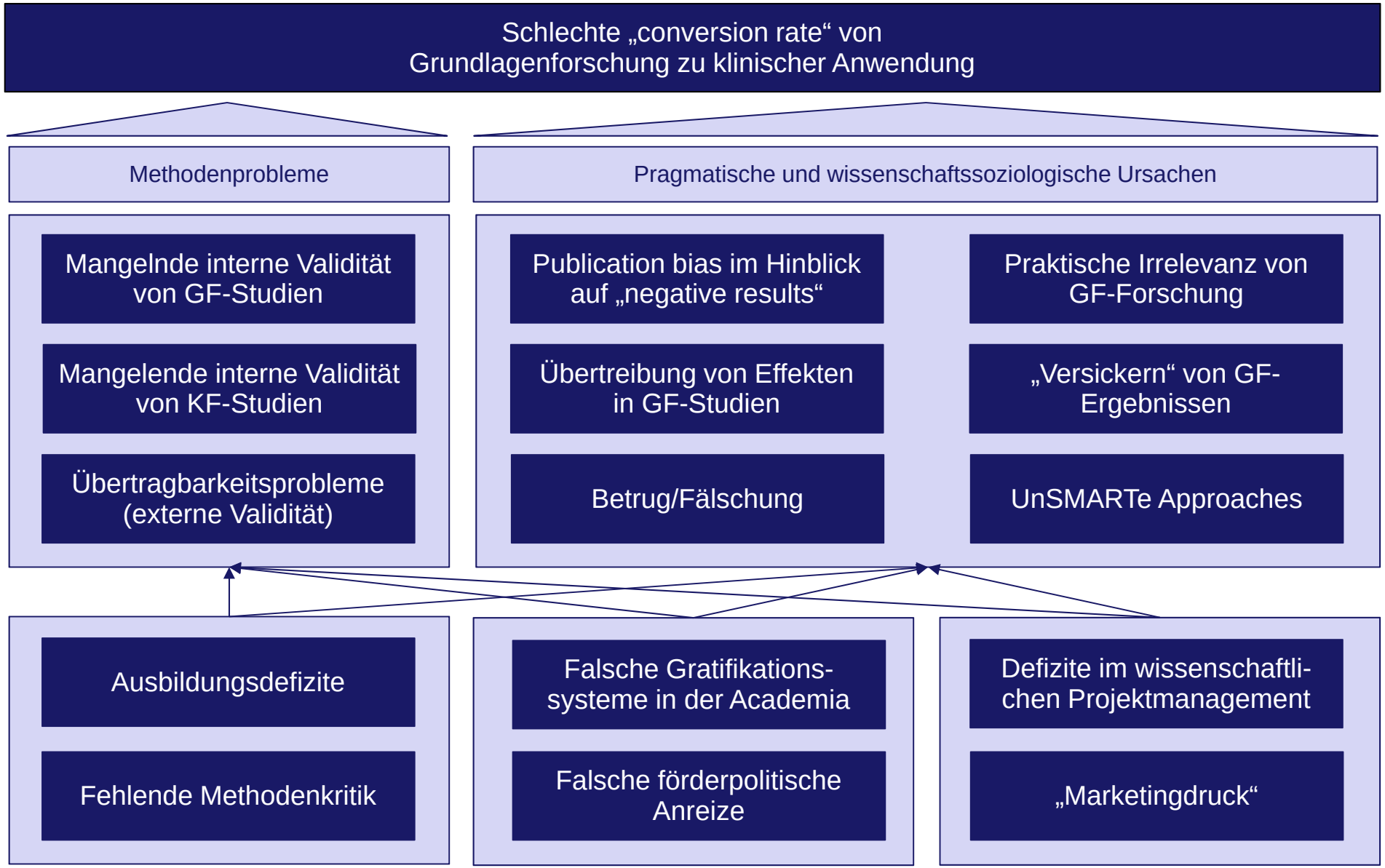
0002-9343/03/$–see front matter 477
doi:10.1016/S0002-9343(03)00013-5

Contopoulos-Ioannidis et al. 2003

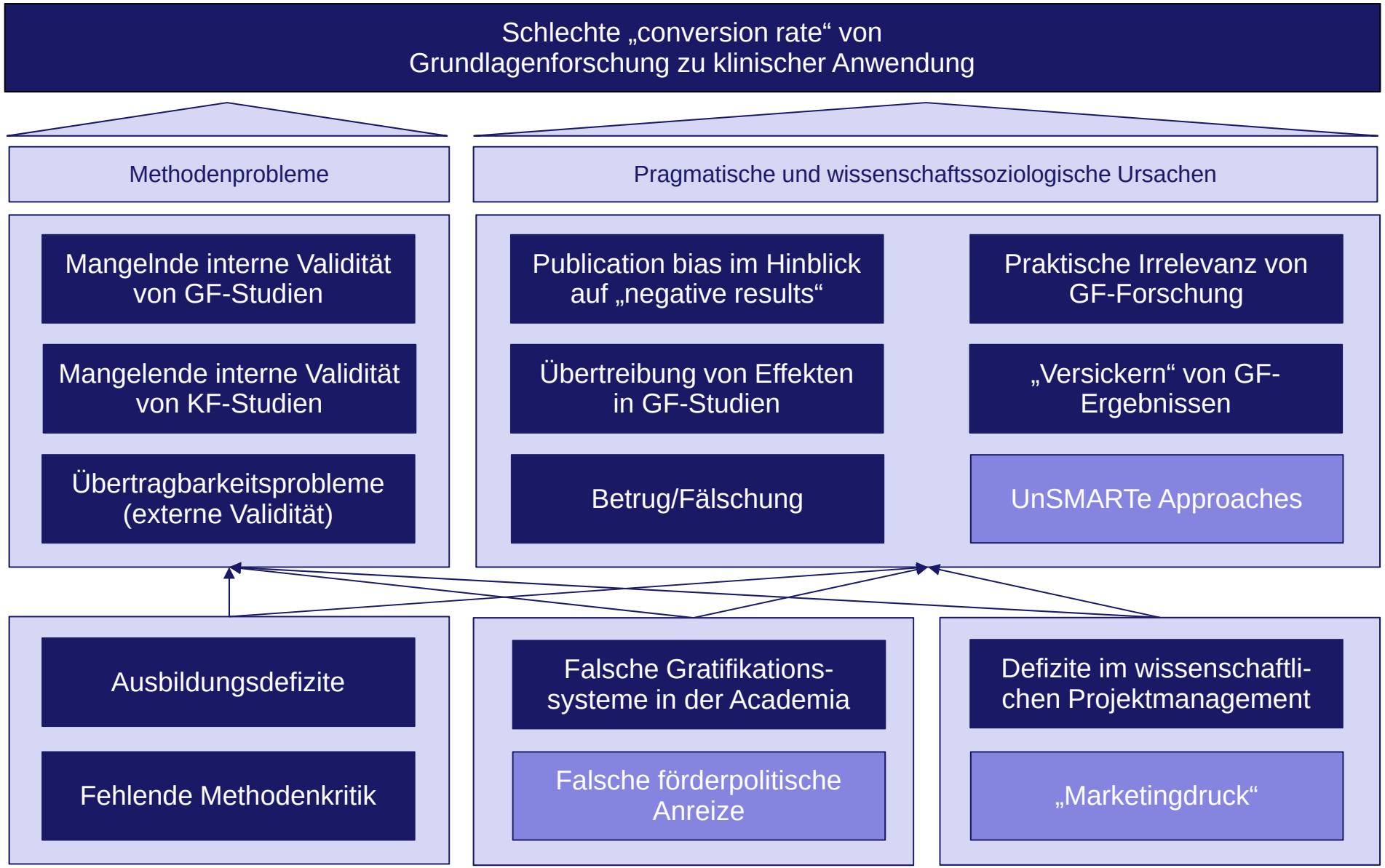
1. Hintergrund – Normative Aspekte der Conversionsproblematik



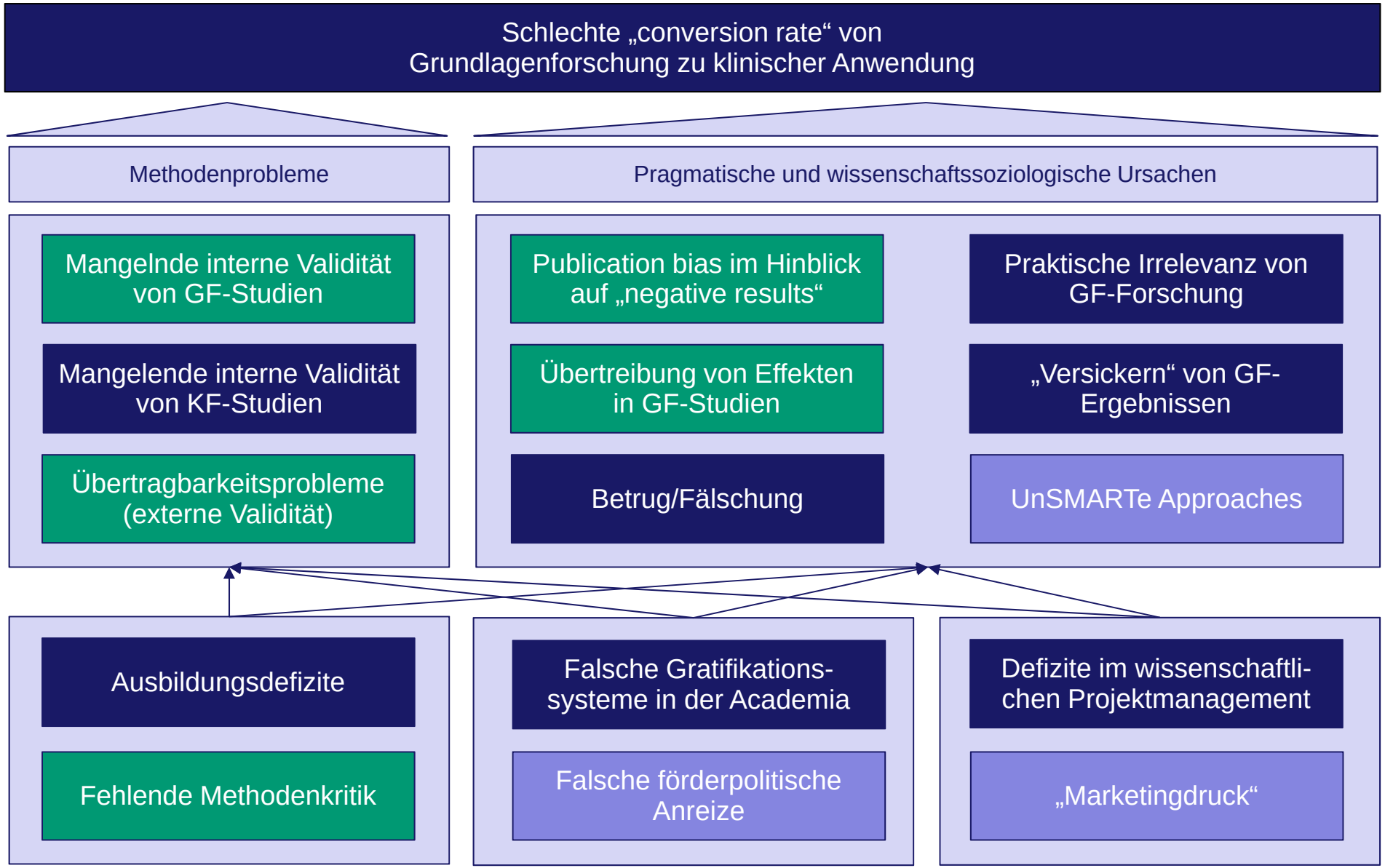
1. Hintergrund – Ursachen der Conversionsproblematik (Auswahl)



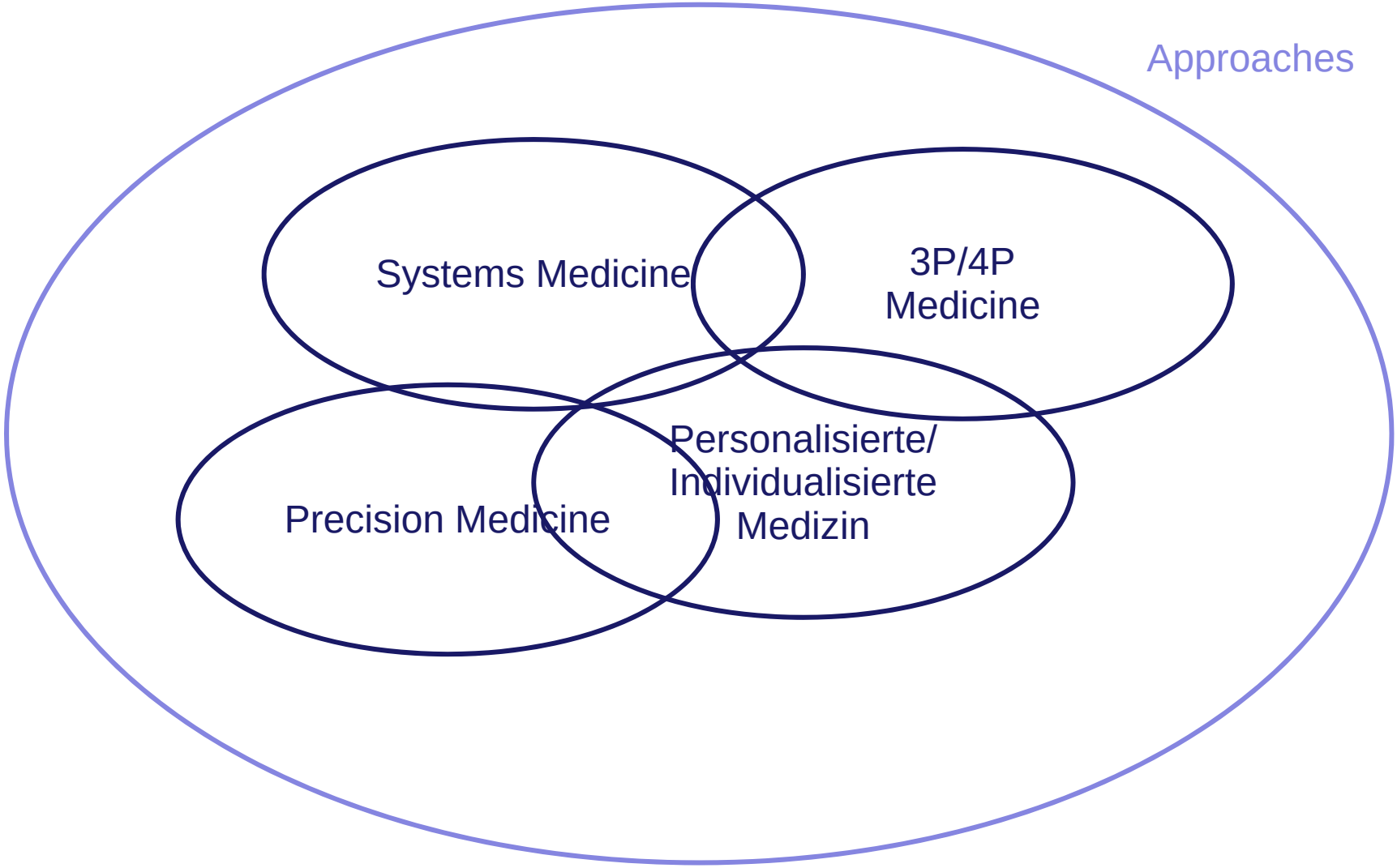
1. Hintergrund – Ursachen der Conversionsproblematik (Auswahl)



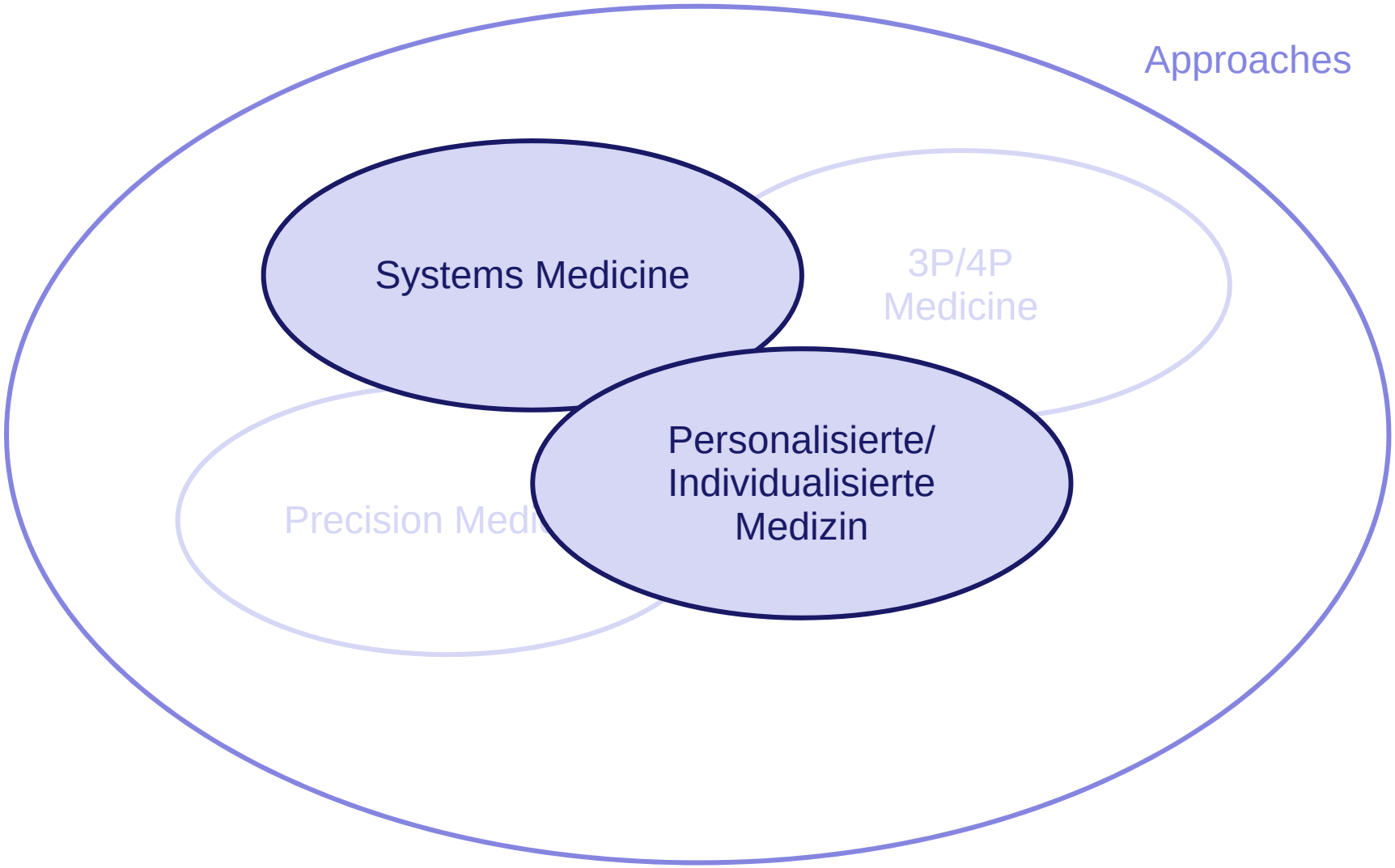
1. Hintergrund – Ursachen der Conversionsproblematik (Auswahl)



2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

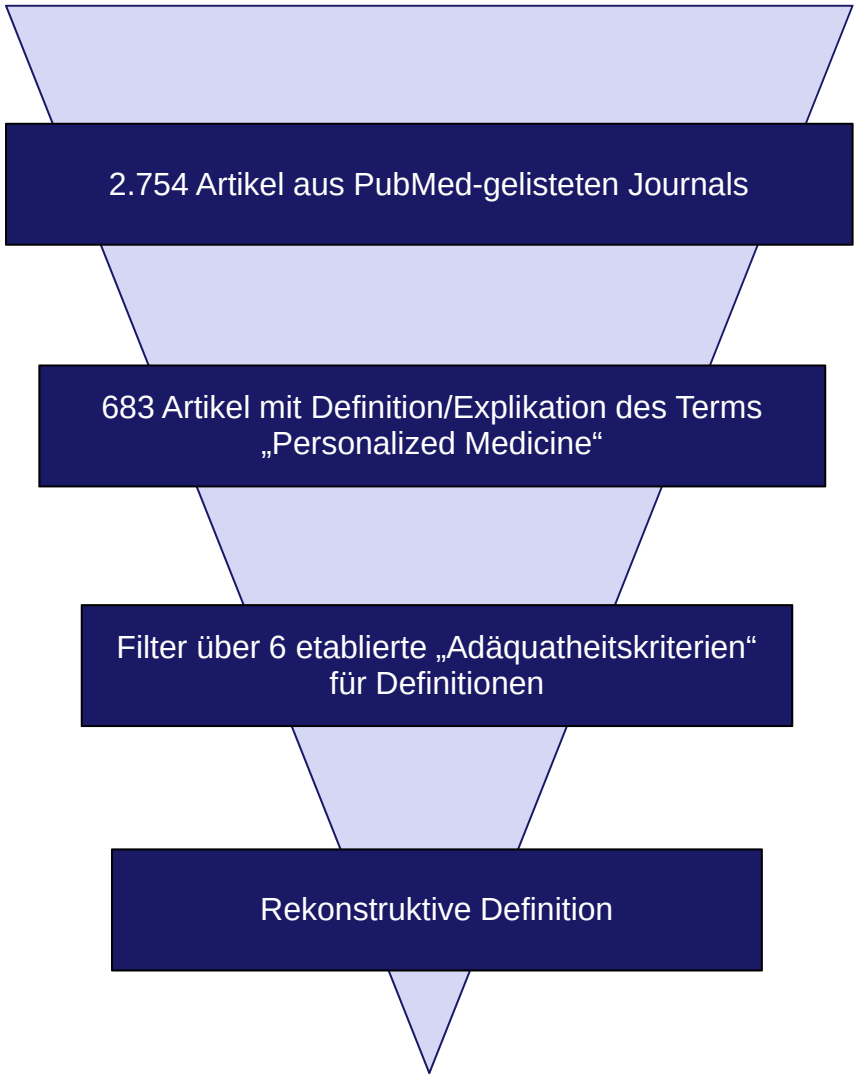


2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung



2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Was ist Personalisierte/Individualisierte Medizin?



Schleidgen *et al. BMC Medical Ethics* 2013, **14**:55
http://www.biomedcentral.com/1472-6939/14/55

BMC Medical Ethics

RESEARCH ARTICLE Open Access

What is personalized medicine: sharpening a vague term based on a systematic literature review

Sebastian Schleidgen[1*], Corinna Klingler[1], Teresa Bertram[1], Wolf H Rogowski[2,3] and Georg Marckmann[1]

Abstract

Background: Recently, individualized or personalized medicine (PM) has become a buzz word in the academic as well as public debate surrounding health care. However, PM lacks a clear definition and is open to interpretation. This conceptual vagueness complicates public discourse on chances, risks and limits of PM. Furthermore, stakeholders might use it to further their respective interests and preferences. For these reasons it is important to have a shared understanding of PM. In this paper, we present a sufficiently precise as well as adequate definition of PM with the potential of wide acceptance.

Methods: For this purpose, in a first step a systematic literature review was conducted to understand how PM is actually used in scientific practice. PubMed was searched using the keywords "individualized medicine", "individualised medicine", "personalized medicine" and "personalised medicine" connected by the Boolean operator OR. A data extraction tabloid was developed putting forward a means/ends-division. Full-texts of articles containing the search terms in title or abstract were screened for definitions. Definitions were extracted; according to the means/ends distinction their elements were assigned to the corresponding category. To reduce complexity of the resulting list, summary categories were developed inductively from the data using thematic analysis. In a second step, six well-known criteria for adequate definitions were applied to these categories to derive a so-called precising definition.

Results: We identified 2457 articles containing the terms PM in title or abstract. Of those 683 contained a definition of PM and were thus included in our review. 1459 ends and 1025 means were found in the definitions. From these we derived the precising definition: PM seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics.

Conclusions: Our definition includes the aspects that are specific for developments labeled as PM while, on the other hand, recognizing the limits of these developments. Furthermore, it is supported by the quantitative analysis of PM definitions in the literature, which suggests that it it is widely acceptable and thus has the potential to avoid the above mentioned issues.

Keywords: Biomarkers, Conceptual vagueness, Definition, Individualized medicine, Stratification, Timing

Background

In recent years, individualized or personalized medicine (IM/PM)[4] has become a buzz word in the academic as well as public debate surrounding health care. Promising to make health care more effective and efficient by tailored medical interventions it has become one of the core areas of public research funding and pharmaceutical research investment [1]. However, PM lacks a clear definition and is open to interpretation [2]. Consequently, a whole continuum of PM understandings exists, in which three main positions can be identified: (a) PM is not a new concept as medicine has always been individualized, (b) PM is holistic health care centered around the needs of the individual patient and (c) PM is treatment targeted at stratified subgroups (e.g. pharmacogenetics) [3].

The prevailing vagueness of the term poses several problems. First and foremost, it unduly complicates public discourse on chances, risks and limits of PM: if the meaning of a term like PM is not clearly defined, it is trivially impossible to debate questions of its matter as well as its (future) handling. As a consequence, it is

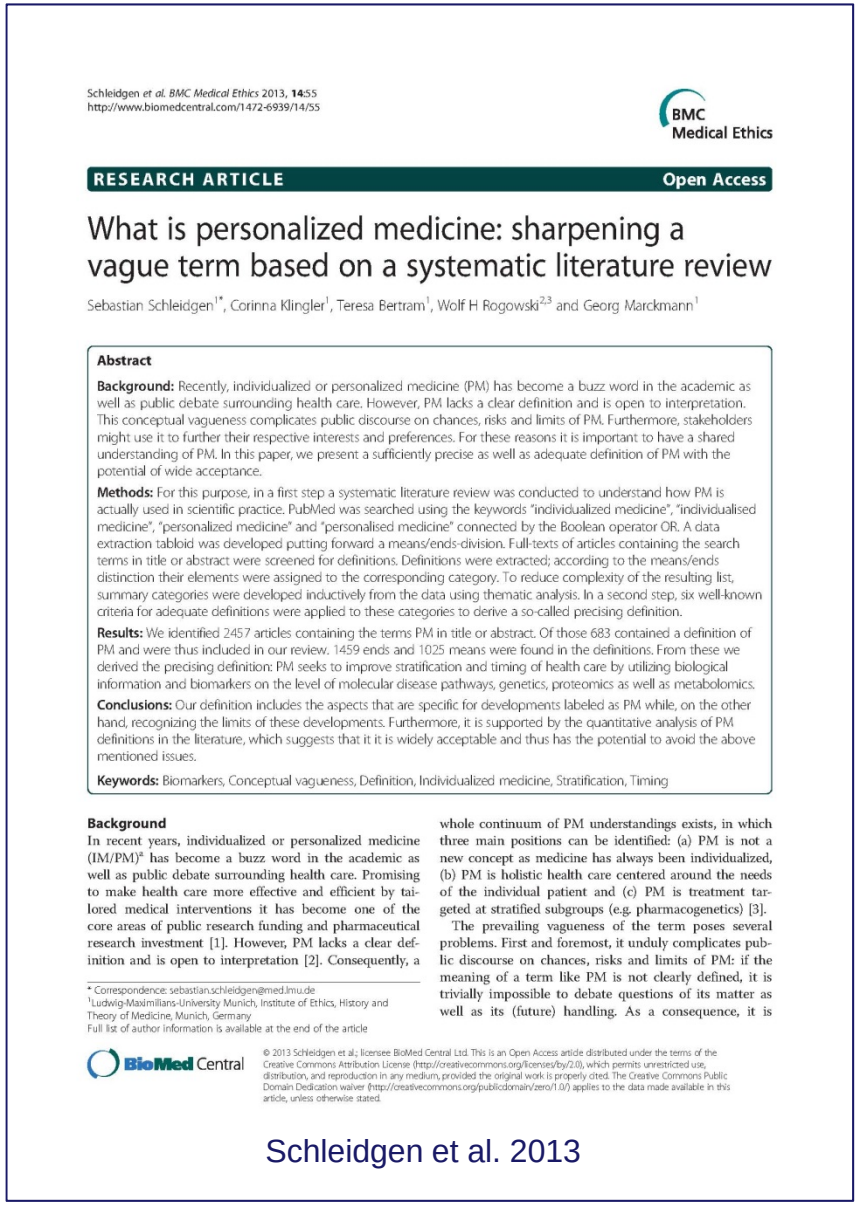
* Correspondence: sebastian.schleidgen@med.lmu.de
[1]Ludwig-Maximilians-University Munich, Institute of Ethics, History and Theory of Medicine, Munich, Germany
Full list of author information is available at the end of the article

Schleidgen et al. 2013

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Was ist Personalisierte/Individualisierte Medizin?

“PM seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics.”



Schleidgen *et al. BMC Medical Ethics* 2013, **14**:55
http://www.biomedcentral.com/1472-6939/14/55

BMC Medical Ethics

RESEARCH ARTICLE Open Access

What is personalized medicine: sharpening a vague term based on a systematic literature review

Sebastian Schleidgen[1*], Corinna Klingler[1], Teresa Bertram[1], Wolf H Rogowski[2,3] and Georg Marckmann[1]

Abstract

Background: Recently, individualized or personalized medicine (PM) has become a buzz word in the academic as well as public debate surrounding health care. However, PM lacks a clear definition and is open to interpretation. This conceptual vagueness complicates public discourse on chances, risks and limits of PM. Furthermore, stakeholders might use it to further their respective interests and preferences. For these reasons it is important to have a shared understanding of PM. In this paper, we present a sufficiently precise as well as adequate definition of PM with the potential of wide acceptance.

Methods: For this purpose, in a first step a systematic literature review was conducted to understand how PM is actually used in scientific practice. PubMed was searched using the keywords “individualized medicine”, “individualised medicine”, “personalized medicine” and “personalised medicine” connected by the Boolean operator OR. A data extraction tabloid was developed putting forward a means/ends-division. Full-texts of articles containing the search terms in title or abstract were screened for definitions. Definitions were extracted; according to the means/ends distinction their elements were assigned to the corresponding category. To reduce complexity of the resulting list, summary categories were developed inductively from the data using thematic analysis. In a second step, six well-known criteria for adequate definitions were applied to these categories to derive a so-called precising definition.

Results: We identified 2457 articles containing the terms PM in title or abstract. Of those 683 contained a definition of PM and were thus included in our review. 1459 ends and 1025 means were found in the definitions. From these we derived the precising definition: PM seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics.

Conclusions: Our definition includes the aspects that are specific for developments labeled as PM while, on the other hand, recognizing the limits of these developments. Furthermore, it is supported by the quantitative analysis of PM definitions in the literature, which suggests that it it is widely acceptable and thus has the potential to avoid the above mentioned issues.

Keywords: Biomarkers, Conceptual vagueness, Definition, Individualized medicine, Stratification, Timing

Background

In recent years, individualized or personalized medicine (IM/PM)[4] has become a buzz word in the academic as well as public debate surrounding health care. Promising to make health care more effective and efficient by tailored medical interventions it has become one of the core areas of public research funding and pharmaceutical research investment [1]. However, PM lacks a clear definition and is open to interpretation [2]. Consequently, a whole continuum of PM understandings exists, in which three main positions can be identified: (a) PM is not a new concept as medicine has always been individualized, (b) PM is holistic health care centered around the needs of the individual patient and (c) PM is treatment targeted at stratified subgroups (e.g. pharmacogenetics) [3].

The prevailing vagueness of the term poses several problems. First and foremost, it unduly complicates public discourse on chances, risks and limits of PM: if the meaning of a term like PM is not clearly defined, it is trivially impossible to debate questions of its matter as well as its (future) handling. As a consequence, it is

* Correspondence: sebastian.schleidgen@med.lmu.de
[1]Ludwig-Maximilians-University Munich, Institute of Ethics, History and Theory of Medicine, Munich, Germany
Full list of author information is available at the end of the article

Schleidgen et al. 2013

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

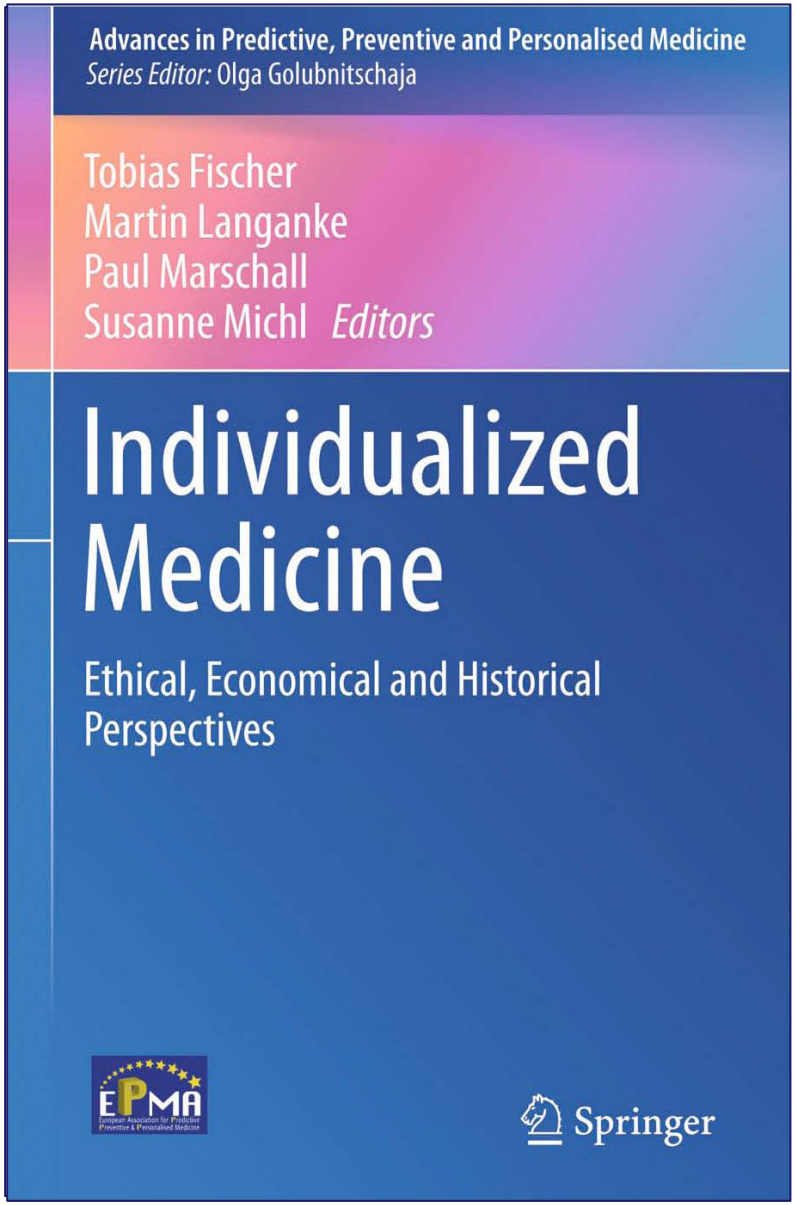
Was ist Personalisierte/Individualisierte Medizin?

Zweckreflexion

Logisch-semantisch und pragmatisch angeleitete Analyse ausgewählter Definitionen/Explikationen

Logische und semantische Rekonstruktion

Stipulativ-rekonstruktive Definition



2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Was ist Personalisierte/Individualisierte Medizin?

“Individualized Medicine” (“IM” in short) stands for current medical fields of research on the one hand, and for health care practices on the other hand.
Medical fields of research can be subsumed under the term Individualized Medicine if they aim at identifying, validating and integrating biomarkers into the clinical routine which allow to predict the outbreak or the course of diseases and/or the effect of therapies or their unwanted effects for certain patient groups in a better way.
Preventive, therapeutic or rehabilitative health care practices, which can be included in the term of Individualized Medicine, are characterized by the fact that they use biomarkers for a systematic prediction of risks or courses of diseases and/or for the prediction of the effect of therapies or their unwanted effects.”

Chapter 2
The Meaning of “Individualized Medicine”: A Terminological Adjustment of a Perplexing Term

Martin Langanke, Wolfgang Lieb, Pia Erdmann, Marcus Dörr, Tobias Fischer, Heyo K. Kroemer, Steffen Flessa and Heinrich Assel

Abstract This chapter introduces “Individualized Medicine” as a technical term. In order to do this the chapter first gives a precise, logical and conceptual analysis of relevant explanations and definitions from English and German speaking areas. It secondly presents a definition according to which the term “Individualized Medicine” should be used for describing research approaches and health care practices, when the

M. Langanke (✉) · P. Erdmann · H. Assel
Theologische Fakultät, Lehrstuhl für Systematische Theologie, Ernst-Moritz-Arndt-Universität Greifswald, Am Rubenowplatz 2–3, 17487 Greifswald, Germany
e-mail: langanke@uni-greifswald.de

P. Erdmann
e-mail: pia.erdmann@uni-greifswald.de

H. Assel
e-mail: assel@uni-greifswald.de

W. Lieb
Institut für Epidemiologie, Christian-Albrechts-Universität zu Kiel, Campus UKSH, Niemannsweg 11, 24105 Kiel, Germany
e-mail: wolfgang.lieb@epi.uni-kiel.de

M. Dörr
Klinik und Poliklinik für Innere Medizin B, Universitätsmedizin Greifswald, F.-Sauerbruch-Straße, 17475 Greifswald, Germany
e-mail: mdoerr@uni-greifswald.de

T. Fischer
Department für Ethik, Theorie und Geschichte der Lebenswissenschaften, Universitätsmedizin Greifswald, Walther-Rathenau-Str. 48, 17475 Greifswald, Germany
e-mail: tobias.fischer@uni-greifswald.de

H. K. Kroemer
Universitätsmedizin Göttingen, Grete-Henry-Str. 9, 37085 Göttingen, Germany
e-mail: heyo.kroemer@med.uni-goettingen.de

S. Flessa
Rechts- und Staatswissenschaftliche Fakultät, Lehrstuhl für Allgemeine Betriebswirtschaftslehre und Gesundheitsmanagement, Ernst-Moritz-Arndt-Universität Greifswald, Friedrich-Loeffler-Str. 70, 17487 Greifswald, Germany
e-mail: steffen.flessa@uni-greifswald.de

T. Fischer et al. (eds.), *Individualized Medicine*, Advances in Predictive, Preventive and Personalised Medicine 7, DOI 10.1007/978-3-319-11719-5_2

Langanke et al. 2015

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Abacavir	n-o	2008	
Anastrozol		o	1996
Arsentrioxid		o	2002
Azathioprin		n-o	2002
Bosutinib		o	2013
Brentuximab vedotin		o	2012
Carbamazepin		n-o	2012
Cetuximab		o	2008
Crizotinib		o	2012
Dasatinib		o	2006
Erlotinib	o	2011	
Everolimus		o	2009
Exemestan		o	1999
Fulvestrant		o	2004
Gefitinib	o	2009	
Imatinib	o	2001	
Ivacaftor	n-o	2012	
Lapatinib		o	2008
Letrozol	o	1997	
Maraviroc		n-o	2007
Mercaptopurin		o	< 2009
Natalizumab		n-o	2011
Nilotinib	o	2007	
Oxarbazepin		n-o	2012
Patinumumab		o	2007
Pertuzumab		o	2013
Tamoxifen		o	< 2009
Toremifen		o	1996
Trastuzumab		o	2000
Vandetanib		o	2012
Vemurafenib		o	2012

IM/PM-Applikationen sind vor allem Test-Präparat-Anwendungen in der Onkologie

Zugelassene Test-Präparat-Kombinationen in Deutschland

Quelle: http://www.vfa.de/de/download-manager/_individualisierte-medizin.pdf

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Was ist Personalisierte/Individualisierte Medizin?

- Muss ein PM/IM-Marker molekular sein? Was ist mit „mixed scores“?
- Ist PM/IM ein Forschungs- und/oder ein Versorgungsansatz?
- Wurde in der Medizin nicht immer schon stratifiziert?
- Wollen nicht alle medizinischen Ansätze die Krankenversorgung verbessern?
- Ist PM/IM gegenüber anderen „approaches“ scharf abgrenzbar oder nur graduell?

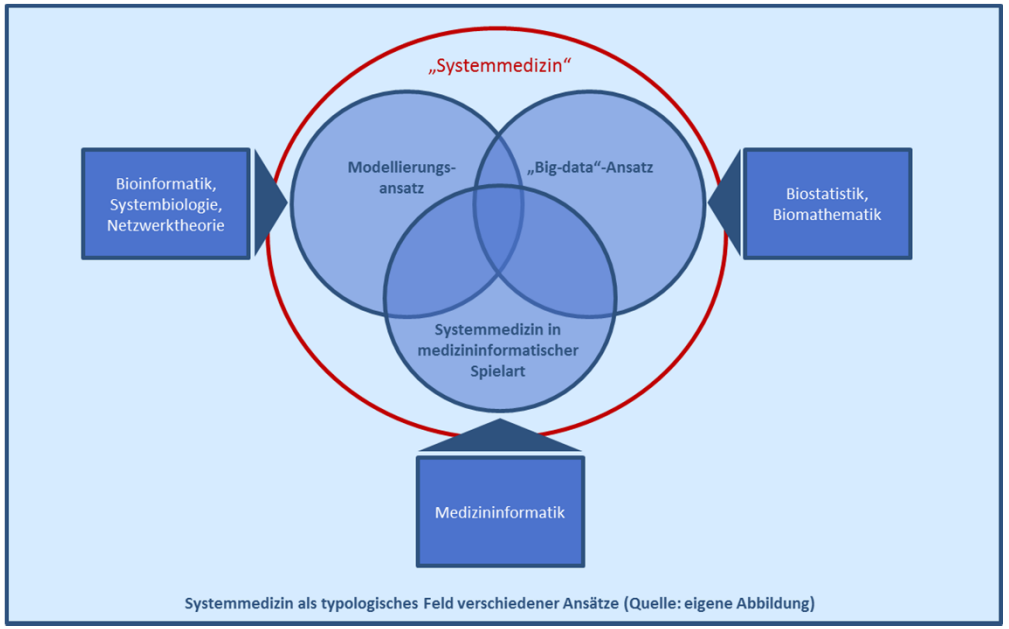
“Eine ökonomische Bewertung der personlisierten oder individualisierten Medizin als Konzept scheiterte regelmäßig an der unscharfen Definition des Terminus. Denn was man nicht definieren kann (oder will), kann man auch nicht abschließend ökonomisch bewerten.”

Erdmann et al. 2015

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Was ist Systemmedizin?

- Big-Data-Ansatz
- Mechanistische bioinformatische Modellierung
- Medizininformatische Ansätze



Systemmedizin als typologisches Feld verschiedener Ansätze (Quelle: eigene Abbildung)

MEDIZINISCHE VERSORGUNG

Systemmedizin: Herausforderungen eines aktuellen Ansatzes

Die Thematisierung ethischer und ökonomischer Herausforderungen, die mit der Translation systemmedizinischer Ansätze in die Versorgung einhergehen, ist keineswegs neu.

Pia Erdmann*, Tobias Fischer*, Susan Raths*, Steffen Fleßa, Martin Langanke

Ein neues Zauberwort prägt die aktuellen Debatten um die Gesundheitsversorgung der Zukunft: „Systemmedizin“. Damit werden zum einen grundlagentheoretische Ansätze bezeichnet, die darauf zielen, die biologischen Mechanismen der Krankheitsentstehung unter Nutzung von Methoden aus „omics“-Forschung, Systembiologie, Informatik und Netzwerktheorie besser zu verstehen. Diese verorten sich zum Teil explizit in der Nachfolge der personalisierten oder individualisierten Medizin (1–6). Zum anderen vermarkten sich auch stärker anwendungsorientierte, „translationale“ Konzepte als Systemmedizin; damit sollen Klinikprozesse unter Nutzung medizininformatischer Tools nachhaltiger und patientenzentrierter gestaltet oder Behandlungsdaten für die wissenschaftliche Auswertung zugänglich gemacht werden (7–9). Allen diesen Ansätzen gemeinsam ist letztlich eine besonders starke Fokussierung auf den Einsatz informationstechnologischer Werkzeuge für medizinische Zwecke. Dabei reicht das Spektrum von hoch anspruchsvollen biomathematischen Modellierungsinstrumenten über hypothesenfreie „Big-Data“-Ansätze bis hin zur IT-Unterstützung von Workflows in Forschung und Klinik.

- Von der Entwicklung bioinformatischer Modelle erhofft man sich, pathologische Prozesse und die dahinter stehenden Mechanismen am Computer simulieren und so gezielt Interventionsoptionen nicht-invasiv auf ihre Effektivität hin prüfen zu können (4, 5). Voraussetzung dafür ist ein kausales Verständnis der jeweils relevanten pathogenen Zusammenhänge. Entsprechend arbeiten Forschergruppen, die sich diesem Strang der Systemmedizin zuordnen, an der Aufklärung von Signalwegen und molekularen *pathways* zwischen Genom und Krankheits-Phänotyp (10). An dieser Stelle verzahnen sich systemmedizinische Forschungsanliegen häufig mit Fragestellungen, die in den letzten Jahren auch unter den Schlagworten der personalisierten oder individualisierten Medizin verfolgt wurden.

- Wurden schon im Rahmen der personalisierten Medizin zum Zweck der Biomarker-Bestimmung große (klinisch-)epidemiologische Studien wie GANI_MED durchgeführt (11–12), so stößt die Assoziationsforschung im Bereich der Systemmedizin in eine neue wissenschaftstheoretische Dimension vor. Denn mit „Big Data“ wird der innermedizinisch etablierten Vorgehensweise, falsifizierbar formulierte Hypothesen experimentell zu überprüfen, ein rein auf statistische Assoziationen medizinisch relevanter Merkmale abstellender und damit wesentlich probabilistischer Ansatz methodisch vorangestellt (13). Die kausale Überprüfung erfolgt der „Big-Data“-Methodologie zufolge erst im Nachgang zur Identifikation von statistisch signifikanten Zusammenhängen (13–14). Zur Generierung einer ausreichend starken Power sind dabei häufig sehr große Datenmengen („n = alles“) aus verschiedensten Quellen erfor-

Dr. rer. med. Erdmann, Dr. phil. Langanke: Theologische Fakultät, Lehrstuhl für Systematische Theologie, Ernst-Moritz-Arndt-Universität Greifswald
Dr. rer. med. Fischer: Department für Ethik, Theorie und Geschichte der Lebenswissenschaften, Universitätsmedizin Greifswald
Raths, M. Sc., Prof. Dr. rer. pol. Fleßa: Rechts- und Staatswissenschaftliche Fakultät, Lehrstuhl für Allgemeine Betriebswirtschaftslehre und Gesundheitsmanagement, Ernst-Moritz-Arndt-Universität Greifswald
*Die Autoren haben zu gleichen Anteilen an der Manuskripterstellung mitgewirkt.

Erdmann et al. 2015

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Was ist Systemmedizin?

[Mit „Systemmedizin“] werden zum einen grundlagentheoretische Ansätze bezeichnet, die darauf zielen, die biologischen Mechanismen der Krankheitsentstehung unter Nutzung von Methoden aus „omics“-Forschung, Systembiologie, Informatik und Netzwerktheorie besser zu verstehen. Diese verorten sich zum Teil explizit in der Nachfolge der personalisierten oder individualisierten Medizin

Erdmann et al. 2015

Eine Dokumentenanalyse der Abstracts der im Rahmen der e:med-Initiative des BMBF unter dem Label „Systemmedizin“ geförderten Vorhaben zeigt, dass diese Vorhaben PM/IM-Projekte mit (etwas) stärkerem Fokus auf den Einsatz informationstechnologischer Werkzeuge (der Datenauswertung) sind.

MEDIZINISCHE VERSORGUNG

Systemmedizin: Herausforderungen eines aktuellen Ansatzes

Die Thematisierung ethischer und ökonomischer Herausforderungen, die mit der Translation systemmedizinischer Ansätze in die Versorgung einhergehen, ist keineswegs neu.

Pia Erdmann*, Tobias Fischer*, Susan Raths*, Steffen Fleßa, Martin Langanke

Ein neues Zauberwort prägt die aktuellen Debatten um die Gesundheitsversorgung der Zukunft: „Systemmedizin“. Damit werden zum einen grundlagentheoretische Ansätze bezeichnet, die darauf zielen, die biologischen Mechanismen der Krankheitsentstehung unter Nutzung von Methoden aus „omics“-Forschung, Systembiologie, Informatik und Netzwerktheorie besser zu verstehen. Diese verorten sich zum Teil explizit in der Nachfolge der personalisierten oder individualisierten Medizin (1–6). Zum anderen vermarkten sich auch stärker anwendungsorientierte, „translationale“ Konzepte als Systemmedizin; damit sollen Klinikprozesse unter Nutzung medizininformatischer Tools nachhaltiger und patientenzentrierter gestaltet oder Behandlungsdaten für die wissenschaftliche Auswertung zugänglich gemacht werden (7–9). Allen diesen Ansätzen gemeinsam ist letztlich eine besonders starke Fokussierung auf den Einsatz informationstechnologischer Werkzeuge für medizinische Zwecke. Dabei reicht das Spektrum von hoch anspruchsvollen biomathematischen Modellierungsinstrumenten über hypothesenfreie „Big-Data“-Ansätze bis hin zur IT-Unterstützung von Workflows in Forschung und Klinik.

- Von der Entwicklung bioinformatischer Modelle erhofft man sich, pathologische Prozesse und die dahinter stehenden Mechanismen am Computer simulieren und so gezielt Interventionsoptionen nicht-invasiv auf ihre Effektivität hin prüfen zu können (4, 5). Voraussetzung dafür ist ein kausales Verständnis der jeweils relevanten pathogenen Zusammenhänge. Entsprechend arbeiten Forschergruppen, die sich diesem Strang der Systemmedizin zuordnen, an der Aufklärung von Signalwegen und molekularen *pathways* zwischen Genom und Krankheits-Phänotyp (10). An dieser Stelle verzahnen sich systemmedizinische Forschungsanliegen häufig mit Fragestellungen, die in den letzten Jahren auch unter den Schlagworten der personalisierten oder individualisierten Medizin verfolgt wurden.

- Wurden schon im Rahmen der personalisierten Medizin zum Zweck der Biomarker-Bestimmung große (klinisch-)epidemiologische Studien wie GANI_MED durchgeführt (11–12), so stößt die Assoziationsforschung im Bereich der Systemmedizin in eine neue wissenschaftstheoretische Dimension vor. Denn mit „Big Data“ wird der innermedizinisch etablierten Vorgehensweise, falsifizierbar formulierte Hypothesen experimentell zu überprüfen, ein rein auf statistische Assoziationen medizinisch relevanter Merkmale abstellender und damit wesentlich probabilistischer Ansatz methodisch vorangestellt (13). Die kausale Überprüfung erfolgt der „Big-Data“-Methodologie zufolge erst im Nachgang zur Identifikation von statistisch signifikanten Zusammenhängen (13–14). Zur Generierung einer ausreichend starken Power sind dabei häufig sehr große Datenmengen („n = alles“) aus verschiedensten Quellen erfor-

Dr. rer. med. Erdmann, Dr. phil. Langanke, Theologische Fakultät, Lehrstuhl für Systematische Theologie, Ernst-Moritz-Arndt-Universität Greifswald
Dr. rer. med. Fischer, Department für Ethik, Theorie und Geschichte der Lebenswissenschaften, Universitätsmedizin Greifswald
Raths, M. Sc., Prof. Dr. rer. pol. Fleßa, Rechts- und Staatswissenschaftliche Fakultät, Lehrstuhl für Allgemeine Betriebswirtschaftslehre und Gesundheitsmanagement, Ernst-Moritz-Arndt-Universität Greifswald

*Die Autoren haben zu gleichen Anteilen an der Manuskripterstellung mitgewirkt.

Erdmann et al. 2015

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

1	Die verschiedenen „approaches“ in der modernen Biomedizin sind unscharf gegeneinander abgegrenzt.
2	Sie verfolgen vielfach Ziele, die Ziele medizinischer Forschung überhaupt sind.
3	Die „approaches“ sind methodisch häufig disparate „Felder“ oder die eingesetzten Methoden sind unspezifisch.
4	Diese mehrfache Unspezifität führt dazu, dass eine Bewertung der „appoaches“ erschwert ist, ihr Erfolg oder Misserfolg kann kaum klar eingeschätzt werden.

Die Frage nach dem translationalen Erfolg biomedizinsicher Forschungsvorhaben ist auf der Ebene der „approches“ methodisch kaum abschließend zu klären. Dies hat wesentlich (auch) begrifflich-definitorische und damit a-priorische Ursachen.

Das Translationsproblem wird aber nicht gelöst, wenn wir uns der Beurteilbarkeit in der Biomedizin einfach durch Rekurs auf unspezifische „Konzepte“ entziehen.

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Stattdessen?		
S	**Spezifisch**	Ziele müssen eindeutig definiert sein (nicht vage, sondern so präzise wie möglich).
M	**Messbar**	Ziele müssen messbar sein (Messbarkeitskriterien).
A	**Akzeptiert**	Ziele müssen von den Empfängern akzeptiert werden/sein (auch: angemessen, attraktiv, abgestimmt ausführbar oder **anspruchsvoll**).
R	**Realistisch**	Ziele müssen möglich sein.
T	**Terminiert**	Zu jedem Ziel gehört eine klare Terminvorgabe.

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Stattdessen?		
S	**Spezifisch**	Ziele einzelner biomedizinischer Forschungsvorhaben müssen spezifischer formuliert sein als das allgemeine Ziel der Medizin („Verbesserung der Gesundheitsversorgung“). Forschung sollte direkt operationalisierbare Ziele angeben, d.h. Ziele, die mit spezifischen Methoden verfolgt werden können.

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Stattdessen?		
M	**Messbar**	Ziele biomedizinischer Forschung müssen messbar sein. Die Ziele abstrakter „approaches“ sind häufig so unspezifisch, dass über die Zielerreichung nicht präzise entschieden werden kann.

2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Stattdessen?		
A	**Akzeptiert**	Die gesellschaftliche Öffentlichkeit gibt dem Sektor der biomedizinischen Forschung großen Kredit. Diese breite Akzeptanz sollte jedoch nicht dazu führen, die Öffentlichkeit mit wohlfeilen Phrasen („wir verbessern Ihre Behandlung“) abzuspeisen. Denn solche Ziele sind nicht wissenschaftlich **anspruchsvol**l. **Anspruchsvoll** sind in der Wissenschaft nur Ziele, die operationalisiert und daher auch klar verfehlt werden können.

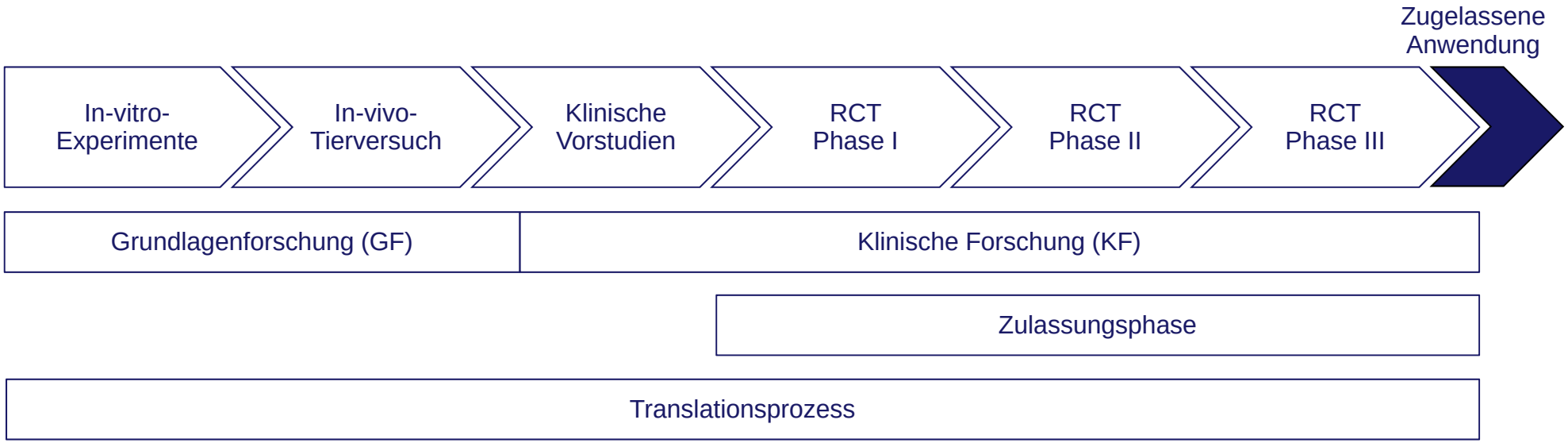
2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Stattdessen?		
R	**Realistisch**	Ziele biomedizinischer Forschung müssen so formuliert sein, dass sie mit den Methoden der biomedizinischen Disziplinen auch direkt angegangen werden können. Die Lösung gesellschaftlicher „Großaufgaben“ (z.B. „das demographische Problem“) ist kein sinnvoller methodischer Selbstanspruch der Biomedizin. Denn solche Ziele kann die Biomedizin mit ihrem Methodenspektrum nicht direkt adressieren.

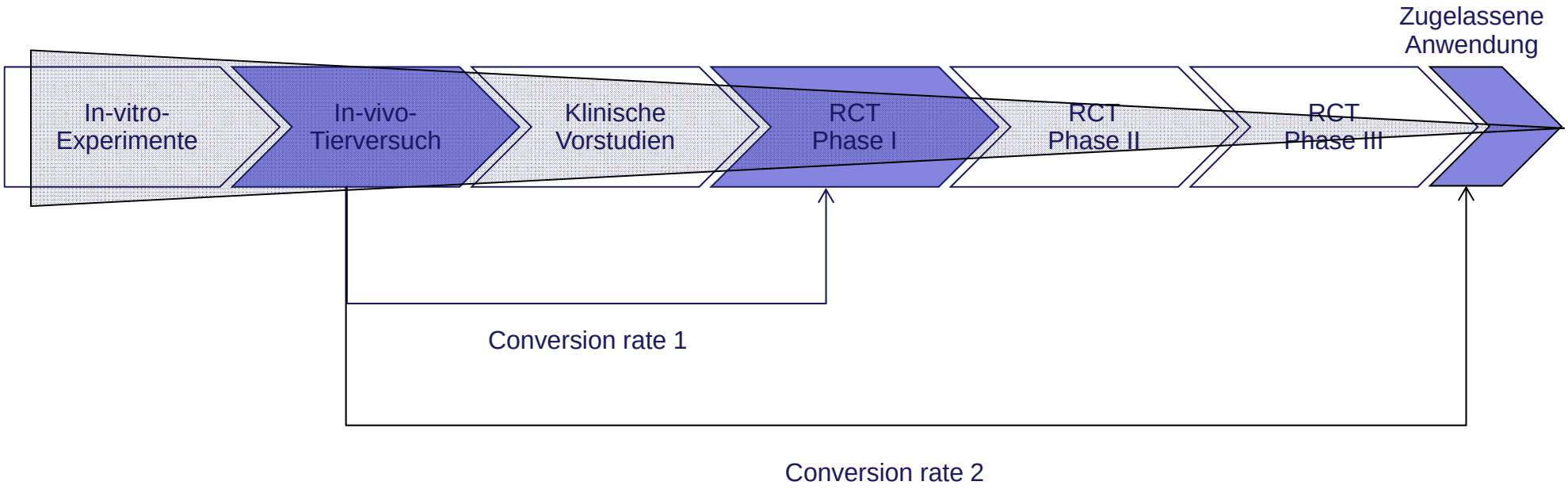
2. „Approaches“ in der Biomedizin – UnSMARTe Ziele in der Forschung

Stattdessen?		
T	**Terminiert**	Wissenschaftlicher Fortschritt ist nicht oder nur sehr eingeschränkt planbar. Retrospektive Daten zur durchschnittlichen Dauer von Translationsprozessen sollten jedoch zur Kenntnis genommen werden, um unhaltbare Versprechungen bzgl. der Zeitkorridore zu vermeiden.

3. Tiermodelle – Translation von Tierversuchsergebnissen



3. Tiermodelle – Translation von Tierversuchsergebnissen



3. Tiermodelle – Translation von Tierversuchsergebnissen

2000 Artikel aus 7 High-Impact-Journals im Erscheinungszeitraum 1980 bis 2000

76 Tierversuchsstudien als Sample gemäß Einschlusskriterien

28 in randomisierten KF-Studien repliziert

Hypothesen aus 14 Studien für Humananwendungen in KF „durchgefallen“

8 replizierte KF-Anwendungen für Patienten zugelassen

LETTERS

Smoking was an exclusion criterion for controls, whereas 4 of the 21 cases were regular smokers of 2 to 10 cigarettes per day. Mean urinary excretion rates of 8-iso-$PGF_{2\alpha}$ were similar in the 4 smokers (404 pg/mg of creatinine) and in the 21 cases considered as a whole (482 pg/mg of creatinine). Urine albumin excretion rates were not tested. There was only a small glucose variability between each day (day 1 mean amplitude of glycemic excursions [MAGE], 74 mg/dL; day 2 MAGE, 76 mg/dL), and MAGE values on day 1 and day 2 were highly correlated ($r=0.87$; $P<.001$).

Finally, conflicting observations in the study by O'Byrne et al[4] could have resulted from the use of different methods in different groups of patients at different ages: enzyme immunoassay in our study (21 patients with type 2 diabetes; mean age of 64 years) vs stable isotope dilution mass spectrometry assay in O'Byrne et al (13 patients with type 1 diabetes; mean age of 36 years).

Louis Monnier, MD
l-monnier@chu-montpellier.fr
Department of Metabolic Diseases
Lapeyronie Hospital
Emilie Mas, PhD
University Institute of Clinical Research
University of Montpellier
Christine Ginet, MD
Department of Metabolic Diseases
Lapeyronie Hospital
Françoise Michel, MD
University Institute of Clinical Research
University of Montpellier
Laetitia Villon, MD
Department of Metabolic Diseases
Lapeyronie Hospital
Jean-Paul Cristol, MD
Claude Colette, PhD
University Institute of Clinical Research
University of Montpellier
Montpellier, France

Financial Disclosures: None reported.

1. Buckingham B, Block J, Wilson DM. Continuous glucose monitoring. *Curr Opin Endocrinol Diabetes*. 2005;12:273-279.
2. Roberts LJ, Morrows JD. Measurement of F(2)-isoprostanes as an index of oxidative stress in vivo. *Free Radic Biol Med*. 2000;28:505-513.
3. Brownlee M, Hirsch IB. Glycemic variability: a hemoglobin A_{1c}-independent risk factor for diabetic complications. *JAMA*. 2006;295:1707-1708.
4. O'Byrne S, Forte P, Roberts J, et al. Nitric oxide synthesis and isoprostane production in subjects with type 1 diabetes and normal urinary albumin excretion. *Diabetes*. 2000;49:857-862.

RESEARCH LETTER

Translation of Research Evidence From Animals to Humans

To the Editor: Most medical therapies in use today were initially developed and tested in animals,[1] yet animal experiments often fail to replicate when tested in rigorous human trials.[2,3] We conducted a systematic review to determine how often highly cited animal studies translate into successful human research.

Figure 1. Methodological Quality of Animal Trials (n=76)

Methods. The 7 leading scientific journals by citation impact factor (Journal Citation Reports, Thomson Scientific, Philadelphia, Pa, 2004) that regularly publish original animal studies were searched: *Science*, *Nature*, *Cell*, *Nature Medicine*, *Nature Genetics*, *Nature Immunology*, and *Nature Biotechnology*. Articles with more than 500 citations were retrieved under the assumption that such prominent findings would more likely be tested in subsequent human trials.[4] A total of 2000 articles published between 1980 and 2000 were screened, reflecting advances in molecular biology and recombinant genetics. Articles were included if they investigated a preventive or therapeutic intervention in an in vivo animal model. When there were multiple animal studies of the same intervention, the most cited study was retained. Power calculations ($\alpha=0.05$, $\beta=0.05$) estimated that 49 articles were needed to exclude a translation rate below 5%.

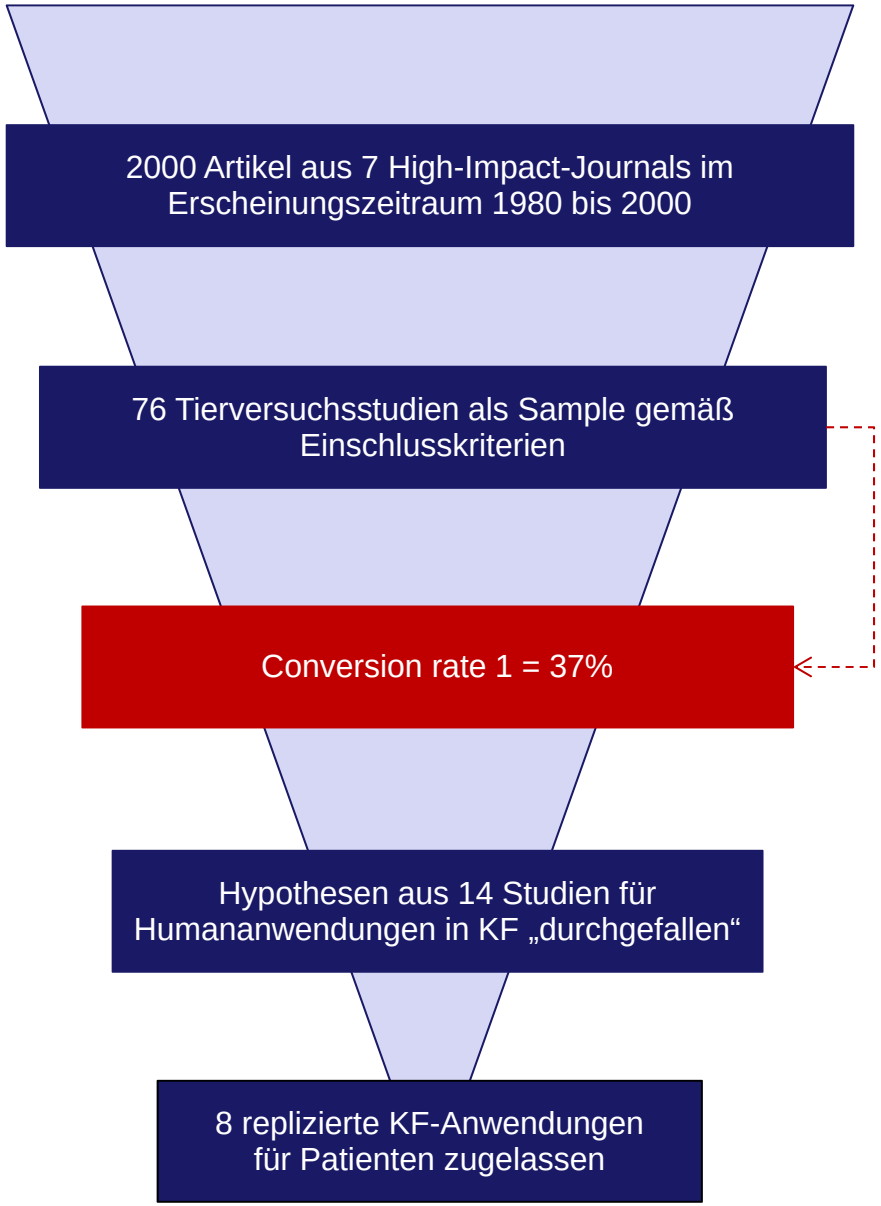
For each included study, a literature search identified human studies that translated the animal evidence. Successful translation was defined as replication in a randomized trial yielding results that were statistically positive according to primary outcome. Interventions and diseases analogous to those studied in the animal study were allowed.

MEDLINE, EMBASE, the Cochrane Central Register of Controlled Trials, the Cochrane Database of Systematic Reviews, the National Institutes of Health Clinical Trials Database, BIOSIS Previews, and the International Pharmaceutical Abstracts Database were searched from their inception through May 2006. Bibliographies of topic-specific review articles were manually searched for additional studies and experts were contacted if the search was negative.

The quality of the studies was assessed based on adapted standards for the conduct of animal research (FIGURE 1).[5] Good quality was defined as a global methodology score of 50% or higher. Multivariable logistic regression was used to assess predictors of translation. The Pearson correlation test was used to determine if methodological quality of ani-

Hackam et al. 2006

3. Tiermodelle – Translation von Tierversuchsergebnissen



LETTERS

Smoking was an exclusion criterion for controls, whereas 4 of the 21 cases were regular smokers of 2 to 10 cigarettes per day. Mean urinary excretion rates of 8-iso-$PGF_{2\alpha}$ were similar in the 4 smokers (404 pg/mg of creatinine) and in the 21 cases considered as a whole (482 pg/mg of creatinine). Urine albumin excretion rates were not tested. There was only a small glucose variability between each day (day 1 mean amplitude of glycemic excursions [MAGE], 74 mg/dL; day 2 MAGE, 76 mg/dL), and MAGE values on day 1 and day 2 were highly correlated ($r=0.87$; $P<.001$).

Finally, conflicting observations in the study by O'Byrne et al[4] could have resulted from the use of different methods in different groups of patients at different ages: enzyme immunoassay in our study (21 patients with type 2 diabetes; mean age of 64 years) vs stable isotope dilution mass spectrometry assay in O'Byrne et al (13 patients with type 1 diabetes; mean age of 36 years).

Louis Monnier, MD
l-monnier@chu-montpellier.fr
Department of Metabolic Diseases
Lapeyronie Hospital
Emilie Mas, PhD
University Institute of Clinical Research
University of Montpellier
Christine Ginet, MD
Department of Metabolic Diseases
Lapeyronie Hospital
Françoise Michel, MD
University Institute of Clinical Research
University of Montpellier
Laetitia Villon, MD
Department of Metabolic Diseases
Lapeyronie Hospital
Jean-Paul Cristol, MD
Claude Colette, PhD
University Institute of Clinical Research
University of Montpellier
Montpellier, France

Financial Disclosures: None reported.

1. Buckingham B, Block J, Wilson DM. Continuous glucose monitoring. *Curr Opin Endocrinol Diabetes*. 2005;12:273-279.
2. Roberts LJ, Morrows JD. Measurement of F(2)-isoprostanes as an index of oxidative stress in vivo. *Free Radic Biol Med*. 2000;28:505-513.
3. Brownlee M, Hirsch IB. Glycemic variability: a hemoglobin A_{1c}-independent risk factor for diabetic complications. *JAMA*. 2006;295:1707-1708.
4. O'Byrne S, Forte P, Roberts J, et al. Nitric oxide synthesis and isoprostane production in subjects with type 1 diabetes and normal urinary albumin excretion. *Diabetes*. 2000;49:857-862.

RESEARCH LETTER

Translation of Research Evidence From Animals to Humans

To the Editor: Most medical therapies in use today were initially developed and tested in animals,[1] yet animal experiments often fail to replicate when tested in rigorous human trials.[2,3] We conducted a systematic review to determine how often highly cited animal studies translate into successful human research.

Figure 1. Methodological Quality of Animal Trials (n=76)

Methods. The 7 leading scientific journals by citation impact factor (Journal Citation Reports, Thomson Scientific, Philadelphia, Pa, 2004) that regularly publish original animal studies were searched: *Science*, *Nature*, *Cell*, *Nature Medicine*, *Nature Genetics*, *Nature Immunology*, and *Nature Biotechnology*. Articles with more than 500 citations were retrieved under the assumption that such prominent findings would more likely be tested in subsequent human trials.[4] A total of 2000 articles published between 1980 and 2000 were screened, reflecting advances in molecular biology and recombinant genetics. Articles were included if they investigated a preventive or therapeutic intervention in an in vivo animal model. When there were multiple animal studies of the same intervention, the most cited study was retained. Power calculations ($\alpha=0.05$, $\beta=0.05$) estimated that 49 articles were needed to exclude a translation rate below 5%.

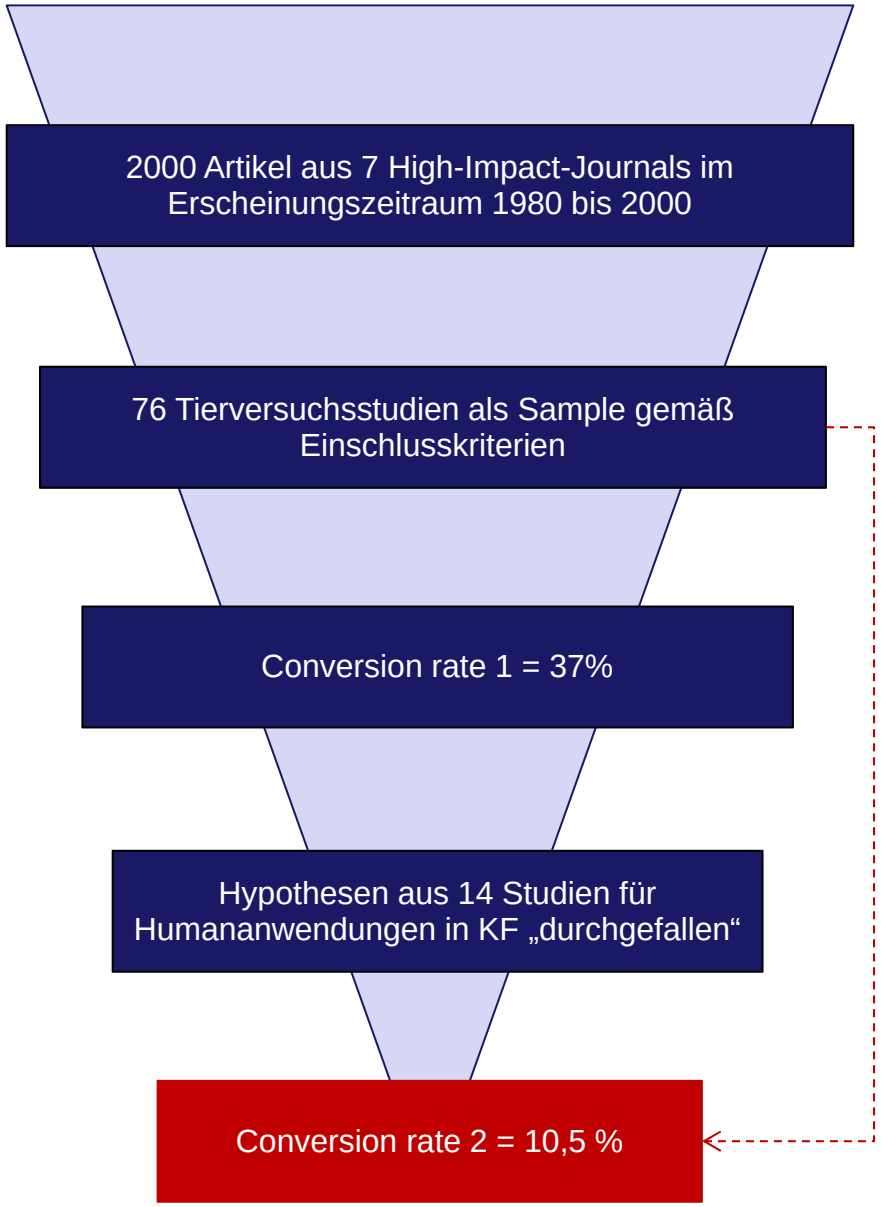
For each included study, a literature search identified human studies that translated the animal evidence. Successful translation was defined as replication in a randomized trial yielding results that were statistically positive according to primary outcome. Interventions and diseases analogous to those studied in the animal study were allowed.

MEDLINE, EMBASE, the Cochrane Central Register of Controlled Trials, the Cochrane Database of Systematic Reviews, the National Institutes of Health Clinical Trials Database, BIOSIS Previews, and the International Pharmaceutical Abstracts Database were searched from their inception through May 2006. Bibliographies of topic-specific review articles were manually searched for additional studies and experts were contacted if the search was negative.

The quality of the studies was assessed based on adapted standards for the conduct of animal research (FIGURE 1).[5] Good quality was defined as a global methodology score of 50% or higher. Multivariable logistic regression was used to assess predictors of translation. The Pearson correlation test was used to determine if methodological quality of ani-

Hackam et al. 2006

3. Tiermodelle – Translation von Tierversuchsergebnissen



LETTERS

Smoking was an exclusion criterion for controls, whereas 4 of the 21 cases were regular smokers of 2 to 10 cigarettes per day. Mean urinary excretion rates of 8-iso-$PGF_{2\alpha}$ were similar in the 4 smokers (404 pg/mg of creatinine) and in the 21 cases considered as a whole (482 pg/mg of creatinine). Urine albumin excretion rates were not tested. There was only a small glucose variability between each day (day 1 mean amplitude of glycemic excursions [MAGE], 74 mg/dL; day 2 MAGE, 76 mg/dL), and MAGE values on day 1 and day 2 were highly correlated ($r=0.87$; $P<.001$).

Finally, conflicting observations in the study by O'Byrne et al[4] could have resulted from the use of different methods in different groups of patients at different ages: enzyme immunoassay in our study (21 patients with type 2 diabetes; mean age of 64 years) vs stable isotope dilution mass spectrometry assay in O'Byrne et al (13 patients with type 1 diabetes; mean age of 36 years).

Louis Monnier, MD
l-monnier@chu-montpellier.fr
Department of Metabolic Diseases
Lapeyronie Hospital
Emilie Mas, PhD
University Institute of Clinical Research
University of Montpellier
Christine Ginet, MD
Department of Metabolic Diseases
Lapeyronie Hospital
Françoise Michel, MD
University Institute of Clinical Research
University of Montpellier
Laetitia Villon, MD
Department of Metabolic Diseases
Lapeyronie Hospital
Jean-Paul Cristol, MD
Claude Colette, PhD
University Institute of Clinical Research
University of Montpellier
Montpellier, France

Financial Disclosures: None reported.

1. Buckingham B, Block J, Wilson DM. Continuous glucose monitoring. *Curr Opin Endocrinol Diabetes*. 2005;12:273-279.
2. Roberts LJ, Morrows JD. Measurement of F(2)-isoprostanes as an index of oxidative stress in vivo. *Free Radic Biol Med*. 2000;28:505-513.
3. Brownlee M, Hirsch IB. Glycemic variability: a hemoglobin A_{1c}-independent risk factor for diabetic complications. *JAMA*. 2006;295:1707-1708.
4. O'Byrne S, Forte P, Roberts J, et al. Nitric oxide synthesis and isoprostane production in subjects with type 1 diabetes and normal urinary albumin excretion. *Diabetes*. 2000;49:857-862.

RESEARCH LETTER

Translation of Research Evidence From Animals to Humans

To the Editor: Most medical therapies in use today were initially developed and tested in animals,[1] yet animal experiments often fail to replicate when tested in rigorous human trials.[2,3] We conducted a systematic review to determine how often highly cited animal studies translate into successful human research.

Figure 1. Methodological Quality of Animal Trials (n=76)

Methods. The 7 leading scientific journals by citation impact factor (Journal Citation Reports, Thomson Scientific, Philadelphia, Pa, 2004) that regularly publish original animal studies were searched: *Science*, *Nature*, *Cell*, *Nature Medicine*, *Nature Genetics*, *Nature Immunology*, and *Nature Biotechnology*. Articles with more than 500 citations were retrieved under the assumption that such prominent findings would more likely be tested in subsequent human trials.[4] A total of 2000 articles published between 1980 and 2000 were screened, reflecting advances in molecular biology and recombinant genetics. Articles were included if they investigated a preventive or therapeutic intervention in an in vivo animal model. When there were multiple animal studies of the same intervention, the most cited study was retained. Power calculations ($\alpha=0.05$, $\beta=0.05$) estimated that 49 articles were needed to exclude a translation rate below 5%.

For each included study, a literature search identified human studies that translated the animal evidence. Successful translation was defined as replication in a randomized trial yielding results that were statistically positive according to primary outcome. Interventions and diseases analogous to those studied in the animal study were allowed.

MEDLINE, EMBASE, the Cochrane Central Register of Controlled Trials, the Cochrane Database of Systematic Reviews, the National Institutes of Health Clinical Trials Database, BIOSIS Previews, and the International Pharmaceutical Abstracts Database were searched from their inception through May 2006. Bibliographies of topic-specific review articles were manually searched for additional studies and experts were contacted if the search was negative.

The quality of the studies was assessed based on adapted standards for the conduct of animal research (FIGURE 1).[3] Good quality was defined as a global methodology score of 50% or higher. Multivariable logistic regression was used to assess predictors of translation. The Pearson correlation test was used to determine if methodological quality of ani-

Hackam et al. 2006

3. Tiermodelle – Translation von Tierversuchsergebnissen

51 eingereichte und genehmigte Tierversuchsvorhaben aus einer TVK im Zeitraum 1991 bis 1993

16 Vorhaben mit humanmedizinischer Perspektive und „positive results“

63 Primärpublikationen

1.183 Referenzierungen der 63 Primärpublikationen bis 2005

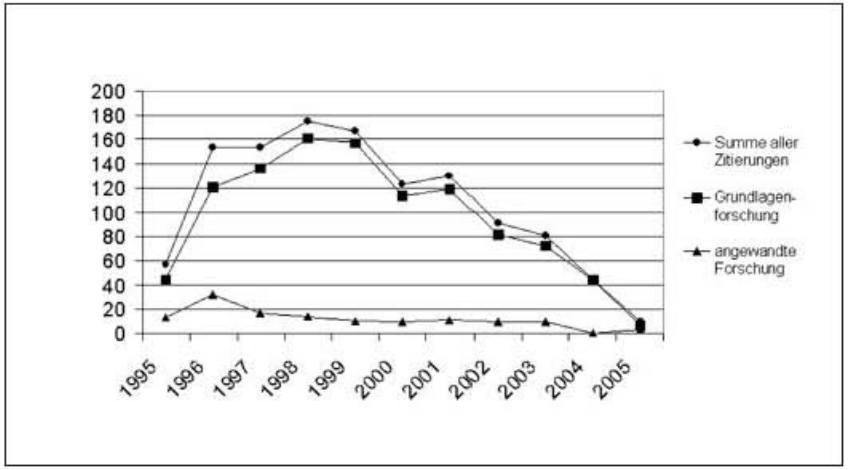


Abb. 1: Zeitlicher Verlauf sowie Häufigkeit aller Zitierungen innerhalb von 10 Jahren.

Tierversuche in der biomedizinischen Forschung

Eine Bestandsaufnahme der klinischen Relevanz von genehmigten Tierversuchsvorhaben: Nach 10 Jahren keine Umsetzung in der Humanmedizin nachweisbar.

Toni Lindl[1], Manfred Völkel[2] und Roman Kolar[3]

[1]Institut für angewandte Zellkultur, D-München, [2]Tierversuchskommission Nordbayern, Regierung von Unterfranken, D-Würzburg, [3]Tierschutzakademie Neubiberg, D-Neubiberg

Zusammenfassung

Tierexperimentell tätige Wissenschaftler müssen nach den gesetzlichen Vorschriften der Bundesrepublik Deutschland in ihren Anträgen auf Genehmigung eines Tierversuchsvorhabens begründen, inwieweit diese Tierversuche ethisch und wissenschaftlich gerechtfertigt sind. Als Begründung wird meist auf das fehlende Verständnis der Zusammenhänge bei der Entstehung von Krankheiten bzw. mit dem Fehlen entsprechender Therapien am Menschen hingewiesen.

Die Basis für die vorliegende Studie waren die bei den Genehmigungsbehörden eingereichten Forschungsanträge biomedizinischer Arbeitsgruppen aus drei Universitäten in Bayern. Einbezogen wurden 16 Anträge, die von 1991 bis 1993 in einer Tierversuchskommission nach § 15 des Tierschutzgesetzes eingereicht, bewilligt und in einer vorausgegangenen Studie (Toni Lindl et al., 2001) als erfolgreich eingestuft wurden.

Untersucht wurden die Zitierhäufigkeit, der Zitierverlauf und die Frage, in welche Forschungen die Primärzitate eingegangen sind: ob in weiteren tierexperimentellen Studien, in in vitro-Studien, in klinischen Studien oder in Übersichtsartikeln (sog. Reviews). Von ausschließlichem Interesse war, ob die Wissenschaftler das in den Anträgen postulierte Versuchsziel, eine neue Therapie oder überhaupt klinisch Relevantes zu entwickeln, erreichen konnten.

Das Ergebnis war enttäuschend: Es konnten zwar 97 klinisch orientierte Veröffentlichungen ermittelt werden, welche die oben erwähnten Publikationen zitierten (8% aller Zitierungen), aber nur bei 4 Studien (0,3%) wurde ein direkter Zusammenhang zwischen den tierexperimentellen Befunden der Antragsteller und den gefundenen Ergebnissen am Menschen hergestellt. Doch selbst hier konnte die im Tierversuch bestätigte Hypothese klinisch nicht in eine neue Therapie am Menschen umgesetzt werden. Entweder war kein therapeutischer Effekt nachweisbar, oder die Befunde am Menschen widersprachen sogar den Ergebnissen am Tier.

Als Konsequenz dieser Studie wird gefordert, die gesetzlichen Begründungen für einen Tierversuch durch die Behörde strenger zu prüfen und projektspezifische Argumente zu fordern, anstatt pauschale Begründungen zu dulden. Ferner müssen die Kompetenzen der prüfenden Behörden und der beratenden Kommissionen nach § 15 des TSchG dringend erweitert werden.

Summary: Animal experiments in biomedical research. An evaluation of the clinical relevance of approved animal experimental projects: No evident implementation in human medicine within 10 years.

According to the German Animal Welfare Act, scientists in Germany must provide an ethical and scientific justification for their application to the licensing authority prior to undertaking an animal experiment. Such justifications commonly include lack of knowledge on the development of human diseases or the need for better or new therapies for humans.

The present literature research is based on applications to perform animal experiments from biomedical study groups of three universities in Bavaria (Germany) between 1991 and 1993. These applications were classified as successful in the animal model in the respective publications (Lindl et al. ALTEX, 18, 171-178, 2001).

We investigated the frequency of citations, the course of citations, and in which type of research the primary publications were cited: subsequent animal-based studies, in vitro studies, review articles or clinical studies. The criterion we applied was whether the scientists succeeded in reaching the goal they postulated in their applications, i.e. to contribute to new therapies or to gain results with direct clinical impact. The outcome was unambiguous: even though 97 clinically orientated publications containing citations of the above-mentioned publications were found (8 % of all citations), only 4 publications evidenced a direct correlation between the results from animal experiments and observations in humans (0,3 %). However, even in these 4 cases the hypotheses that had been verified successfully in the animal experiment failed in every respect.

The implications of our findings may lead to demands concerning improvement of the licensing practice in Germany.

Keywords: ethical aspects, evaluation, animal experiments, clinical relevance

Das Manuskript wurde am 1. 6. 2005 eingereicht; am 4. 8. 2005 wurde die revidierte Fassung zum Druck angenommen

Lindl et al. 2005

3. Tiermodelle – Translation von Tierversuchsergebnissen

Literatur-Review bzgl. ausgewählter Indikationsfelder

- Sepsis
- Acquired Immunodeficience Syndrom
- Stroke
- Traumatic Brain Injury
- Spinal Cord Injury
- Asthma
- Cancer
- Neurodegenerative diseases

Conversion rate (2) für untersuchte Indikationsfelder : 0 – 5 %

Die erwartete Conversion rate liegt bei „health care workers“: > 40%

Eine Conversion rate < 20% wird von „health care workers“ als unzureichend bewertet.

Joffe et al. BMC Medical Ethics (2015) 16:29
DOI 10.1186/s12910-015-0024-x

BMC Medical Ethics

RESEARCH ARTICLE Open Access

Expectations for methodology and translation of animal research: a survey of health care workers

Ari R Joffe[1,2*], Meredith Bara[3], Natalie Anton[1] and Nathan Nobis[4,5]

Abstract

Background: Health care workers (HCW) often perform, promote, and advocate use of public funds for animal research (AR); therefore, an awareness of the empirical costs and benefits of animal research is an important issue for HCW. We aim to determine what health-care-workers consider should be acceptable standards of AR methodology and translation rate to humans.

Methods: After development and validation, an e-mail survey was sent to all pediatricians and pediatric intensive care unit nurses and respiratory-therapists (RTs) affiliated with a Canadian University. We presented questions about demographics, methodology of AR, and expectations from AR. Responses of pediatricians and nurses/RTs were compared using Chi-square, with $P < .05$ considered significant.

Results: Response rate was 44/114(39%) (pediatricians), and 69/120 (58%) (nurses/RTs). Asked about methodological quality, most respondents expect that: AR is done to high quality; costs and difficulty are not acceptable justifications for low quality; findings should be reproducible between laboratories and strains of the same species; and guidelines for AR funded with public money should be consistent with these expectations. Asked about benefits of AR, most thought that there are sometimes/often large benefits to humans from AR, and disagreed that "AR rarely produces benefit to humans." Asked about expectations of translation to humans (of toxicity, carcinogenicity, teratogenicity, and treatment findings), most expect translation >40% of the time; thought that misleading AR results should occur <21% of the time; and that if translation was to occur <20% of the time, they would be less supportive of AR. There were few differences between pediatricians and nurses/RTs.

Conclusions: HCW have high expectations for the methodological quality of, and the translation rate to humans of findings from AR. These expectations are higher than the empirical data show having been achieved. Unless these areas of AR significantly improve, HCW support of AR may be tenuous.

Keywords: Animal models, Animal research, Ethics, Methodology

Background

Biomedical animal research (AR) involves some harm to sentient animals including distress (due to confinement, boredom, isolation, and fear), pain, and early death [1-3]. AR is said to be morally permissible because the balance of these costs (harms to the animals) and benefits (to human medical care, quality of life, and survival) is favorable [4]. It is generally assumed that the benefits are great to human medicine [5]. An awareness of the empirical costs and benefits of AR is an important issue in medicine for several reasons. Health care workers (HCW) often perform (and are expected to perform) AR, promote AR directly with trainees and indirectly as role models, and advocate for use of public funds (from granting agencies and charitable foundations) toward medical related AR.

There is a growing literature that raises concerns about the empirical practice of AR in at least two domains. First, the methodological quality of AR is often poor in both experimental design and animal welfare aspects [6-12]. AR publications rarely report the use of eligibility criteria, randomization, allocation concealment, blinding, sample size calculation, primary outcome specification, and study

* Correspondence: ari.joffe@albertahealthservices.ca
[1]Department of Pediatrics, University of Alberta, Edmonton Clinic Health Academy, 4-546 11405 87 Ave, Edmonton, AB, Canada
[2]John Dossetor Health Ethics Center, University of Alberta, Edmonton, AB, Canada
Full list of author information is available at the end of the article

Joffe et al. 2015

3. Tiermodelle – Translation von Tierversuchsergebnissen

Forschungslage zur Translation von Ergebnissen aus Tierversuchen

1	Wenige Studien
2	Unterschiedliches methodisches Vorgehen, schlechte Vergleichbarkeit
3	Methodische Limitationen bei jeder einzelnen Studie

Abschätzung

1	*Hackam et al. 2006* derzeit methodisch beste verfügbare Studie, aber bzgl. der Conversion rates 1 und 2 sicherlich zu hoch (nur vielzitierte Studien aus High-Impact-Journals im Sample).
2	Die Ergebnisse von *Contopoulos-Ioannidis et al. 2003* legen für die Translation von GF zu KF in der medizinischen Forschung überhaupt eine Conversion rate 2 im unteren einstelligen %-Bereich nahe.

Conversion rate 1 für Tierversuche liegt bei maximal 3 : 1.
Conversion rate 2 für Tierversuche liegt im kleinen einstelligen %-Bereich (~ 5%).

3. Tiermodelle – Translation von Tierversuchsergebnissen

Ursachen für die unzureichende Translation aus dem Tierversuchswesen in die Humanmedizin

Ursachen mangelnder interner Validität

- Fehlende Randomisierung
- Fehlendes blinded outcome assessment
- Fehlendes allocation concealment
- Fehlende unabhängige study-conduct-Kontrollen
- Methodisch-statistische Auswertungsfehler
- Falsche oder fehlende Stichproben-Kalkualtion

OPEN ACCESS Freely available online PLoS MEDICINE

Research in Translation

Can Animal Models of Disease Reliably Inform Human Studies?

H. Bart van der Worp[1]*, David W. Howells[2], Emily S. Sena[2,3], Michelle J. Porritt[2], Sarah Rewell[2], Victoria O'Collins[2], Malcolm R. Macleod[3]

1 Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Utrecht, The Netherlands, 2 National Stroke Research Institute & University of Melbourne Department of Medicine, Austin Health, Melbourne, Australia, 3 Department of Clinical Neurosciences, University of Edinburgh, Edinburgh, United Kingdom

Animal experiments have contributed much to our understanding of mechanisms of disease, but their value in predicting the effectiveness of treatment strategies in clinical trials has remained controversial [1–3]. In fact, clinical trials are essential because animal studies do not predict with sufficient certainty what will happen in humans. In a review of animal studies published in seven leading scientific journals of high impact, about one-third of the studies translated at the level of human randomised trials, and one-tenth of the interventions, were subsequently approved for use in patients [1]. However, these were studies of high impact (median citation count, 889), and less frequently cited animal research probably has a lower likelihood of translation to the clinic. Depending on one's perspective, this attrition rate of 90% may be viewed as either a failure or as a success, but it serves to illustrate the magnitude of the difficulties in translation that beset even findings of high impact.

Recent examples of therapies that failed in large randomised clinical trials despite substantial reported benefit in a range of animal studies include enteral probiotics for the prevention of infectious complications of acute pancreatitis, NXY-059 for acute ischemic stroke, and a range of strategies to reduce lethal reperfusion injury in patients with acute myocardial infarction [4–7]. In animal models of acute ischemic stroke, about 500 "neuroprotective" treatment strategies have been reported to improve outcome, but only aspirin and very early intravenous thrombolysis with alteplase (recombinant tissue-plasminogen activator) have proved effective in patients, despite numerous clinical trials of other treatment strategies [8,9].

Linked Research Article

This Research in Translation discusses the following new study published in *PLoS Biology*:

Sena ES, van der Worp HB, Bath PMW, Howells DW, Macleod MR (2010) Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biol 8(3): e1000344. doi:10.1371/journal.pbio.1000344

Publication bias confounds attempts to use systematic reviews to assess the efficacy of various interventions tested in experiments modeling acute ischemic stroke, leading to a 30% overstatement of efficacy of interventions tested in animals.

Causes of Failed Translation

The disparity between the results of animal models and clinical trials may in part be explained by shortcomings of the clinical trials. For instance, these may have had insufficient statistical power to detect a true benefit of the treatment under study. For practical or commercial purposes, the designs of some clinical trials have also failed to acknowledge the limitations of efficacy observed in animal studies, for example by allowing therapy at later time points when the window of opportunity has passed [10,11]. Secondly, the failure of apparently promising interventions to translate to the clinic may also be caused by inadequate animal data and overoptimistic conclusions about efficacy drawn from methodologically flawed animal studies. A third possible explanation is the lack of external validity, or generalisability, of some animal models; in other words, that these do not sufficiently reflect disease in humans. Finally, neutral or negative animal studies may be more likely to remain unpublished than neutral clinical trials, giving the impression that the first are more often positive than the second. This article aims to address the possible sources of bias that threaten the internal and external validity of animal studies, to provide solutions to improve the reliability of such studies, and thereby to improve their translation to the clinic.

Internal Validity

Adequate internal validity of an animal experiment implies that the differences observed between groups of animals

Research in Translation discusses health interventions in the context of translation from basic to clinical research, or from clinical evidence to practice.

Citation: van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, et al. (2010) Can Animal Models of Disease Reliably Inform Human Studies? PLoS Med 7(3): e1000245. doi:10.1371/journal.pmed.1000245

Published March 30, 2010

Funding: This work was supported in part by the MRC Trials Methodology Hub and the National Health and Medical Research Council. The funders played no role in the decision to submit the article nor in its preparation.

Competing Interests: Malcolm R. MacLeod is on the Editorial Board of *PLoS Medicine*.

Abbreviations: ALS, amyotrophic lateral sclerosis; CAMARADES, Collaborative Approach to Meta-Analysis And Review of Animal Data from Experimental Stroke; CONSORT, CONsolidated Standards Of Reporting Trials

* E-mail: H.B.vanderWorp@umcutrecht.nl

Provenance: Commissioned; externally peer reviewed.

van der Worp et al. 2010

3. Tiermodelle – Translation von Tierversuchsergebnissen

Ursachen für die unzureichende Translation aus dem Tierversuchswesen in die Humanmedizin

Ursachen mangelnder externer Validität

- Fehlende Berücksichtigung von Alter und Comorbiditäts-Load bei Zielpatienten
- Übertragung von Effekten aus einer homogenen Tierpopulation auf heterogene Patientengruppen
- Fehlende Berücksichtigung des Geschlechtsaspekts
- Mangelnde Ähnlichkeit zwischen Modell- und Zielorganismus
- „Unrealistische“ Interventionsszenarien in den Tierversuchen
- Unterschiede bei den Outcome-Parametern sowie beim Zeitpunkt der Outcome-Analyse

OPEN ACCESS Freely available online PLoS MEDICINE

Research in Translation

Can Animal Models of Disease Reliably Inform Human Studies?

H. Bart van der Worp[1]*, David W. Howells[2], Emily S. Sena[2,3], Michelle J. Porritt[2], Sarah Rewell[2], Victoria O'Collins[2], Malcolm R. Macleod[3]

1 Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Utrecht, The Netherlands, 2 National Stroke Research Institute & University of Melbourne Department of Medicine, Austin Health, Melbourne, Australia, 3 Department of Clinical Neurosciences, University of Edinburgh, Edinburgh, United Kingdom

Animal experiments have contributed much to our understanding of mechanisms of disease, but their value in predicting the effectiveness of treatment strategies in clinical trials has remained controversial [1–3]. In fact, clinical trials are essential because animal studies do not predict with sufficient certainty what will happen in humans. In a review of animal studies published in seven leading scientific journals of high impact, about one-third of the studies translated at the level of human randomised trials, and one-tenth of the interventions, were subsequently approved for use in patients [1]. However, these were studies of high impact (median citation count, 889), and less frequently cited animal research probably has a lower likelihood of translation to the clinic. Depending on one's perspective, this attrition rate of 90% may be viewed as either a failure or as a success, but it serves to illustrate the magnitude of the difficulties in translation that beset even findings of high impact.

Recent examples of therapies that failed in large randomised clinical trials despite substantial reported benefit in a range of animal studies include enteral probiotics for the prevention of infectious complications of acute pancreatitis, NXY-059 for acute ischemic stroke, and a range of strategies to reduce lethal reperfusion injury in patients with acute myocardial infarction [4–7]. In animal models of acute ischemic stroke, about 500 "neuroprotective" treatment strategies have been reported to improve outcome, but only aspirin and very early intravenous thrombolysis with alteplase (recombinant tissue-plasminogen activator) have proved effective in patients, despite numerous clinical trials of other treatment strategies [8,9].

Linked Research Article

This Research in Translation discusses the following new study published in *PLoS Biology*:

Sena ES, van der Worp HB, Bath PMW, Howells DW, Macleod MR (2010) Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biol 8(3): e1000344. doi:10.1371/journal.pbio.1000344

Publication bias confounds attempts to use systematic reviews to assess the efficacy of various interventions tested in experiments modeling acute ischemic stroke, leading to a 30% overstatement of efficacy of interventions tested in animals.

Causes of Failed Translation

The disparity between the results of animal models and clinical trials may in part be explained by shortcomings of the clinical trials. For instance, these may have had insufficient statistical power to detect a true benefit of the treatment under study. For practical or commercial purposes, the designs of some clinical trials have also failed to acknowledge the limitations of efficacy observed in animal studies, for example by allowing therapy at later time points when the window of opportunity has passed [10,11]. Secondly, the failure of apparently promising interventions to translate to the clinic may also be caused by inadequate animal data and overoptimistic conclusions about efficacy drawn from methodologically flawed animal studies. A third possible explanation is the lack of external validity, or generalisability, of some animal models; in other words, that these do not sufficiently reflect disease in humans. Finally, neutral or negative animal studies may be more likely to remain unpublished than neutral clinical trials, giving the impression that the first are more often positive than the second. This article aims to address the possible sources of bias that threaten the internal and external validity of animal studies, to provide solutions to improve the reliability of such studies, and thereby to improve their translation to the clinic.

Internal Validity

Adequate internal validity of an animal experiment implies that the differences observed between groups of animals

Research in Translation discusses health interventions in the context of translation from basic to clinical research, or from clinical evidence to practice.

Citation: van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, et al. (2010) Can Animal Models of Disease Reliably Inform Human Studies? PLoS Med 7(3): e1000245. doi:10.1371/journal.pmed.1000245

Published March 30, 2010

Funding: This work was supported in part by the MRC Trials Methodology Hub and the National Health and Medical Research Council. The funders played no role in the decision to submit the article nor in its preparation.

Competing Interests: Malcolm R. MacLeod is on the Editorial Board of *PLoS Medicine*.

Abbreviations: ALS, amyotrophic lateral sclerosis; CAMARADES, Collaborative Approach to Meta-Analysis And Review of Animal Data from Experimental Stroke; CONSORT, CONsolidated Standards Of Reporting Trials

* E-mail: H.B.vanderWorp@umcutrecht.nl

Provenance: Commissioned; externally peer reviewed.

van der Worp et al. 2010

3. Tiermodelle – Translation von Tierversuchsergebnissen

Ursachen für die unzureichende Translation aus dem Tierversuchswesen in die Humanmedizin

525 Einzelveröffentlichungen zu stroke-bezogenen Tierversuchen über 16 Systematic Reviews erschlossen

Nur 10 Publikationen reporten keinen signifikanten Effekt der getesteten Intervention auf das Infarktvolumen

Nur 6 Veröffentlichungen reporten überhaupt keinen signifikanten Effekt

Egger- und Trimm-and-fill-Analysen begründen die These, dass rund 1/3 der berichteten Effekte auf einen publication bias, hier: auf das Unpubliziert-Bleiben von „negative results“ zurückzuführen sind.

OPEN ACCESS Freely available online PLOS BIOLOGY

Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy

Emily S. Sena[1,2,3], H. Bart van der Worp[4], Philip M. W. Bath[5], David W. Howells[2,3], Malcolm R. Macleod[1,6]*

1 Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom, 2 National Stroke Research Institute, Austin Health, University of Melbourne, Melbourne, Victoria, Australia, 3 Department of Medicine, Austin Health, University of Melbourne, Melbourne, Victoria, Australia, 4 Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, The Netherlands, 5 Stroke Trials Unit, University of Nottingham, Nottingham, England, United Kingdom, 6 Department of Neurology, NHS Forth Valley, Stirling, Scotland, United Kingdom

Abstract

The consolidation of scientific knowledge proceeds through the interpretation and then distillation of data presented in research reports, first in review articles and then in textbooks and undergraduate courses, until truths become accepted as such both amongst "experts" and in the public understanding. Where data are collected but remain unpublished, they cannot contribute to this distillation of knowledge. If these unpublished data differ substantially from published work, conclusions may not reflect adequately the underlying biological effects being described. The existence and any impact of such "publication bias" in the laboratory sciences have not been described. Using the CAMARADES (Collaborative Approach to Meta-analysis and Review of Animal Data in Experimental Studies) database we identified 16 systematic reviews of interventions tested in animal studies of acute ischaemic stroke involving 525 unique publications. Only ten publications (2%) reported no significant effects on infarct volume and only six (1.2%) did not report at least one significant finding. Egger regression and trim-and-fill analysis suggested that publication bias was highly prevalent (present in the literature for 16 and ten interventions, respectively) in animal studies modelling stroke. Trim-and-fill analysis suggested that publication bias might account for around one-third of the efficacy reported in systematic reviews, with reported efficacy falling from 31.3% to 23.8% after adjustment for publication bias. We estimate that a further 214 experiments (in addition to the 1,359 identified through rigorous systematic review; non publication rate 14%) have been conducted but not reported. It is probable that publication bias has an important impact in other animal disease models, and more broadly in the life sciences.

Citation: Sena ES, van der Worp HB, Bath PMW, Howells DW, Macleod MR (2010) Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy. PLoS Biol 8(3): e1000344. doi:10.1371/journal.pbio.1000344

Academic Editor: Ian Roberts, London School of Hygiene and Tropical Medicine, United Kingdom

Received August 24, 2009; **Accepted** February 18, 2010; **Published** March 30, 2010

Funding: We acknowledge financial support from the Scottish Chief Scientists' Office. MRM acknowledges the support of the Edinburgh MRC Trials Methodology Hub. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: CAMARADES, Collaborative Approach to Meta-analysis and Review of Animal Data in Experimental Studies; tPA, tissue plasminogen activator; IL1-RA, interleukin 1 receptor antagonist

* E-mail: malcolm.macleod@ed.ac.uk

Introduction

Few publications describing natural phenomena are in themselves sufficient to change our understanding of the world, and knowledge advances through the summarising of data in conference presentations, review articles, and books. Traditionally this process has been rather haphazard, with sometimes partisan experts using narrative review articles to emphasise their own particular perspective. Attempts have been made to account for this bias using the technique of systematic review, in which there is prespecification of the biological question being addressed, the methods through which contributing data will be identified, and the criteria that will be used to select which data are included in the analysis [1]. While systematic reviewers often go to some lengths to identify unpublished data sources, both approaches are potentially confounded by the ability to include only available data. If experiments have been conducted but are not available to reviewers, and if the results of these experiments as a group are not the same as results from experiments that were published, then both narrative and systematic reviews, and the resulting expert opinion and public understanding, will be biased. This is the "file drawer problem" [2,3]: at its most extreme, the 95% of studies that were truly neutral (that is, which reported no significant effects) remain in the files of the investigators, the 5% of experiments that were falsely positive are published, and reviewers conclude falsely that the literature represents biological truth.

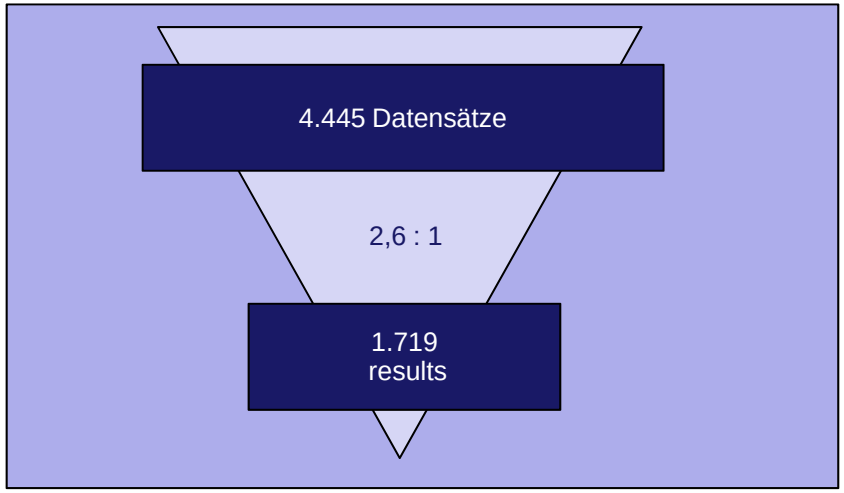
The consequences of the drawing of erroneous conclusions would be troubling if it involved, for instance, the interpretation of data from clinical trials; indeed, the recognition of a substantial publication bias in this literature has led to the introduction of clinical trial registration systems to ensure that those summarising research findings are at least aware of all relevant clinical trials that have been performed [4]. Publication bias has also been observed in reports of genetic association studies [5] and in ecology and evolution, in which 40% of meta-analyses were confounded by publication bias, and adjusting for publication bias might have altered the conclusions in around one-third of cases [6]. A related group of biases, the citation biases [7], can be addressed through rigorous systematic review, in that an attempt is made to include all relevant publications describing data meeting predefined

Sena et al. 2010

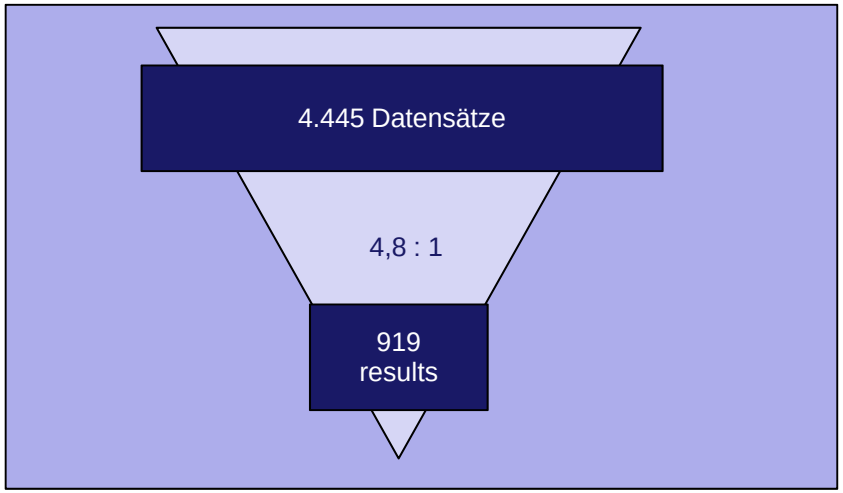
3. Tiermodelle – Translation von Tierversuchsergebnissen

Ursachen für die unzureichende Translation aus dem Tierversuchswesen in die Humanmedizin

Beobachtet



Statistisch zu erwarten



OPEN ACCESS Freely available online PLOS | BIOLOGY

Evaluation of Excess Significance Bias in Animal Studies of Neurological Diseases

Konstantinos K. Tsilidis[1☯], Orestis A. Panagiotou[1☯], Emily S. Sena[2,3], Eleni Aretouli[4,5], Evangelos Evangelou[1], David W. Howells[3], Rustam Al-Shahi Salman[2], Malcolm R. Macleod[2], John P. A. Ioannidis[6*]

1 Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, **2** Department of Clinical Neurosciences, University of Edinburgh, Edinburgh, United Kingdom, **3** The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Victoria, Australia, **4** Department of Methods and Experimental Psychology, University of Deusto, Bilbao, Spain, **5** Laboratory of Cognitive Neuroscience, School of Psychology, Aristotle University of Thessaloniki, Thessaloniki, Greece, **6** Stanford Prevention Research Center, Department of Medicine, and Department of Health Research and Policy, Stanford University School of Medicine, and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, United States of America

Abstract

Animal studies generate valuable hypotheses that lead to the conduct of preventive or therapeutic clinical trials. We assessed whether there is evidence for excess statistical significance in results of animal studies on neurological disorders, suggesting biases. We used data from meta-analyses of interventions deposited in Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Studies (CAMARADES). The number of observed studies with statistically significant results (O) was compared with the expected number (E), based on the statistical power of each study under different assumptions for the plausible effect size. We assessed 4,445 datasets synthesized in 160 meta-analyses on Alzheimer disease ($n = 2$), experimental autoimmune encephalomyelitis ($n = 34$), focal ischemia ($n = 16$), intracerebral hemorrhage ($n = 61$), Parkinson disease ($n = 45$), and spinal cord injury ($n = 2$). 112 meta-analyses (70%) found nominally ($p \leq 0.05$) statistically significant summary fixed effects. Assuming the effect size in the most precise study to be a plausible effect, 919 out of 4,445 nominally significant results were expected versus 1,719 observed ($p < 10^{-4}$). Excess significance was present across all neurological disorders, in all subgroups defined by methodological characteristics, and also according to alternative plausible effects. Asymmetry tests also showed evidence of small-study effects in 74 (46%) meta-analyses. Significantly effective interventions with more than 500 animals, and no hints of bias were seen in eight (5%) meta-analyses. Overall, there are too many animal studies with statistically significant results in the literature of neurological disorders. This observation suggests strong biases, with selective analysis and outcome reporting biases being plausible explanations, and provides novel evidence on how these biases might influence the whole research domain of neurological animal literature.

Citation: Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, et al. (2013) Evaluation of Excess Significance Bias in Animal Studies of Neurological Diseases. PLoS Biol 11(7): e1001609. doi:10.1371/journal.pbio.1001609

Academic Editor: Lisa Bero, University of California San Francisco, United States of America

Received February 13, 2013; **Accepted** June 6, 2013; **Published** July 16, 2013

Funding: No specific and/or direct funding was received for this study. No funding bodies played any role in the design, writing or decision to publish this manuscript. The authors were personally salaried by their institutions during the period of writing though no specific salary was set aside or given for the writing of this paper. There are no current external funding sources for this study.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: AD, Alzheimer disease; CAMARADES, Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Studies; E, number of expected studies with statistically significant results; EAE, experimental autoimmune encephalomyelitis; ICH, intracerebral hemorrhage; IQR, interquartile range; MBP, myelin basic protein; NOS, nitric oxide species; O, number of observed studies with statistically significant results; PD, Parkinson disease; RCT, randomized clinical trial; SCI, spinal cord injury

* E-mail: jioannid@stanford.edu

☯ These authors contributed equally to this work.

Introduction

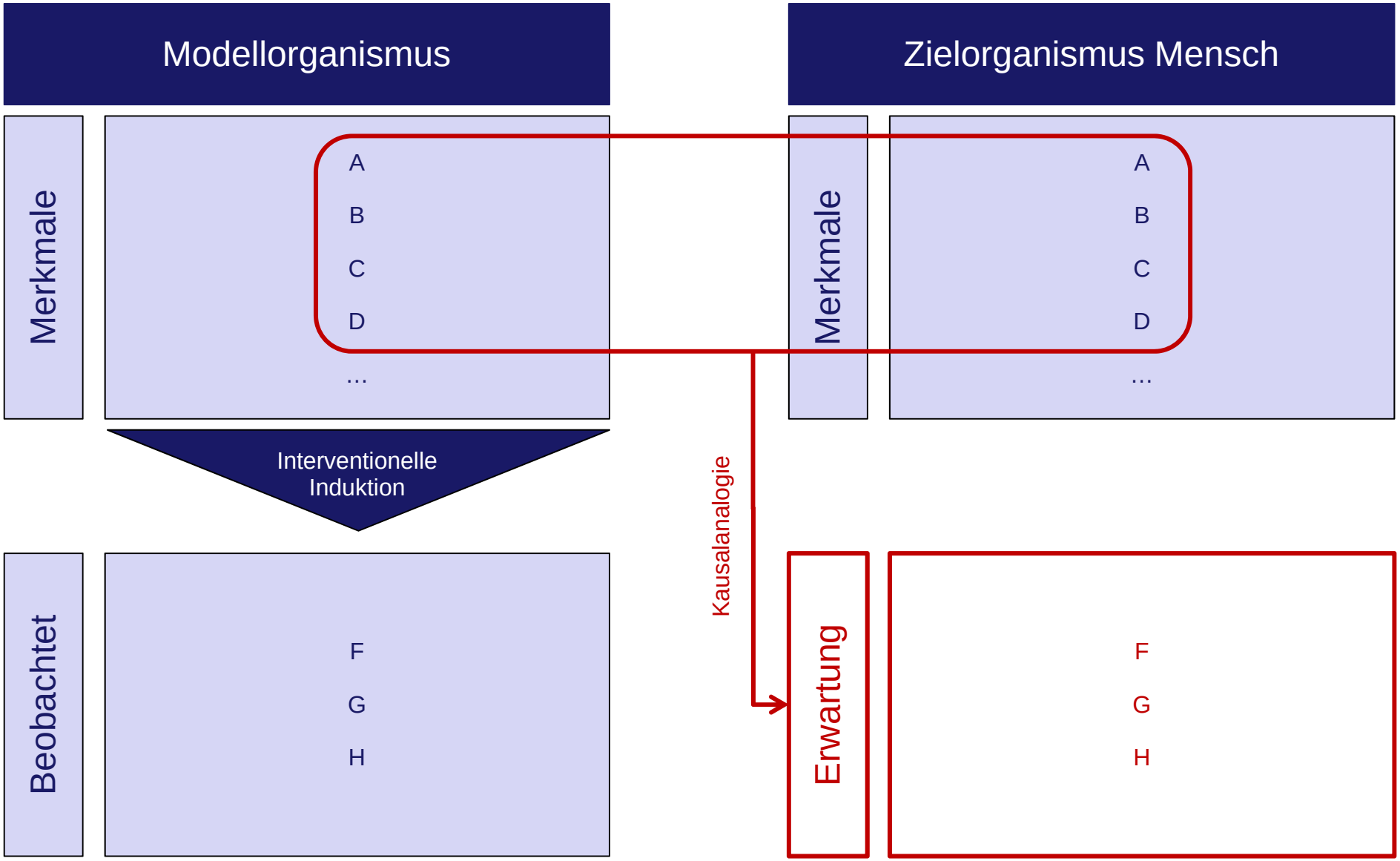
Animal research studies make a valuable contribution in the generation of hypotheses that might be tested in preventative or therapeutic clinical trials of new interventions. These data may establish that there is a reasonable prospect of efficacy in human disease, which justifies the risk to trial participants.

Several empirical evaluations of the preclinical animal literature have shown limited concordance between treatment effects in animal experiments and subsequent clinical trials in humans [1–4]. Systematic assessments of the quality of animal studies have attributed this translational failure, at least in part, to shortcomings in experimental design and in the reporting of results [5]. Lack of randomization, blinding, inadequate application of inclusion and exclusion criteria, inadequate statistical power, and inappropriate statistical analysis may compromise internal validity [6,7].

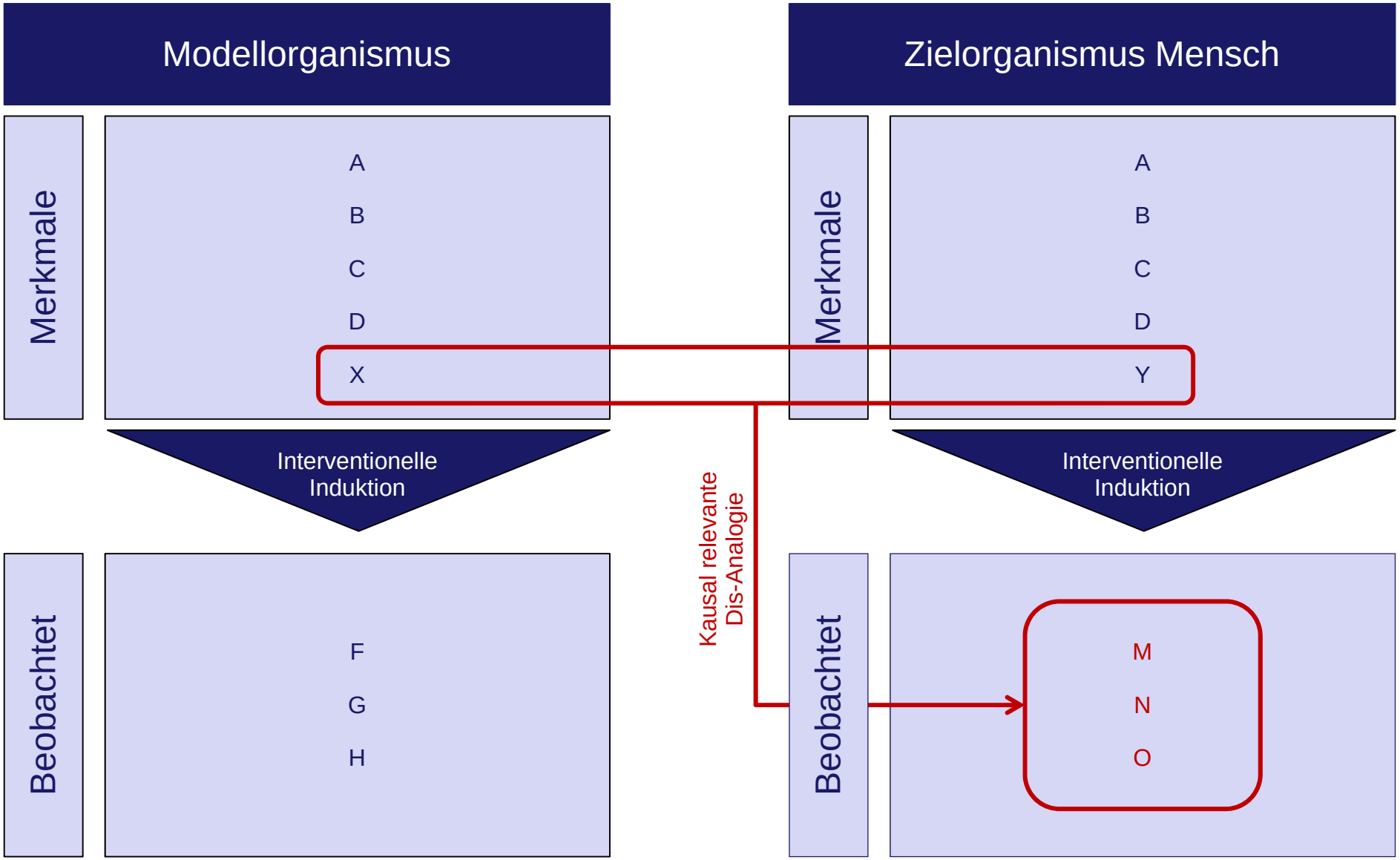
These problems are compounded by different types of reporting biases [8]. First, bias against publication of "negative" results (publication bias) or publication after considerable delay (time lag bias) may exist [9]. Such findings may not be published at all, published with considerable delay, or published in low impact or low visibility national journals in comparison to studies with "positive" findings. Second, selective analysis and outcome reporting biases may emerge when there are many analyses that can be performed, but only the analysis with the "best" results is presented resulting in potentially misleading findings [10]. This can take many different

Tsilidis et al. 2013

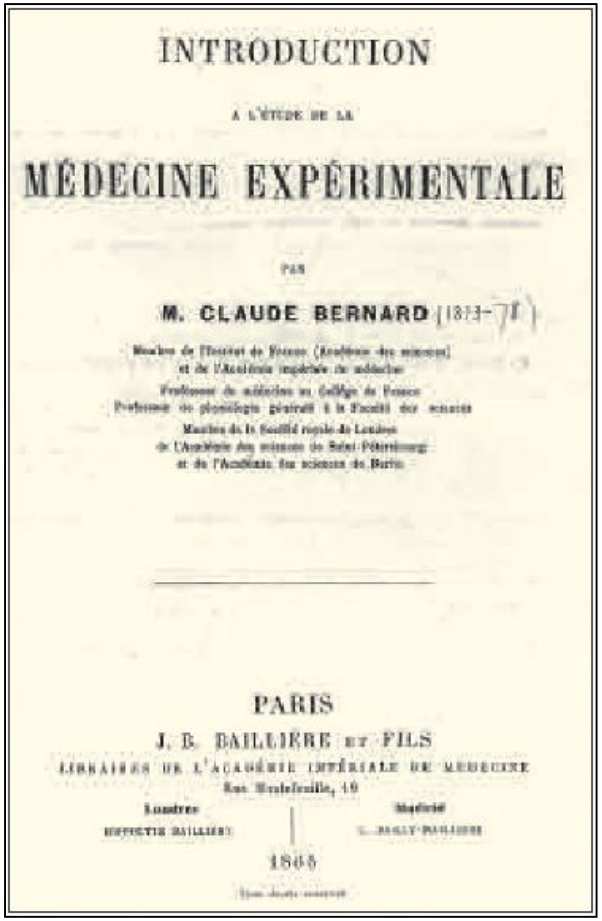
3. Tiermodelle – Translation von Tierversuchsergebnissen



3. Tiermodelle – Translation von Tierversuchsergebnissen



3. Tiermodelle – Translation von Tierversuchsergebnissen



INTRODUCTION

A L'ÉTUDE DE LA

MÉDECINE EXPÉRIMENTALE

PAR

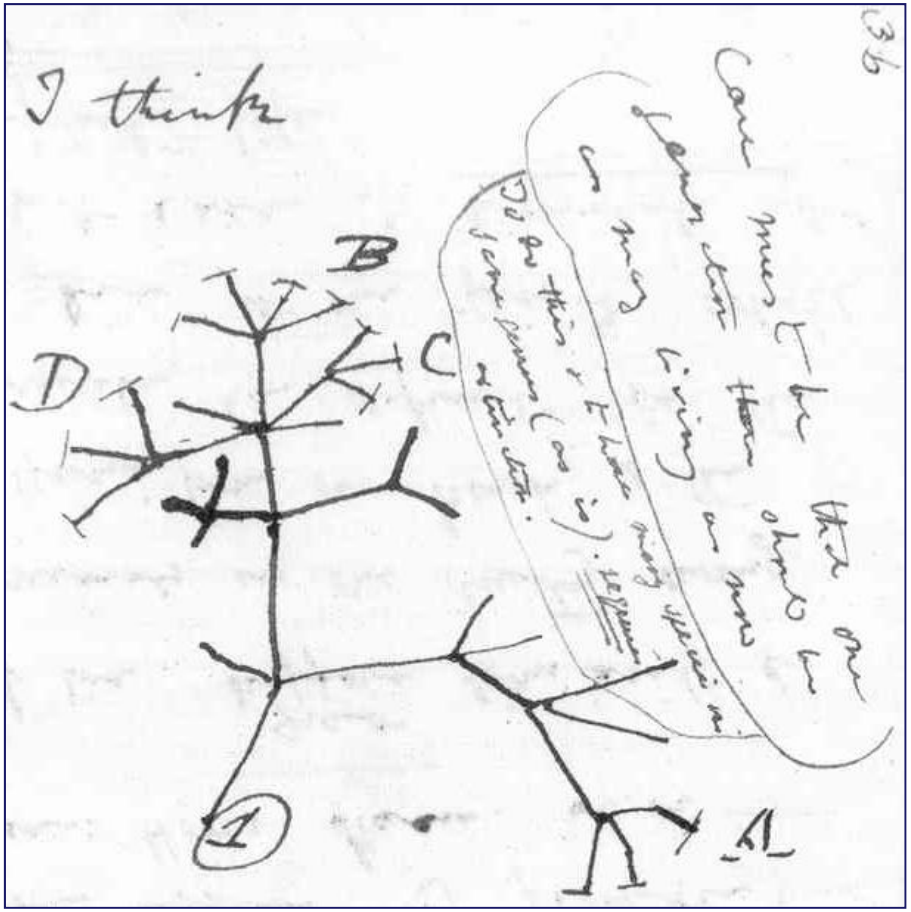
M. CLAUDE BERNARD (1813-78)

PARIS

J. B. BAILLIÈRE ET FILS

1865

Titelblatt von Claude Bernards
„Einführung in das Studium der Experimentellen Medizin“
(1865)



Darwins erste Skizze
eines evolutionären Stammbaums
(1837)

4. Empfehlungen

1	Statt Förderung nicht direkt operationalisierbarer Visionen, Förderung SMARTer Ziele!
2	Systematische krankheitsspezifische Überprüfung „etablierter“ Tiermodelle auf externe Validität durch retrospektive (Meta-)Studien!

5. Kontakt

Dr. phil. Martin Langanke, M.A.

Ernst-Moritz-Arndt-Universität Greifswald
Theologische Fakultät
Lehrstuhl für Systematische Theologie
Am Rubenowplatz 2-3
17487 Greifswald

T: 03834 86-2504
F: 03834 86-2520
E: langanke@uni-greifswald.de