

Higher Education Research
and Science Studies

RESEARCH

Sophie Biesenbender
Judith Hartstein *Hrsg.*

Qualitätsmessung als Prisma

Forschungsevaluation in der Medizin

DZHW.

OPEN ACCESS



Springer VS

Higher Education Research and Science Studies

Reihe herausgegeben von

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH,
Hannover, Niedersachsen, Deutschland

In der Reihe „Higher Education Research and Science Studies“ (HERSS) werden Monografien und referierte Sammelbände in deutscher oder englischer Sprache im Themenspektrum der Hochschul- und Wissenschaftsforschung veröffentlicht. Sie trägt mit der Fokussierung auf interdisziplinäre und international anschlussfähige Forschung insbesondere zur innovativen Entwicklung dieses Forschungsfeldes in der Schnittmenge von Hochschul- und Wissenschaftsforschung bei. Herausgegeben wird die Reihe HERSS vom Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), einem nationalen und internationalen Kompetenzzentrum für die Hochschul- und Wissenschaftsforschung. Das DZHW betreibt erkenntnis- und problemorientierte Forschung zu aktuellen und langfristigen Entwicklungen auf allen Ebenen des Hochschul- und Wissenschaftssystems.

Sophie Biesenbender · Judith Hartstein
(Hrsg.)

Qualitätsmessung als Prisma

Forschungsevaluation in der Medizin

Hrsg.

Sophie Biesenbender 
Deutsches Zentrum für Hochschul- und
Wissenschaftsforschung GmbH (DZHW)
Berlin, Deutschland

Judith Hartstein 
Robert K. Merton Zentrum für
Wissenschaftsforschung
Humboldt-Universität zu Berlin
Berlin, Deutschland

Deutsches Zentrum für Hochschul- und
Wissenschaftsforschung GmbH (DZHW)
Berlin, Deutschland



ISSN 2662-5709

ISSN 2662-5717 (electronic)

Higher Education Research and Science Studies

ISBN 978-3-658-43682-7

ISBN 978-3-658-43683-4 (eBook)

<https://doi.org/10.1007/978-3-658-43683-4>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://portal.dnb.de> abrufbar.

Die Veröffentlichung wurde gefördert aus dem Open-Access-Publikationsfonds der Humboldt-Universität zu Berlin.

© Der/die Herausgeber bzw. der/die Autor(en) 2024. Dieses Buch ist eine Open-Access-Publikation.

Open Access Dieses Buch wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Buch enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten. Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Marija Kojic

Springer VS ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Das Papier dieses Produkts ist recycelbar.

Qualität als Mäander – Geleitwort

Augustinus hat in den „*confessiones*“ seine Überlegungen zur Frage, wie denn die „Zeit“ zu definieren sei, mit einer klugen Bemerkung eingeleitet: „Wenn mich niemand danach fragt, weiß ich es; will ich einem Fragenden es erklären, weiß ich es nicht.“

Mit dem Versuch „Qualität“ zu definieren verhält es sich ähnlich, nur sind hier sehr viel mehr Interessen im Spiel als im Falle der Zeitdefinition. Für die Medizin gilt dies in besonderem Maße, denn neben den forschungstypischen Interessen nach robusten, replizierbaren Erkenntnissen steht das Patientenwohl, dessen Erhalt meist nicht unter den *Ceteris-Paribus*-Bedingungen des Laborexperiments möglich ist und wenn – wie in der Medizin der Fall – Qualitätsmaße schließlich in die Steuerung von Mittelverteilungssystemen einfließen, geht es auch um Geld und Reputation.

Dass sich unterschiedliche Interessen und Perspektiven so mit dem Qualitätsbegriff verschränken können, ist dem Umstand geschuldet, dass Qualität kein statisches oder einfach objektivierbares, sondern ein kontextsensitives und mehrdimensionales Konstrukt ist, das sich zudem nur indirekt über Indikatoren, die ihre jeweils eigenen Beschränkungen haben, abbilden lässt. Und nicht zuletzt enthält der Qualitätsdiskurs ein unhintergebar normatives Element, wenn es um die Auswahl, Gewichtung, Messung und Verrechnung einzelner Aspekte geht.

Es ist daher nicht verwunderlich, dass um den Qualitätsbegriff gestritten wird und die Akteure je nach Interessenlage unterschiedliche Kritikpunkte vortragen. Verwunderlich ist eher, dass dieser Diskurs häufig aus dem Anekdotischen schöpft, mit vorschnellen Kausalattributionen operiert, überholte Frontstellungen zwischen quantitativen und qualitativen Verfahren reaktiviert, kaum historische Tiefenschärfe hat und methodisch nicht zu rechtfertigende Gewissheiten propagiert. Sorgfältige empirische Analysen, wie sie in diesem Band vorgestellt

werden, sind eher die Ausnahme als die Regel. Zugleich lässt sich aber – trotz oft massiver Kritik – eine große Beharrlichkeit etablierter Praktiken und Empfehlungen feststellen. Einmal entwickelte Kompromisse – oft von kontroversen und konfliktreichen Auseinandersetzungen begleitet – fungieren dann als Quasi-Standard.

Welche Auswirkungen derartige Quasi-Standards haben, ist schwer zu beurteilen. Ihr Steuerungsimpuls ist einer unter vielen auf den Forscher und Organisationen reagieren. Viele populäre Vermutungen, wie etwa das „publish or perish“ sei eine Folge der in Mittelverteilungssystemen verwendeten Metriken, ignorieren völlig, dass dieser Aphorismus seit den späten 1920er Jahren in der Wissenschaft kursiert. Derartige Simplifizierungen führen offenkundig in die Irre, was aber nicht heißt, dass die eingesetzten Steuerungssysteme im kompetitiven Wissenschaftsbetrieb wirkungslos bleiben – im Positiven wie im Negativen; nur sind die Wirkungen meist nicht flächendeckend, nicht linear, nicht für alle Statusgruppen gleich und auch nicht unabhängig von der (Sub-)Disziplin oder lokalen Bedingungen, und sie interagieren unter Umständen auf unerwartete Weise.

Problematisch erscheint, dass etablierte Verfahren der Qualitätssicherung nur langsam oder gar nicht auf veränderte Zielsetzungen reagieren. In den letzten Jahrzehnten haben sich die Erwartungen an die Wissenschaft deutlich verändert. Unter dem Dach von „third mission“ finden sich Themen wie Wissenstransfer, Anwendungsorientierung, Bearbeitung großer gesellschaftlicher Herausforderungen, Nachwuchsqualifizierung, soziale Innovation und vieles mehr. Die Coronapandemie hat diese Dimensionen der Forschung nachdrücklich in die öffentliche Debatte transportiert. Valide und reliable Messinstrumente für diese Aktivitäten existieren allerdings bisher kaum.

Herausforderungen für die Qualitätsmessung und -sicherung ergeben sich aber nicht nur aus veränderten gesellschaftlichen Anforderungen. Auch wissenschaftsinterne Entwicklungen werfen Fragen auf: Ist ein vielzitiertes „systematic review“ möglicherweise überbewertet angesichts schnell wachsender maschineller Möglichkeiten derartiges zu erledigen? Ist die Fokussierung auf Grundlagenforschung zu eng angesichts der Koordinations- und Kooperationsnotwendigkeiten einer individualisierten Medizin? Ist die für eine nachhaltige Forschung notwendige Infrastruktur in den Qualitätskriterien unterbelichtet?

Ein dynamisch verstandenes Qualitätskonzept und darauf aufbauende Qualitätssicherungssysteme verlangen angesichts solcher Fragen eine beständige kritische Revision und ggf. Veränderung. Dafür aber reichen *common sense* und Bauchgefühl nicht aus. Es bedarf vielmehr einer systematischen Forschung, die

Bestehendes überprüft und für neue Fragestellungen neue Methoden und empirische Zugänge entwickelt und *last not least* daraus praxistaugliche Modelle entwickelt.

Der vorliegende Band ist ein exemplarischer Schritt in diese Richtung, der nicht nur Facetten des Qualitätskonzepts beleuchtet, sondern auch die Zusammenarbeit mit Praxispartnern erprobt.

Berlin, Deutschland
24. Juli 2024

Prof. a.D. Dr. Stefan Hornbostel

Danksagung

Besonderer Dank gilt Prof. a.D. Dr. Stefan Hornbostel und Prof. Dr. Christoph Herrmann-Lingen für die Anbahnung und laufende Begleitung des Projekts QuaMedFo sowie allen Mitarbeiter*innen der vier Teilprojekte von QuaMedFo für die gelungene und wertschätzende Zusammenarbeit. Weiterer Dank gilt den Pilotfakultäten des Projekts QuaMedFo für die Kooperation über die gesamte Projektlaufzeit. Für die interne Unterstützung während der Projektlaufzeit danken wir insbesondere Anh Nguyen Xuan, Camilo Almendrales und Valmira Xharavina. Schließlich danken wir dem Bundesministerium für Bildung und Forschung für die Förderung dieses Verbundvorhabens in vier Teilprojekten (Förderkennzeichen 16PU17011A, 16PU17011B, 16PU17011C und 16PU17011D).

Qualitätsmessung als Prisma – Einleitung

Die medizinische Wissenschaft steht ständig unter Beobachtung – durch die wissenschaftliche Community, Politik und Wirtschaft sowie die breite Öffentlichkeit. Im Zentrum stehen dabei Fragen der Kommunikation und des Transfers wissenschaftlicher Ergebnisse, der Translation in die medizinische Praxis, aber auch – wie im Zuge aktueller gesellschaftlicher Herausforderungen immer wieder deutlich wird – die Erwartungen, die aus verschiedenen Richtungen an die Wissenschaft und ihre Kompetenz für die Lösung gesellschaftlicher Probleme und Herausforderungen gestellt werden.

Die Betrachtung, Kategorisierung und Bewertung von Wissensproduktion und Forschungsleistungen erfolgt daher aus unterschiedlichen Perspektiven mit verschiedenen Zielstellungen, Erwartungen und Ansätzen. Bewertungsanlässe umfassen dabei auch Steuerungsanliegen auf verschiedenen Ebenen und letztlich die Beantwortung der Frage, unter welchen Bedingungen sich welche Themen bestmöglich beforschen lassen, um das Patientenwohl und die öffentliche Gesundheit zu fördern sowie Wissenschaft und Gesellschaft voranzubringen.

Die verschiedenen Perspektiven eint das Anliegen, die Qualität erfolgter Forschungsaktivitäten zu bemessen, um sie auf diese Weise greifbar zu machen und unterschiedliche zugrunde liegende Prozesse in Kontext zu setzen. Dafür werden unterschiedliche Informationen und Daten *über* Forschung (sogenannte Forschungsinformationen) benötigt, welche verschiedene Aspekte von Forschung bzw. Forschungsleistung abbilden und sich – je nach Verwendungszweck und Zielstellung – in vielfältige Kennzahlen oder Indikatoren überführen lassen. Die Vielfalt verfügbarer Indikatoren und Messkonzepte, lebhaft geführte Auseinandersetzungen innerhalb der Wissenschaftsforschung und der erweiterten wissenschaftlichen Community über Sinn und Unsinn bestimmter Kennzahlen

und Indikatoren bzw. ihr oftmals mangelnder Bezug zur eigentlichen Zielstellung, unbefriedigende Ausgangslagen bezüglich der Verfügbarkeit und Qualität von Forschungsinformationen sowie sich ändernde bzw. unklare Definitionen von Qualität stellen eine besondere Herausforderung für eine faire und sinnvolle Betrachtung und Bewertung von Forschung dar. In dieser dynamischen Gemengelage lassen sich keine objektivierbaren Empfehlungen für die Ausgestaltung von Monitoring- und Bewertungsinstrumenten ableiten. Jeder Kontext kann als individuell betrachtet werden, und es scheint ein aussichtsloses Unterfangen, vor diesem Hintergrund Standards und Empfehlungen zu identifizieren. Dennoch lassen sich einige Ansprüche und Beispiele guter Praxis für einen angemessenen, verantwortungsvollen und fairen Umgang mit Forschungsinformationen und Evaluationskontexten identifizieren, die in dem vorliegenden Band am Beispiel der medizinischen Forschung beleuchtet werden sollen.

Für diesen Band haben wir den Titel „Qualitätsmessung als Prisma“ gewählt: So wie ein optisches Prisma die verschiedenen Farbanteile einer Lichtquelle sichtbar macht, so legt die Qualitätsmessung gewissermaßen ganz unterschiedliche Bestandteile medizinischer Forschungsleistung frei – eine allumfassende Betrachtung, die alle Blickwinkel vereint erscheint unmöglich. Und so, wie Licht nicht einfach nur hell oder dunkel ist, so ist auch Forschungsleistung nicht einfach nur gut oder schlecht – sie erscheint in unterschiedlichen Farben, die je nach Kontext ihre Bedeutung entfalten. Auf diese Weise kommen je nach Messinstrument verschiedene Qualitäten besonders zur Geltung: etwa die Reputation in der wissenschaftlichen Gemeinschaft, der Praxistransfer oder die Relevanz für die gesellschaftliche Diskussion.

Im Verbundprojekt „**Qualitätsmaße zur Evaluation medizinischer Forschung**“ (QuaMedFo) in dessen Kontext dieser Band entstanden ist, kam der prismatische Charakter der Messung von Forschungsleistung besonders zum Tragen. Unser Regenbogen umfasst **Publikationen, Rezeption, Transfer und Effizienz** – befördert durch die multidisziplinäre Einbettung konnten diese vielen verschiedenen Dimensionen von Forschungsleistung behandelt werden.

Multidisziplinäre Perspektiven

Die **Medizin** als Startpunkt der Analyse ist in diesem Band mit Prof. Dr. Christoph Herrmann-Lingen und mit Christopher Traylor aus der Klinik für Psychosomatische Medizin und Psychotherapie der Universitätsmedizin Göttingen, der Prof. Dr. Herrmann-Lingen auch als Direktor vorsteht, vertreten: Jegliche Qualitätsmessung medizinischer Forschung muss sich am Stand der Debatte in

der medizinischen Fachgemeinschaft orientieren und so war es im Verbundprojekt maßgeblich Prof. Dr. Herrmann-Lingen zu verdanken, dass ein regelmäßiger Austausch zwischen den QuaMedFo-Verbundpartnern und den Praxispartner*innen und ein laufender Abgleich mit der Binnenperspektive stattfinden konnte. Auch durch langjährige Mitarbeit in der Arbeitsgemeinschaft Wissenschaftlicher Medizinischer Fachgesellschaften (AWMF) konnte er zusätzlich den nötigen Überblick über die medizinischen Teildisziplinen einbringen.

Die **Informationswissenschaften** sind mit den Bereichen Bibliometrie, Web Science und Data Science sehr breit vertreten:

Im Bereich der **Bibliometrie**, genauer der Indikatorik, bewegen sich die Beiträge von Valeria Aman vom Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) in Berlin. Sie hat sich auch um die Metadaten-Erschließung der Publikationskorpora der Praxispartner verdient gemacht, sodass andere Beiträge auf ihren kuratierten Publikationslisten aufbauen konnten.

Der Bereich **Web Science**, genauer der Altmetrics, ist in diesem Band durch Prof. Dr. Isabella Peters, Dr. Steffen Lemke und Anne Witthake vom Leibniz-Informationszentrum Wirtschaft (ZBW) in Kiel repräsentiert. Ihnen ist der Einbezug neuer web-basierter Metriken zu verdanken, die von klassischen bibliometrischen Indikatoren unterbelichtete Bereiche von Forschungsimpact ausleuchten.

Der Bereich **Data Science**, genauer der semantischen Suchtechnologie, ist hier mit Prof. Dr. Konrad Förstner und Dr. Klaus Lippert vom Informationszentrum Lebenswissenschaften (ZB MED) in Köln vertreten. Mit einer neuartigen Methode, medizinisch-wissenschaftliche Publikationen mit Patentdatenbanken zu verknüpfen ermöglichen sie die bessere Beforschung des Transfers medizinischer Forschung in die Wirtschaft.

Nicht zuletzt sind auch die **Sozialwissenschaften** durch die beiden Herausgeberinnen des Bandes, Dr. Sophie Biesenbender und Judith Hartstein, bei der Gestaltung von Projekt wie Sammelband involviert. Dr. Biesenbender, die im zweiten Projektjahr als Verbundleiterin von QuaMedFo auf Prof. a.D. Dr. Stefan Hornbostel folgte, stellte auch durch ihre Einbindung in wissenschaftliche Standardisierungsgremien die Anschlussfähigkeit des Projekts an die breitere Debatte um Forschungsevaluation in Deutschland sicher und gewährleistete den breiten Überblick. Judith Hartstein übernahm die wirtschaftswissenschaftlich motivierte Aufgabe der Effizienzmessung.

An wen richtet sich dieser Band?

Der vorliegende Sammelband gibt einerseits einen breiten Überblick über die Dimensionen und Möglichkeiten der Qualitätsmessung medizinischer Forschung, andererseits werden die verwendeten Metriken jeweils an konkreten Fällen vorgestellt. Sowohl für **medizinisch Forschende** als auch für diejenigen, die in der **Forschungsadministration** praktisch mit der Umsetzung von Evaluation betraut sind, ist somit ein – wie wir meinen sehr wertvolles – Handbuch entstanden, welches auch als Nachschlagewerk genutzt werden kann.

Bei der Zusammenstellung der Beiträge haben wir darauf geachtet, dass diese möglichst jeweils aus sich heraus verständlich sind. Die hier vorgestellten Aspekte medizinischer Forschungsevaluation bedienen sich verschiedener auch neuartiger methodischer Ansätze, deren Kenntnis wir nicht in jedem Fall voraussetzen können. Entsprechend werden die verwendeten Methoden jeweils kurz eingeführt, wobei sich in den Referenzen der Beiträge natürlich auch vertiefende Literatur findet für diejenigen, die eine solche Methode selbst verwenden möchten oder einfach interessiert sind.

Durch seine Konzeption scheint uns dieses Buch über die vorgenannten Zielgruppen hinaus auch geeignet, als Grundlage für **akademische Lehrveranstaltungen** zu dienen sowie die **politische Diskussion** zu informieren. Kurzum – wir hoffen, es finden sich viele Leserinnen und Leser.

Die Struktur dieses Bandes

Die Beiträge in diesem Band sind drei Bereichen zugeordnet: „**Motivation und Kontext**“, „**Dimensionen der Forschungsevaluation**“ und „**Ressourcen aus dem Projektkontext QuaMedFo**“.

Motivation und Kontext

Nach einem Geleitwort des inzwischen emeritierten Initiators des Projekts Qua-MedFo, Herrn Prof. a.D. Dr. Stefan Hornbostel zum Begriff der Qualität im Forschungskontext und zu seinen Dynamiken („Qualität als Mäander“), finden sich zwei Beiträge, die das Thema des Bandes – die Qualitätsmessung medizinischer Forschung – aus zwei komplementären Perspektiven einführen.

Im ersten Beitrag **„Verantwortungsvolle Wissenschaftsbewertung in der Medizin – was zählt?“** (Herrmann-Lingen) fasst Prof. Dr. Christoph Herrmann-Lingen den Stand der Debatte in der medizinischen Gemeinschaft zusammen, betrachtet die praktischen Herausforderungen bei der Evaluation medizinischer Forschung und beleuchtet Erfahrungen und Eindrücke der von den Evaluationen betroffenen Forscher*innen. Dabei stellt er besonders heraus, dass publikationsbasierte Leistungsindikatoren in der leistungsorientierten Mittelvergabe bzw. Indikatorgestützten Mittelallokation unintendierte Effekte hervorrufen können, und entwickelt daraus Desiderata für die Weiterentwicklung der Leistungsmessung in der Medizin.

Im zweiten Beitrag **„Dimensionen und Qualitätsmaße zur Evaluation medizinischer Forschung – Kontext und Ansätze des Projekts QuaMedFo“** (Biesenbender) nimmt Dr. Sophie Biesenbender den Faden auf und erläutert, wie die bestehenden Desiderata für die Forschungsevaluation in der Medizin im Projekt QuaMedFo (Qualitätsmaße zur Evaluation medizinischer Forschung) aufgegriffen und produktiv umgesetzt wurden, sodass dessen Erkenntnisse schließlich in den hier vorliegenden Sammelband münden konnten. Dabei fasst sie auch noch einmal die wesentlichen Projektergebnisse zusammen, deren vertiefende Darstellung dann in den weiteren Beiträgen dieses Bandes erfolgt.

Dimensionen der Forschungsevaluation

In diesem gedachten Abschnitt sind die empirischen Arbeitsergebnisse aus dem Projekt QuaMedFo jeweils detailliert dargestellt. Den Anfang macht der Beitrag **„Publikations- und Rezeptionsanalyse von drei medizinischen Fakultäten unter Berücksichtigung der LOM“** (Aman) von Valeria Aman, welcher die Publikationslisten von drei deutschen medizinischen Fakultäten erschließt und hinsichtlich Publikationsumfang und Rezeption in der wissenschaftlichen Gemeinschaft anhand von Zitationszahlen analysiert. Dabei wird das Spektrum der etablierten Leistungsindikatoren durch empirische Analysen nachvollziehbar abgedeckt.

Ebenfalls mit der Rezeption wissenschaftlicher Publikationen dieser Fakultäten beschäftigt sich der Beitrag **„Altmetrics zur Evaluation medizinischer Forschung in Deutschland“** (Lemke et al.) von Dr. Steffen Lemke, Anne Witthake und Prof. Dr. Isabella Peters – hier bezogen auf die Rezeption in Gesellschaft und Politik. Das Hauptaugenmerk liegt dabei auf der Abdeckung

von Publikationen durch Erwähnungen in den Sozialen Medien sowie in Policy-Dokumenten, wobei insbesondere bei letzteren die Unterschiede zwischen den medizinischen Teildisziplinen stark zum Tragen kommen.

Eine weitere Dimension der Rezeption wissenschaftlicher Publikationen, die auch ein Ausdruck von Praxistransfer ist, wird im Beitrag **„Welche Charakteristika weist die in Leitlinien zitierte Literatur auf? – Eine Fallstudie anhand einer S2e- und einer S3-Leitlinie“** (Traylor und Aman) von Christopher Traylor und Valeria Aman betrachtet. Der Beitrag geht von der Prämisse aus, dass die Bekanntheit von Publikationen ihrer Zitation voraus gehen muss, und untersucht, unter welchen Umständen die Zitation von wissenschaftlichen Publikationen (z. B. in Fachjournals) in medizinischen Leitlinien erfolgt. Dabei werden Faktoren wie die Sichtbarkeit von Publikationen allgemein bzw. die persönliche Kenntnis von Publikationen durch die Leitlinienautor*innen als mögliche Einflussgrößen geprüft.

Der Transfer medizinischer Forschungsergebnisse in die Wirtschaft ist Gegenstand des Beitrags **„Nutzung von Patentdaten zur Erfassung der wirtschaftlichen Verwertung von Forschung“** (Lippert und Förstner) von Dr. Klaus Lippert und Prof. Dr. Konrad Förstner. Der Beitrag behandelt das Auffinden von Patenten, die als Derivate aus medizinischen Forschungsergebnissen hervorgegangen sind. Transferprofile der Fakultäten können so über die Zuordnung von Patenten zu wissenschaftlichen Publikationen unter Nutzung spezifisch medizinischer Klassifikationsinstrumente angewendet auf Patentklassen dargestellt werden.

Die wirtschaftliche Betrachtung medizinischer Fakultäten als Produzenten wissenschaftlichen Wissens ist dann Thema im Beitrag **„Input-Output-Relationen – Zu den Voraussetzungen einer Effizienzmessung medizinischer Fakultäten anhand von Drittmitteln“** (Hartstein) von Judith Hartstein. Der Beitrag widmet sich der Frage, ob und wie die flankierende Finanzierung medizinischer Forschung durch Drittmittel in Effizienzbetrachtungen einbezogen werden kann und sollte.

Den Abschluss der Forschungsbeiträge in diesem Band bildet der Beitrag **„Qualitätsmaße zur Evaluierung medizinischer Forschung: Erkenntnisse und Empfehlungen aus dem Projekt QuaMedFo“** (Peters und Biesenbender) von Prof. Dr. Isabella Peters und Dr. Sophie Biesenbender. Sie reflektieren den Forschungsprozess im gut dreijährigen Projekt QuaMedFo, welches in seiner multidisziplinären Konstellation sehr erfolgreich war, in empirischen Fragen jedoch einige Herausforderungen überwinden musste. Im Beitrag sind Lektionen zusammengetragen, die für Praktiker*innen in der Forschungsevaluation sowie für Forschende mit ähnlichen Vorhaben hilfreich sein können.

Ressourcen aus dem Projektkontext QuaMedFo

Im Laufe des Projekts, dessen Abschluss dieser Sammelband bildet, sind Ressourcen erarbeitet worden, welche anderen Forschenden zur Verfügung stehen und weiter genutzt werden können und sollen. In diesem letzten Teil des Buches haben wir deshalb Kurzbeschreibungen folgender Ressourcen beigefügt:

- QuaMedFo-Klassifikation wissenschaftlicher Teildisziplinen der Medizin, welche übergreifend in den Beiträgen verwendet wurde (s. QuaMedFo-Fächerklassifikation in diesem Band),
- Konzept und Datenzugangsweg zur QuaMedFo-Umfrage unter medizinischen und administrativen Beschäftigten an medizinischen Fakultäten in Deutschland (s. Datensatz für Sekundäranalysen: Befragung von wissenschaftlichen und administrativen Beschäftigten an medizinischen Fakultäten im Projekt QuaMedFo in diesem Band),
- Literaturempfehlungen aus dem Projektkontext (s. Literaturempfehlungen aus dem Projektkontext in diesem Band).

Ein Stichwortverzeichnis im Anhang schließlich ermöglicht die punktuelle und gezielte Vertiefung bestimmter Themen und Sachverhalte rund um die Themen Evaluation und Qualitätsmessung in der medizinischen Forschung.

Judith Hartstein
Sophie Biesenbender

Inhaltsverzeichnis

Verantwortungsvolle Wissenschaftsbewertung in der Medizin – was zählt?	1
Christoph Herrmann-Lingen	
Dimensionen und Qualitätsmaße zur Evaluation medizinischer Forschung – Kontext und Ansätze des Projekts QuaMedFo	19
Sophie Biesenbender	
Publikations- und Rezeptionsanalyse von drei medizinischen Fakultäten unter Berücksichtigung der LOM	43
Valeria Aman	
Altmetrics zur Evaluation medizinischer Forschung in Deutschland	65
Steffen Lemke, Anne Witthake und Isabella Peters	
Welche Charakteristika weist die in Leitlinien zitierte Literatur auf? – Eine Fallstudie anhand einer S2e- und einer S3-Leitlinie	85
Christopher Traylor und Valeria Aman	
Nutzung von Patentdaten zur Erfassung der wirtschaftlichen Verwertung von Forschung	97
Klaus Lippert und Konrad U. Förstner	
Input-Output-Relationen – Zu den Voraussetzungen einer Effizienzmessung medizinischer Fakultäten anhand von Drittmitteln	109
Judith Hartstein	

Qualitätsmaße zur Evaluierung medizinischer Forschung – Erkenntnisse und Empfehlungen aus dem Projekt QuaMedFo	129
Isabella Peters und Sophie Biesenbender	
QuaMedFo-Fächerklassifikation	143
Literaturempfehlungen aus dem Projektkontext	149
Stichwortverzeichnis	151

Herausgeberinnen- und Autor*innenverzeichnis

Über die Herausgeberinnen

Sophie Biesenbender ist promoviert in Politik- und Verwaltungswissenschaften. Sie forscht und arbeitet seit Oktober 2012 am DZHW (bzw. der Vorgängerorganisation, dem Institut für Forschungsinformation und Qualitätssicherung – iFQ) in Berlin mit einem Fokus auf die Themen Forschungsinformationen und Indikatoren. Zwischen 2020 und Anfang 2022 war sie kommissarische Leiterin der Abteilung „Forschungssystem und Wissenschaftsdynamik“ des DZHW. Seit 2022 ist sie Leiterin der Geschäftsstelle der Kommission für Forschungsinformationen in Deutschland (KFID).

Judith Hartstein forscht seit 2016 am DZHW zu Bewertungspraktiken in der Wissenschaft und zu digitalen Infrastrukturen. Seit 2019 promoviert sie an der Humboldt-Universität zu Berlin im Fach Soziologie. Zuvor hat sie einen M.A. in Wissenschaftsforschung an der Humboldt-Universität erworben sowie einen B.Sc. in Wirtschaftsmathematik an der Hochschule für Technik und Wirtschaft Berlin.

Autor*innenverzeichnis

Valeria Aman Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Berlin, Deutschland

Sophie Biesenbender Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Berlin, Deutschland

Konrad U. Förstner ZB MED – Informationszentrum Lebenswissenschaften,
Köln, Deutschland;
Institut für Informationswissenschaft, TH Köln, Köln, Deutschland

Judith Hartstein Robert K. Merton Zentrum für Wissenschaftsforschung,
Humboldt-Universität zu Berlin, Berlin, Deutschland;
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW),
Berlin, Deutschland

Prof. Dr. Christoph Herrmann-Lingen Klinik für Psychosomatische Medizin
und Psychotherapie, Universitätsmedizin Göttingen – Georg-August-Universität,
Göttingen, Deutschland

Steffen Lemke ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel,
Deutschland;
Christian-Albrechts-Universität zu Kiel, Kiel, Deutschland

Dr. Klaus Lippert ZB MED – Informationszentrum Lebenswissenschaften,
Köln, Deutschland

Isabella Peters ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel,
Deutschland;
Christian-Albrechts-Universität zu Kiel, Kiel, Deutschland

Christopher Traylor Klinik für Psychosomatische Medizin und
Psychotherapie, Universitätsmedizin Göttingen – Georg-August-Universität,
Göttingen, Deutschland

Anne Witthake ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel,
Deutschland



Verantwortungsvolle Wissenschaftsbewertung in der Medizin – was zählt?

Christoph Herrmann-Lingen

1 Einleitung

Der vorliegende Beitrag betrachtet aus der Sicht eines Praktikers die Zielsetzungen und Probleme der Evaluation medizinischer Forschung und der hierauf basierenden Leistungsorientierten Mittelvergabe (LOM) bzw. der Indikatorgestützten Mittelallokation (IMA). Nach einer allgemeinen Einführung ins Thema stellt er, insbesondere unter Bezugnahme auf eine aktuelle Stellungnahme der Deutschen Forschungsgemeinschaft (DFG), methodische Probleme der Publikationsbewertung als eines zentralen Leistungsindicators und unerwünschte Effekte bibliometrisch definierter Anreiz- und Steuerungssysteme dar. In einem Exkurs wird dabei auch auf die durch extrinsische Incentivierungen erzeugten Interessenkonflikte und ihre potenziell schädlichen Effekte auf das Wissenschaftssystem eingegangen. Im zweiten Teil werden Probleme aktueller Anreiz- und Incentivierungssysteme im spezifischen Kontext medizinischer Forschung skizziert. Er geht aus von einem vom Autor federführend mitverfassten Positionspapier der Arbeitsgemeinschaft Wissenschaftlicher Medizinischer Fachgesellschaften (AWMF) hervor und bezieht Befunde des vom Bundesministerium für Bildung und Forschung geförderten Verbundprojekts QuaMedFo, die im vorliegenden Sammelband zusammengetragen werden, sowie eine aktuelle Stellungnahme des Medizinischen Fakultätentags (MFT) mit ein. Dabei werden sowohl Besonderheiten der Bewertung medizinischer Publikationen, etwa klinischer Leitlinien als

C. Herrmann-Lingen (✉)

Klinik für Psychosomatische Medizin und Psychotherapie, Universitätsmedizin
Göttingen – Georg-August-Universität, Göttingen, Deutschland

E-Mail: cherrma@gwdg.de

© Der/die Autor(en) 2024

S. Biesenbender und J. Hartstein (Hrsg.), *Qualitätsmessung als Prisma*, Higher
Education Research and Science Studies,

https://doi.org/10.1007/978-3-658-43683-4_1

auch mögliche Indikatoren diskutiert, die sich auf Aspekte von Wissenstransfer und Drittmiteleinwerbungen stützen, hinsichtlich Aussagekraft und möglicher Steuerungswirkung erörtert. Abschließend werden Anregungen für eine Weiterentwicklung wissenschaftsadäquater und multidimensionaler Evaluations- und Anreizsysteme formuliert.

2 Leistungsbewertung und Incentivierung medizinischer Forschung

Die Bewertung wissenschaftlicher Leistung in der Medizin verfolgt unterschiedliche Zielsetzungen: Sie dient der nachträglichen Überprüfung des zweckmäßigen Einsatzes öffentlicher Forschungsmittel, der prospektiven Entscheidung über die Förderung von Forschungsprojekten oder über strategische Schwerpunktsetzungen an Forschungseinrichtungen sowie dem Auswahlprozess für akademische Positionen. Sie ergänzt idealtypischerweise die in der Regel hohe intrinsische Motivation der Forschenden durch externes Feedback und ggfs. durch hierauf basierende materielle Anreize, etwa im Rahmen der LOM bzw. IMA (Medizinischer Fakultätentag, 2022).

Sie wendet dabei derzeit vorwiegend eine ursprünglich aus den Naturwissenschaften bzw. der Ökonomie stammende Methodik auf wissenschaftliche Prozesse und Ergebnisse an, was vielfach kritisiert wurde. So spricht Binswanger (2011) von einem künstlichen Wettbewerb ohne Markt. Typische Merkmale von Märkten wie z. B. Transparenz und eine klare Preisbildung fehlen aber in der Forschung. Anders als in der Ökonomie, in der es um Profitmaximierung einzelner Akteure geht, ist Wissenschaft zudem geprägt durch die Produktion öffentlicher Güter und durch eine grundlegende Ungewissheit bzgl. der Ergebnisse (Osterloh et al., 2015). Binswanger (2015) verweist auf den durch Evaluationen und Incentivierungen erzeugten Publikationsdruck sowie die hierdurch erzeugten Fehlanreize, möglichst vieles aber ggfs. Unsinniges zu publizieren. So werde es immer schwieriger, in der wachsenden Masse an Publikationen wirklich relevante Arbeiten zu identifizieren.

Die LOM spielt mittlerweile trotz dieser mahnenden Worte eine feste Rolle in der Ressourcenallokation an medizinischen Fakultäten. Dabei ist bemerkenswert, dass die damals vorgeschlagenen Indikatoren, die von der Deutschen Forschungsgemeinschaft (Deutsche Forschungsgemeinschaft, 2004) explizit einfach gewählt wurden und nicht als Patentrezepte verstanden werden sollten, noch nach fast 20 Jahren als quasi Fixgrößen der Evaluation betrachtet werden. Dabei war von der DFG intendiert worden, sie als lediglich vorläufige Indikatoren für

einen Prozess zur Entwicklung besser geeigneter Metriken zur Diskussion zu stellen. Dass an der damaligen DFG-Empfehlung bis heute weitgehend festgehalten wird, verwundert umso mehr als völlig ungewiss ist, ob eine hierauf basierende LOM überhaupt die Qualität und Effizienz medizinischer Forschung erhöht (Krempkow et al., 2013), zumal auch unerwünschte Effekte der einseitigen Betonung von Journal Impact Factor (JIF) und Drittmittelwerbungen in der DFG-Empfehlung von 2004 bereits seit längerem bekannt sind und selbst von der DFG mittlerweile kritisch konstatiert werden (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022).

Vor diesem Hintergrund steht eine verantwortungsvolle Wissenschaftsbewertung, wie sie in diesem Zusammenhang von der Deutschen Forschungsgemeinschaft (2022) aktuell gefordert wird, vor mehreren grundsätzlichen Fragen:

1. Kann die Qualität der Forschung überhaupt durch quantitative Indikatoren bewertet werden?
2. Wenn quantifizierende Metriken herangezogen werden sollen, worauf ist hierbei zu achten? Und:
3. Welche Effekte ergeben sich aus den hierauf aufbauenden Anreizsystemen?

Diesen Fragen wird im Weiteren zunächst nachgegangen.

3 Einige methodische Probleme der Publikationsbewertung

Die Deutsche Forschungsgemeinschaft (2022) weist darauf hin, dass es kein eigentliches Ziel wissenschaftlichen Publizierens sei, eine Begutachtungsgrundlage für Zuwendungs- und Personalfinanzierungssysteme zu schaffen. Eigentliche Ziele des wissenschaftlichen Publizierens seien vielmehr die Bekanntmachung, Qualitätssicherung, Dokumentation und Zuschreibung von Urheberschaft. Insofern handelt es sich bei einer Verwendung von Publikationen für Evaluationszwecke um eine Zweckentfremdung.

Die etablierten Metriken sind aus Sicht der DFG in Bezug auf das zu vermittelnde Bild der Wirkung von Forschung unvollständig, da sie völlig von den konkreten Inhalten abstrahieren. Die Bewertung von Publikationen nach der Reputation des Verlags oder dem Impact-Faktor einer Zeitschrift könne sogar gänzlich ohne deren inhaltliche Rezeption erfolgen.

Oft bestünden übertriebene Erwartungen an die Aussagekraft, Gerechtigkeit und Objektivität bibliometrischer Indikatoren. Diese seien aber leicht zu manipulieren. Die Zahl der Zitationen einer Publikation deute zudem nicht automatisch auf ihre Güte hin, sondern sei durch diverse konfundierende Variablen wie Zugänglichkeit und Länge der Publikation, Artikeltyp, Autorennamen, fachspezifische Publikations- und Diskussionskulturen etc. beeinflusst. Der Journal Impact Factor „erbe“ die Probleme der zitationsbasierten Publikationsbewertung und bilde zudem als Mittelwert die breite Streuung der Zitation einzelner Artikel aus derselben Zeitschrift nicht ab.

Wie bei allen Kennzahlssystemen drohe auch bei bibliometrischen Indikatoren das Maß zum Ziel zu werden, insbesondere wenn sie als Grundlage von Zuwendungssystemen eingesetzt würden. Wenn aber eine Metrik zum Ziel werde, höre sie auf, eine gute Metrik zu sein.

Schließlich gebe es Fachgebiete, die sich grundsätzlich nicht zur Verwendung von Metriken zum Vergleich ihres wissenschaftlichen Outputs eignen, etwa hochkollaborative Forschungsfelder wie z. B. die Epidemiologie. Metriken eignen sich auch nicht für Vergleiche über Wissenschaftsgebiete hinweg und sollten hier nicht zum Einsatz kommen.

Kritisch befassen sich die Autor*innen des DFG-Papiers mit der Rolle kommerzieller Anbieter von Metriken und Publikationsplattformen. Sie beobachten Monopolisierungstendenzen auf dem wissenschaftlichen Publikations-, Datenbanken- und Softwaremarkt. Hingewiesen wird auf das *Bias*, das durch elektronische Literatur-Suchsysteme zwangsläufig erzeugt wird, deren Nutzung freilich angesichts der Publikationsflut zwangsläufig erforderlich ist. Neben der Bindung an digitale Tools eines Anbieters können Wissenschaftler*innen Eingriffen in ihre informationelle Selbstbestimmung durch die Anbieter von Forschungs- bzw. Publikationsdienstleistungen ausgesetzt sein. Die Steuerung der Wissenschaft durch von kommerziellen Anbietern erhobene Daten könne daher grundsätzlich kritisch gesehen werden (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022).

Angesichts der erheblichen methodischen Probleme der quantifizierenden Wissenschaftsbewertung entlang der derzeit geläufigen Metriken haben zahlreiche internationale Verbände und Organisationen, etwa im Rahmen der San Francisco Declaration on Research Assessment aus dem Jahr 2012,¹ in den letzten Jahren das vorherrschende Evaluationssystem als ungeeignet kritisiert. Neben den methodischen Mängeln spielen für diese Einschätzung die resultierenden negativen Steuerungseffekte eine entscheidende Rolle. So wird der Ruf nach

¹ <https://sfдора.org/>

einer verantwortungsvollen, also wissenschaftsadäquaten und gesellschaftlich zielführenden Wissenschaftsbewertung und einer Reform der hierauf basierenden Anreizsysteme zunehmend lauter.

4 Unerwünschte Wirkungen bibliometrisch definierter Anreiz- und Steuerungssysteme

Negative Konsequenzen der LOM und anderer auf quantifizierenden Metriken basierenden Anreizsysteme und Entscheidungskulturen, etwa im Rahmen von Karriereentscheidungen und Förderprogrammen, werden seit vielen Jahren kritisch kommentiert. Die Deutsche Forschungsgemeinschaft (2022) benennt hierzu aktuell eine Reihe problematischer Punkte, die sich wie folgt zusammenfassen lassen:

Auf der Ebene von Individuen liefen die Anreize den Standards guter wissenschaftlicher Praxis zuwider, indem sie u. a. die Strategie begünstigten, Manuskripte wiederholt bei Fachjournalen abgestufter Reputation (insbes. nach Journal Impact Factor) einzureichen und damit lange Verzögerungen durch deren Begutachtungs- und Entscheidungsprozesse in Kauf zu nehmen. Forschende würden zur Aufspaltung von Veröffentlichungen in kleinstmögliche Einzelpublikationen und zur Mehrfachpublikation identischer Ergebnisse (mit zunehmender Informationsüberflutung der Rezipient*innen) verführt. Die Marktlogik der Anreizsysteme erzeuge die Gefahr, dass Fragestellung und Dateninterpretation hinsichtlich erwünschter oder erwartbarer Ergebnisse verzerrt und weniger „vermarktungsfähige“ Hypothesen ausgeschlossen würden. Wissenschaftler*innen würden verleitet, dem Mainstream zu folgen, abweichende Meinungen zu unterschlagen, unerwünschte Messwerte wegzulassen oder sogar passende zu erfinden. Auch Selbstzitationen und die Vergabe oder Einforderung von Autor*innenpositionen richte sich neben inhaltlichen Gesichtspunkten zunehmend am Evaluationssystem aus. Insgesamt entstehe der Eindruck, dass sich unter dem Einfluss wirkmächtiger Sachzwänge und Anreize das Publikationswesen auf eine an Surrogaten statt an Inhalten orientierte Wissenschaftsbewertung ausrichte. Die durch bibliometrisch gestützte Wissenschaftsbewertung gesetzten Anreize seien insofern problematisch und könnten eine wissenschaftsadäquate Entwicklung des Publikationswesens und des Wissenschaftssystems insgesamt verhindern.

Wissenschaftler*innen, die der Kennzahlen-orientierten Logik nicht folgten, gerieten rasch ins Hintertreffen, ggfs. einschließlich der von ihnen geleiteten Einrichtungen bzw. der ihnen anvertrauten Nachwuchswissenschaftler*innen (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022).

Hieraus resultieren – nicht nur aus Sicht der DFG – mehrere relevante systemische Folgeprobleme:

Zu nennen ist hier zum einen die vielfach beklagte Replikationskrise in den Lebenswissenschaften (z. B. Begley und Ioannidis 2015), zumal das Bewertungssystem zu wenige Anreize setzt, negative Studienergebnisse zu publizieren (mit der Folge eines *publication bias*) oder Replikationsstudien (mangels „Originalität“, Förderung und erwartbarer „Belohnung“) auch nur durchzuführen. So schätzen Chalmers und Glasziou (2009), dass mehr als 85 % der weltweiten biomedizinischen Forschungsmittel verschwendet sein könnten, indem irrelevante Fragestellungen untersucht, die vorhandene Evidenz nicht angemessen bei der Studienplanung berücksichtigt, ungeeignete Methoden eingesetzt, Protokolle nicht ausreichend beschrieben, Forschungsergebnisse nicht oder unvollständig publiziert sowie Ergebnisse nicht zutreffend interpretiert und in den Kontext der vorhandenen Literatur eingeordnet werden.

5 Exkurs: Anreizsysteme und Interessenkonflikte im Publikationswesen

Indem quantifizierende Anreizsysteme in den meisten Fällen sekundäre Interessen der Forschenden nach Reputation oder finanziellen Vorteilen bedienen, geraten sie in Konflikt mit primären Erkenntnisinteressen und führen somit zu Interessenkonflikten, deren Bedeutung im Kontext industrieller Zuwendungen an die Wissenschaft breit diskutiert wird (z. B. AWMF, 2017; ICMJE, 2021),² da nahe liegt und empirisch gezeigt werden konnte, dass solche Interessenkonflikte auch die publizierten Forschungsergebnisse substanziell beeinflussen können (Lundh et al., 2017). Vor diesem Hintergrund ist es erstaunlich, dass Interessenkonflikte und ihre Auswirkungen im Kontext staatlicher und universitärer Anreizsysteme für die Forschung aber nach wie vor kaum explizit problematisiert werden.

Dabei lassen sich die etwa von der Deutschen Forschungsgemeinschaft (2022) konstatierten Nebenwirkungen bibliometrisch begründeter Anreizsysteme

² So gibt die AWMF Empfehlungen für den Umgang mit Interessenkonflikten bei der Aktivitäten wissenschaftlicher Fachgesellschaften, etwa im Kontext der Leitlinienentwicklung, und das ICMJE definiert den angemessenen Umgang mit Interessenkonflikten bei Publikationen in medizinischen Fachzeitschriften.

im Wissenschafts- und Publikationssystem zu großen Teilen über die hierdurch erzeugten Interessenkonflikte erklären.

Probleme durch Predatory Publishing und predatory-like Journals

Eine Fehlentwicklung stellen die von der DFG zurecht verurteilten Raubveröffentlichungen („predatory publishing“; Moher et al., 2017; Gallent Torres 2022) dar. Hier zeigt sich auch eine Kehrseite der grundsätzlich im Interesse der Forschungstransparenz zu begrüßenden Open-Access-Bewegung:

In klassischen subskriptionsbasierten Journalen werben primär Forschende mit möglichst relevanten und methodisch anspruchsvollen Forschungsarbeiten um die Akzeptanz ihrer Arbeiten durch kritische Gutachter*innen. Da die Herausgeber*innen nur begrenzten Raum für Publikationen zur Verfügung haben, mit denen sie ihrerseits um Leser*innen bzw. Abonnent*innen werben, führt die Begutachtung zu einer – wenn auch wie unten dargelegt methodisch fragwürdigen – „Besten-Auslese“. Dagegen werben Open-Access-Journale mit finanziellem Eigeninteresse und teilweise sehr massiv um Forschungsbeiträge, die dann nach Stückzahl zu bezahlen sind. Bei grundsätzlich nicht begrenzten Online-Publikationsmöglichkeiten legen finanzielle Interessen von Verlagen auf der einen und durch quantifizierende Evaluationen auf Autor*innen einwirkende Anreize auf der anderen Seite eine weniger strenge Auswahl zu publizierender Artikel und die Veröffentlichung möglichst vieler – ggfs. auch qualitativ weniger guter – Artikel nahe. Diese Kollusion kommt im Sinne einer Win-Win-Situation sowohl den materiellen Interessen der Verlage als auch den sekundären Interessen der Autor*innen entgegen.

Dies betrifft nicht ausschließlich und möglicherweise nicht einmal in erster Linie die ganz offensichtlichen Raubjournale, sondern zunehmend auch im Grunde seriöse Open-Access-Journale, die etwa über die zunehmende Herausgabe thematischer Special Issues (Amrein, 2022) gegen Zahlung relativ hoher Publikationsgebühren eine sehr hohe Annahmewahrscheinlichkeit und zugleich einen in der LOM honorierten respektablen Impact-Faktor bieten. Die Aufnahme des Publikationsportals Frontiers in eine verbreitete Liste potenzieller Raubverlage³ führte zu einer kontroversen Debatte über diese Art des Publizierens.⁴ Diese Debatte wurde u. a. in Nature geführt, deren Verlag auch Frontiers herausgibt und somit auch in der Debatte nicht frei von Interessenkonflikten ist. Einem weiteren in der Diskussion befindlichen Portal (MDPI) wurden in einer aktuellen Untersuchung (Oviedo-García, 2021) auch empirisch Merkmale des *Predatory Publishing* bescheinigt.

Auch der klassische Peer-Review-Prozess anerkannter wissenschaftlicher Journale ist nur bedingt in der Lage, wissenschaftliche Qualität zu bewerten. So zeigt sich immer wieder die sehr niedrige Interrater-Reliabilität von Artikelbewertungen durch Reviewer (Walker & Rocha da Silva, 2015). Gutachtenden gelingt es nicht immer, gefälschte oder erfundene Ergebnisse im Begutachtungsprozess zu identifizieren, was selbst in Spitzenjournalen immer wieder zu Retractionen bereits publizierter Artikel führt. Auch hier spielen letztlich Interessenkonflikte

³ <https://bealllist.net/>

⁴ <https://www.nature.com/articles/526613f>

unterschiedlicher Art eine gewisse Rolle, denn auch die Gutachtenden sind nicht frei von Eigeninteressen und Voreingenommenheiten, was zu Verzerrungen von Bewertungen oder zur Aufforderung zur Zitation eigener Werke in den begutachteten Artikeln führen kann. Faktisch besteht ein Interessenkonflikt allein schon darin, dass die i. d. R. kostenlos zu erstellenden Gutachten mit der Arbeit an eigenen Forschungsprojekten um begrenzte Zeitressourcen konkurrieren und so nicht immer die notwendige Sorgfalt erfahren. Auch betrügerische Begutachtungen durch die Autor*innen selbst, die sich unter Pseudonymen und mit teilweise gefälschten E-Mail-Adressen den Herausgeber*innen von Fachzeitschriften als Gutachtende empfahlen und dann für die Annahme ihrer Beiträge plädierten, wurden wiederholt beschrieben (Ferguson et al., 2014). Die Deutsche Forschungsgemeinschaft (2022) verweist zudem auf sogenannte „paper mills“, Netzwerke wissenschaftlich erfahrener Ghostwriter, die gegen Bezahlung Fake-Artikel verfassen. Diese reichen Wissenschaftler*innen dann unter eigenem Namen bei durchaus seriösen Journalen ein, wo sie z. T. erst nach Publikation als Fälschungen erkannt werden und teilweise zu regelrechten Retraktionswellen geführt haben.

6 Besonderheiten der Forschungsevaluation in der Medizin

Die genannten Probleme mit quantifizierenden Evaluationen erlangen insbesondere in der Medizin eine problematische Bedeutung, wenn sie, wie von der DFG konstatiert, Anreize gegen die gute wissenschaftliche Praxis setzen und damit den medizinischen Fortschritt und letztlich die Gesundheit von Menschen beeinträchtigen.

Die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) hat daher bereits 2014 in einem Positionspapier die gängige Evaluationspraxis medizinischer Forschung kritisiert, da sie u. a. vielfach auf unklaren Zielvorstellungen basiere und so zu Fehlsteuerungen führen könne (Herrmann-Lingen et al., 2014).

Der wichtigste Parameter der Evaluation medizinischer Forschung sei letztlich die Bedeutung von Forschungsleistungen für die Weiterentwicklung der wissenschaftlichen Medizin bzw. eines spezifischen medizinischen Fachgebiets.

Hervorgehoben wird die Bedeutung einer a priori klar kommunizierten Zielsetzung der Bewertung und der aus ihr resultierenden Konsequenzen, die auch über die Art, Intensität und Häufigkeiten der Evaluationen entscheiden müsse.

Neben qualitativen, also inhaltlichen Bewertungen einzelner Forschungsbeiträge oder ganzer Forschungsprojekte bzw. Institutionen durch z. B. informierte Peer-Review-Verfahren können aus Sicht der AWMF quantifizierende Bewertungen vor allem im Rahmen formativer Evaluationen und nur unter bestimmten Voraussetzungen auch im Rahmen summativer Evaluationen eingesetzt werden.

Die unterschiedlichen Aspekte medizinischer Forschungsleistung erforderten dabei ein mehrdimensionales Vorgehen und ein ausreichend breites Indikatoren-spektrum.

7 Impact medizinischer Forschung: Mehr als der Journal Impact Factor

Insbesondere der Aspekt des Impacts medizinischer Forschungsleistung muss hier deutlich weiter gefasst werden als die vorherrschende Bewertung von Zeitschriftenaufsätzen anhand des Journal Impact Factors der jeweiligen Zeitschrift. So sollten auch Publikationsformate jenseits der Zeitschriftenpublikation Berücksichtigung finden.

Aman kann allerdings in diesem Band zeigen, dass – mutmaßlich unter dem Einfluss der LOM – die Zahl und der relative Anteil der Journalbeiträge gegenüber anderen Publikationsformen an den untersuchten Pilotfakultäten über die Jahre zugenommen hat. Dies dürfte mutmaßlich zur Benachteiligung solcher medizinischen Fächer beitragen, die klassischerweise auch oder sogar überwiegend in anderen Formaten (etwa Buchbeiträgen oder Monographien) publizieren. Bei seit Einführung der LOM insgesamt geringerer (oder fehlender) Bewertung dieser Formate nimmt ihr relativer Wert in der LOM-Logik weiter ab, wenn bei gedeckelten LOM-Mitteln immer mehr Journalpublikationen den Gegenwert der einzelnen Publikation sinken lassen, was schwerlich ohne Qualitätsverlust durch eine (quantitative) Produktivitätssteigerung bei der Verfassung von Monographien kompensiert werden kann. Fächerunterschiede konnte Aman (in diesem Band) auch hinsichtlich der Rezeption von Forschungsarbeiten belegen, was der seit Jahren u. a. von der AWMF geforderten Berücksichtigung einer Feldnormierung für die quantifizierende Publikationsbewertung Nachdruck verleiht.

Publikationen sollten aus Sicht der AWMF nicht nur auf der Basis ihrer Zitationen durch Zeitschriftenartikel, sondern z. B. auch auf Grundlage ihres tatsächlichen Einflusses auf die klinische Medizin bewertet werden. So kann etwa die Zitation durch hochwertige medizinische Leitlinien als Maß für den Transfer in die medizinische Anwendung berücksichtigt werden.

Im Rahmen des QuaMedFo-Projekts, das auch diesem Sammelband zugrunde liegt (Biesenbender in diesem Band) konnten Traylor und Herrmann-Lingen (2023) kürzlich zeigen, dass die Zitation durch deutsche Leitlinien der höchsten Entwicklungsstufen S2e und S3 weder hinsichtlich des Zeitfensters, dem die zitierten Arbeiten entstammen noch hinsichtlich der Zitationshäufigkeit bestimmter Journale der Logik des Journal Impact Factors folgt und somit als weitgehend von diesem unabhängiger Indikator zu werten ist. Die Fallstudie von Traylor und Aman (in diesem Band) erweitert diese Analysen, indem sie auf Artekelebene zeigt, dass die Zitation durch zwei exemplarisch ausgewählte Leitlinien beim Vergleich mit anderen jeweils im selben Journal und Jahr publizierten Artikeln mit einer deutlich und signifikant überdurchschnittlichen Zitationszahl durch andere Journalbeiträge einhergeht, was ebenfalls unterstreicht, dass nicht der JIF eines Journals, sondern die Resonanz einzelner Forschungsarbeiten entscheidend für ihren Impact in der Praxis ist.

Aber auch andere Beiträge zum Praxistransfer wie z. B. die Ausrichtung von Fachtagungen und industrieunabhängigen Fortbildungsveranstaltungen oder die Publikation von Lehrbüchern sollten aus Sicht der AWMF als Leistungsmerkmale bewertet werden. Zu berücksichtigen sind ferner Aufgaben in der Steuerung bzw. Selbstverwaltung der Wissenschaft, ihrer Organisationen und Fachgesellschaften sowie das Engagement in der Nachwuchsförderung.

8 Drittmittel als Indikatoren

Bei der Bewertung von Drittmitteln sind laut AWMF neben der Fördersumme das Vorhandensein eines kompetitiven Ausschreibungsverfahrens und der aus der Förderung resultierende wissenschaftliche Output zu berücksichtigen.

Unklar bleibt freilich erstens, wie das Drittmittelaufkommen am sinnvollsten in die Evaluation einfließen kann. Wie Hartstein (in diesem Band) ausführt, stellt schon eine einheitliche Verfügbarkeit geeigneter Drittmitteldaten an den Fakultäten eine Herausforderung dar.

Zweitens ist eine Analyse des Zusammenhangs zwischen Personal-„Input“ und Publikations-Output auf die Verfügbarkeit ausreichend differenzierter Angaben zum wissenschaftlichen Personal angewiesen, was an medizinischen Fakultäten zu Problemen führen kann, wenn im Rahmen des Integrationsmodells nicht eindeutig zwischen wissenschaftlichem und klinischem Personal unterschieden werden kann.

Drittens zeigt sich, dass es zwischen den Fächergruppen signifikante Unterschiede im Zusammenhang zwischen der Zahl der Personalstellen und den

Drittmittelausgaben gibt. Hierfür könnten u. a. unterschiedliche faktische Belastungen des wissenschaftlichen Personals durch Lehre und Krankenversorgung eine Rolle spielen: So kann einerseits wissenschaftliches Personal in unterschiedlichem Umfang in der Krankenversorgung gebunden sein, andererseits kann gerade in großen Kliniken das klinische Personal durch (quasi nebenberufliche) Beteiligung an Forschung und Lehre das im engeren Sinne wissenschaftliche Personal unterstützen bzw. von Lehraufgaben entlasten.

Und viertens können die Prämissen klassischer Input–Output-Analysen, etwa die Annahme linearer Zusammenhänge, in komplexen rückgekoppelten Systemen wie sie der dynamischen Realität medizinischer Einrichtungen und ihrer Forschung entsprechen, durchaus infrage gestellt werden. Während Hartstein (in diesem Band) argumentiert, dass die moderate Korrelation zwischen Drittmittelausgaben und Drittmittelpersonal nahelegen könnte, bei der Betrachtung des Input-Output-Verhältnisses auf die Verwendung der Drittmittelausgaben als Input-Indikator zu verzichten, um eine zu komplexe Modellierung zu vermeiden, könnte dieser Ansatz zur Abbildung der realen Dynamiken unterkomplex erscheinen, wie die schematische Darstellung dieser Dynamiken in Abb. 1 dargestellt:

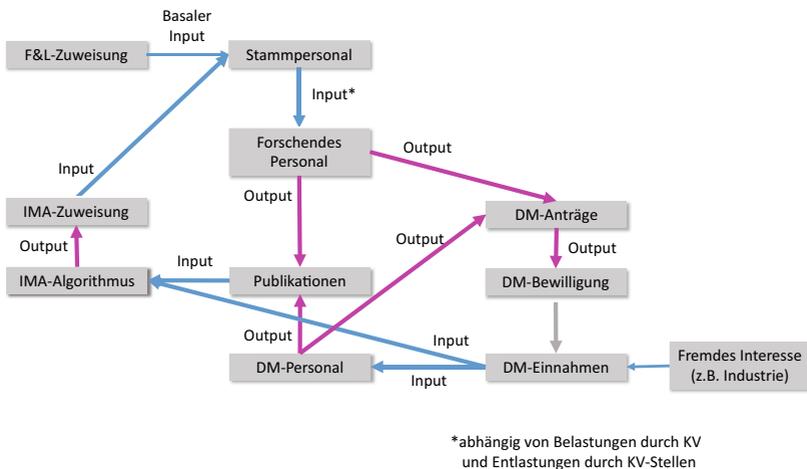


Abb. 1 Input-Output-Beziehungen in der medizinischen Forschung unter Zugrundelegung eines typischen Publikations- und Drittmittel-basierten IMA-Algorithmus. (Blaue Pfeile: Inputs; Violette Pfeile: Outputs. Zu erkennen sind mehrere z. T. überlappende Input–Output-Zirkel. Abkürzungen: DM = Drittmittel; F&L = Forschung und Lehre; IMA = Indikatorgestützte Mittelallokation; KV = Krankenversorgung)

Ein Großteil der verausgabten Drittmittel wird (als Input) für (Drittmittel-)Personal eingesetzt, das dann (als Output) seinerseits sowohl Publikationen verfassen als auch selbst Drittmittel einwerben kann. Da zudem die Drittmittelausgaben via LOM-/IMA-Algorithmus zur erhöhten Zuweisung von Planmitteln führen, nehmen sie als Inputvariable eine wichtige Rolle in einem rückgekoppelten System ein, in dem sich mehrere selbstverstärkende Kreislaufprozesse entwickeln:

Der primäre fakultäre Input in Form von Mittelzuweisungen für Forschung (und ggfs. Lehre) befähigt eine Einrichtung – beeinflusst durch die unter drittens genannten „Störeffekte“ durch das Engagement in Klinik und Lehre und abhängig von ihrer Effizienz in der Nutzung des basalen Inputs – Drittmittelanträge zu stellen (als Output), die bei Bewilligung nach kompetitivem Peer Review als qualifizierter Output anerkannt werden. Dies leuchtet ein, da hierfür u. a. (neben z. B. der Passung zu einem bestimmten Förderinteresse sowie Interessenkonflikten von Gutachtenden) die Qualität des Antrags eine Rolle spielt. Allerdings ist schon hier einschränkend zu erwähnen, dass insbesondere industrielle Drittmittelzuwendungen oft nicht als qualifizierter Output der universitären Einrichtung, sondern als Ausdruck des Eigeninteresses einer Firma zu werten sind und z. T. ja nicht einmal eine eigenständige Antragstellung als „Leistung“ voraussetzen, dennoch aber die Forschungsressourcen einer Einrichtung erhöhen.

Zugleich stellen die auf die Mittelbewilligung folgenden Prozesse, also die Bereitstellung bzw. Verausgabung der Drittmittel (insbesondere, aber keineswegs nur bei Industrie-Drittmitteln) in zweifacher Weise eine Input-Variable dar: Zum einen fließen sie in den LOM-/IMA-Algorithmus und vermittelt darüber in den Input für (mehr) Stammpersonal ein und zum anderen ermöglichen sie erst die Einstellung von drittmittelfinanziertem Personal, das dann – gute Effizienz vorausgesetzt – als Output wie oben angedeutet sowohl direkt neue Drittmittel einwerben als auch indirekt über Publikationen die LOM-Zuweisung erhöhen kann.

In diesen miteinander verbundenen Selbstverstärkungskreisen spielt jedes Element eine wichtige Rolle. Die Effizienz einer Einrichtung würde sich hier in einer umfassenderen Input-Output-Betrachtung als ein Durchschnittswert der wissenschaftlichen Arbeits-Effizienzen von Stamm- und Drittmittelpersonal (bereinigt um die o. g. Be- bzw. Entlastungsaspekte durch klinische Aufgaben bzw. klinisches Personal) ergeben. Nur gemeinsam können diese Prozesse das Geschehen an der Einrichtung angemessen abbilden. Insofern sollten zukünftig mathematische Modelle der Prozessevaluation entwickelt werden, die bei entsprechender Automatisierung eine sehr viel passgenauere Abbildung der Input-Output-Beziehungen ermöglichen würden als eine einfache Korrelationsanalyse, die

insbesondere bei rein querschnittlicher Betrachtung keine hinreichende Differenzierung nach Ursache und Wirkung ermöglicht.

Geeignete Methoden sollten zudem entwickelt werden, um bei interdisziplinären Publikationen oder Mitteleinwerbungen sowohl dem individuellen Beitrag als auch der Gruppenleistung insgesamt (Mehrwert durch Vernetzung, Koordination etc.) Rechnung zu tragen (Herrmann-Lingen et al., 2014).

9 Einige Konsequenzen für Wissenschaft und Forschungsförderung

Die Frage, welche Maßstäbe letztlich in welchem Kontext für die Bewertung medizinischer Forschung in welchem Verhältnis zu gewichten sind, bleibt dabei von den verantwortlichen Akteuren auszuhandeln. Dabei ist wie schon 2014 von der AWMF auch in jüngerer Zeit vielfach betont worden, dass ein breiteres Indikatorenspektrum benötigt wird, das über Maße für Zeitschriftenartikel und Drittmiteleinwerbungen hinausgeht (European University Association, 2022; Medizinischer Fakultätentag, 2022). Hier können und sollten schon aus methodischen Gründen an die Evaluation einzelner Wissenschaftler*innen – etwa im Rahmen von Habilitations- und Berufungsverfahren – andere Kriterien angelegt werden als an die Bewertung ganzer Kliniken, Institute oder Fakultäten.

So könnten zur Evaluation von Einzelpersonen neben individuellen qualitativen Begutachtungen mit den genannten Limitationen auch quantitative Metriken wie Zitationen und kompetitive Drittmiteleinwerbungen benutzt werden. Ergänzt werden sollten sie aber je nach Fachgebiet um etwa die Mitwirkung an und Zitation durch hochwertige Leitlinien, die Anmeldung von Patenten, Aktivitäten zur Förderung der Transparenz in der Wissenschaft (durch z. B. Präregistrierung von Studien und Data Sharing) sowie die Herausgabe von Lehrbüchern oder Sammelbänden. Auch die Ausrichtung von wissenschaftlichen Fachtagungen oder industrieunabhängigen Fortbildungsveranstaltungen und die Mitwirkung in Editorial Boards seriöser Fachzeitschriften, in Fachkollegien oder Gutachtergremien bei Forschungsförderorganisationen oder in der Leitung von Fachgesellschaften sollten Berücksichtigung finden (Herrmann-Lingen et al., 2014). Zusätzlich erscheinen in jüngerer Zeit Aktivitäten zur Verbreitung wissenschaftlicher Erkenntnisse in sozialen Medien oder die Resonanz in diesen Medien als möglichen Indikatoren interessant. Wichtig ist aber die Klärung, wie diese Indikatoren für den praktischen Einsatz mit jeweils vertretbarem Aufwand gewonnen werden können.

Vor diesem Hintergrund erfolgten die Analysen des Projekts QuaMedFo unter Zugrundelegung eines weitgefassten Impact-Begriffs und in Verbindung mit Erwägungen der praktischen Machbarkeit und Anwendbarkeit in konkreten (Evaluations-)Kontexten. Vorschläge zur Berücksichtigung von Patenten machen Lippert und Förstner (in diesem Band) mit ihren Befunden zur automatisierten Zuordnung von Patentanmeldungen zu konkreten Wissenschaftler*innen. Sie legen nahe, dass eine automatisierte Autor*innen-Zuordnung von Patenten vor allem beim Vergleich über Einrichtungen hinweg von Nutzen sein dürfte, während die Zuordnung bei der Bewertung von Einzelpersonen – etwa im Rahmen von Qualifizierungs- oder Berufungsverfahren – ggfs. auch einfach über deren Eigenangaben vorgenommen werden kann. Lemke et al. (in diesem Band) untersuchen die Nutzbarkeit altmetrischer Indikatoren, wobei allerdings auch deren Limitationen deutlich werden. Grundsätzlich könnte ihr Angewiesensein auf große Publikationszahlen (Lemke et al. in diesem Band) zumindest dem Einsatz bei der Bewertung von Individuen im Wege stehen. Hier besteht weiterhin erheblicher Forschungsbedarf.

Während sich die genannten Indikatoren in kumulierter Form auch für die Evaluation von Institutionen eignen dürften, sollten hier zusätzlich stärker als bisher strukturelle Maßnahmen zur Förderung des (klinisch-)wissenschaftlichen Nachwuchses (etwa im Sinne von Forschungsfreistellungen für Nachwuchswissenschaftler*innen oder Clinician-Scientist-Programmen) und deren messbare Ergebnisse (z. B. Zahl erfolgreicher Promotionen und Habilitationen, Drittmittelinwerbungen durch Nachwuchs-Wissenschaftler*innen) betrachtet werden (Herrmann-Lingen et al., 2014).

Dabei ist zu beachten, dass unterschiedliche Fächerkulturen und Belastungsprofile (etwa durch unterschiedlich ausgeprägte Verpflichtungen in Krankenversorgung und Lehre) auch zu einer differenzierten Bewertung der Evaluationsergebnisse und entsprechend angepassten Mittelzuweisungen führen müssen (Medizinischer Fakultätentag, 2022).

Die Deutsche Forschungsgemeinschaft (2022) fordert von der Wissenschaft, neue Formen der Qualitätsprüfung von Veröffentlichungen zu etablieren, zu nutzen und anzuerkennen sowie die Adressatenorientierung wissenschaftlichen Publizierens auszubauen. Den Forschungsförderorganisationen legt sie nahe, die inhaltliche Bewertung wissenschaftlichen Outputs in den Mittelpunkt zu stellen (Deutsche Forschungsgemeinschaft, 2022). Der Medizinische Fakultätentag schlägt vor, den irreführenden Begriff der „Leistungsorientierten Mittelvergabe“ (LOM) durch den passenderen Begriff der „Indikatorgestützten Mittelallokation“ (IMA) zu ersetzen, wobei die Fakultäten es noch stärker als ihre originäre Aufgabe wahrnehmen sollten, auf die Auswahl sachgerechter Indikatoren zu achten

und diese transparent zu kommunizieren (Medizinischer Fakultätentag, 2022). Für die Auswahl geeigneter Indikatoren sollte das QuaMedFo-Projekt (Biesenbender in diesem Band), aus dem dieser Sammelband hervorgegangen ist, einige empirische Befunde bereitstellen.

Allerdings ist wie oben erwähnt nach wie vor nicht geklärt, inwieweit LOM oder IMA als Incentivierungsinstrumente überhaupt ein günstiges Wirkungs-Nebenwirkungsprofil aufweisen (Medizinischer Fakultätentag, 2022) oder ob angesichts der in der Regel hohen intrinsischen Motivation der Forschenden nicht viel stärker (wie auch von der DFG gefordert) auf die inhaltliche Rückmeldung zu den geleisteten wissenschaftlichen Arbeiten und weitere nicht-quantifizierende Aspekte der Reputationszuweisung gesetzt werden sollte. Auf Fakultätsebene könnte das u. a. bedeuten, die verbreitete Post-Hoc-Honorierung stärker auf eine antragsbasierte ex-ante-Förderung aussichtsreicher Forschungsansätze umzustellen.

Letztlich sind es dabei nicht die Zahlen, die zählen, sondern die Frage wie sehr eine medizinische Forschungsleistung tatsächlich zur Weiterentwicklung der Medizin im Allgemeinen und ihres gesellschaftlichen bzw. gesundheitlichen Nutzens beiträgt. Damit kann ihre Bewertung nicht ausschließlich wissenschaftlichen Peers überlassen bleiben, sondern es wäre unter Einbeziehung weiterer Stakeholder wie etwa Patient*innen-Vertretungen und anderer politischer Akteure auch ihr Impact in der medizinischen bzw. gesellschaftlichen Praxis zu ermitteln. Und es wären sowohl geeignete Instrumente zur Erhebung und Auswertung adäquater quantifizierender Metriken als auch neue und praxistaugliche Wege der qualitativen Wissenschaftsbewertung – und eine Kombination beider – zu entwickeln und zu testen, damit ein solches multidimensionales Bewertungssystem Eingang in die Praxis finden kann.

Interessenkonflikte Der Autor berichtet, dass er in den letzten drei Jahren persönliche Vortragshonorare der Firma Novartis sowie verschiedener Kliniken, Aus- und Weiterbildungsinsitute zu psychosomatischen Themen sowie von verschiedenen Verlagen Tantiemen für die deutsche Fassung der Hospital Anxiety and Depression Scale und einige Buchbeiträge erhalten hat. Er hat – u. a. zum Thema der Evaluation medizinischer Forschung – institutionelle Forschungsförderung vom BMBF, der DFG, der EU-Kommission und der Stiftung Nachlass Grun erhalten und erhält von der Universitätsmedizin Göttingen institutionelle LOM-Zuweisungen sowie leistungsabhängige Gehaltsbestandteile. Er ist Mitglied mehrerer medizinischer Fachgesellschaften in zeitweise leitender Funktion und war bis 2021 als Präsidiumsmitglied der AWMF und als Sprecher der Kommission für Qualitätsentwicklung in Forschung und Lehre u. a. mit der Abfassungen von Publikationen zum Thema der Forschungsevaluation befasst.

Literatur

- Amrein M. (2022). Profit statt Qualität. *Neue Zürcher Zeitung* 6.11.2022, 45–46.
- AWMF. (2017). *Empfehlungen der AWMF zum Umgang mit Interessenkonflikten bei Aktivitäten wissenschaftlicher medizinischer Fachgesellschaften*. https://www.awmf.org/fileadmin/user_upload/dateien/publikationen_zu_leitlinien/20171110_AWMF_Empfehlungen_zu_Interessenkonflikten_V2.2_f.pdf.
- Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation research*, 116(1), 116–126. <https://doi.org/10.1161/CIRCRESAHA.114.303819>.
- Binswanger, M. (2011). Sinnlose Wettbewerbe behindern Wissenschaft und Forschung. *Cicero*. Verfügbar unter <https://www.cicero.de/wirtschaft/sinnlose-wettbewerbe-behindern-wissenschaft-und-forschung/41572>. Zugegriffen: 17. Aug. 2022.
- Binswanger, M. (2015). How Nonsense Became Excellence: Forcing Professors to Publish. In: Welpel, I., Wollersheim, J., Ringelhan, S., Osterloh, M. (Hrsg.), *Incentives and Performance* (S. 19–32). Springer, Cham. https://doi.org/10.1007/978-3-319-09785-5_2.
- Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *Lancet (London, England)*, 374(9683), 86–89. [https://doi.org/10.1016/S0140-6736\(09\)60329-9](https://doi.org/10.1016/S0140-6736(09)60329-9).
- Deutsche Forschungsgemeinschaft | AG Publikationswesen. (2022). *Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung*. <https://doi.org/10.5281/ZENODO.6538163>.
- Deutsche Forschungsgemeinschaft. (2004). *Empfehlungen zu einer Leistungsorientierten Mittelvergabe (LOM) an den Medizinischen Fakultäten. Stellungnahme der Senatskommission für Klinische Forschung der Deutschen Forschungsgemeinschaft* (S. 1–23). https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/stellungnahmen_papiere/2004/stellungnahme_klinische_forschung_04.pdf.
- European University Association (EUA). (2022). *Agreement on Reforming Research Assessment*. https://research-and-innovation.ec.europa.eu/document/download/bd53d467-0220-4c9d-8b63-26eb56303ba1_en.
- Ferguson, C., Marcus, A., & Oransky, I. (2014). Publishing: The peer-review scam. *Nature*, 515(7528), 480–482. <https://doi.org/10.1038/515480a>.
- Gallent, T. C. (2022). Editorial misconduct: The case of online predatory journals. *Heliyon*, 8(3), e08999. <https://doi.org/10.1016/j.heliyon.2022.e08999>.
- Herrmann-Lingen, C., Brunner, E., Hildenbrand, S., Loew, T. H., Raupach, T., Spies, C., Treede, R.-D., Vahl, C.-F., & Wenz, H.-J. (2014). Evaluation of medical research performance – position paper of the Association of the Scientific Medical Societies in Germany (AWMF). *GMS German Medical Science; 12:Doc11; ISSN 1612–3174*. <https://doi.org/10.3205/000196>.
- ICMJE. (2021). *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. https://www.icmje.org/recommendations/archives/2021_dec_urm.pdf.
- Krempkow, R., Landrock, U., Neufeld, J., & Schulz, P. (2013). *Intendierte und nicht-intendierte Effekte dezentraler Anreizsysteme am Beispiel der fakultätsinternen leistungsorientierten Mittelvergabe in der Medizin*. iFQ – Institut für Forschungsinformation

- und Qualitätssicherung. <https://www.forschungsinfo.de/projekte/GOMED/GOMED-Abschlussbericht.pdf>.
- Lundh, A., Lexchin, J., Mintzes, B., Schroll, J. B., & Bero, L. (2017). Industry sponsorship and research outcome. *The Cochrane database of systematic reviews*, 2(2), MR000033. <https://doi.org/10.1002/14651858.MR000033.pub3>.
- Medizinischer Fakultätentag. (2022). *Impulspapier: Weiterentwicklung der Indikator-gestützten Mittelallokation für die Forschung in der Hochschulmedizin*. Medizinischer Fakultätentag. https://medizinische-fakultaeten.de/wp-content/uploads/2022/10/MFT-Impulspapier_IMA_LOM_2022.pdf.
- Moher, D., Shamseer, L., Cobey, K. D., Lalu, M. M., Galipeau, J., Avey, M. T., Ahmadzai, N., Alabousi, M., Barbeau, P., Beck, A., Daniel, R., Frank, R., Ghannad, M., Hamel, C., Hersi, M., Hutton, B., Isupov, I., McGrath, T. A., McInnes, M. D. F., Page, M. J., ... Ziai, H. (2017). Stop this waste of people, animals and money. *Nature*, 549(7670), 23–25. <https://doi.org/10.1038/549023a>.
- Osterloh, M., Wollersheim, J., Ringelhan, S., & Welpel, I.M. (2015). Preface: Does Science go wrong?. In: Welpel, I., Wollersheim, J., Ringelhan, S., Osterloh, M. (Hrsg.), *Incentives and Performance* (S. v–xxii). Springer, Cham. <https://doi.org/10.1007/978-3-319-09785-5>.
- Oviedo-García, M. A. (2021). Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI). *Research Evaluation*, 30(3), 405–419. <https://doi.org/10.1093/reseval/rvab020>.
- Traylor, C., & Herrmann-Lingen, C. (2023). Does the journal impact factor reflect the impact of German medical guideline contributions? *Scientometrics*, 128(3), 1951–1962. <https://doi.org/10.1007/s11192-023-04647-z>.
- Walker, R., & Rocha da Silva, P. (2015). Emerging trends in peer review—a survey. *Frontiers in neuroscience*, 9, 169. <https://doi.org/10.3389/fnins.2015.00169>.

Herrmann-Lingen, Christoph Prof. Dr. Christoph Herrmann-Lingen ist Internist und Facharzt für Psychosomatische Medizin und Psychotherapie sowie Direktor der gleichnamigen Klinik an der Universitätsmedizin Göttingen. Sein Haupt-Forschungsschwerpunkt liegt in der Psychokardiologie. Er war Präsident der American Psychosomatic Society (2016–17) und des Deutschen Kollegiums für Psychosomatische Medizin (2018–21). Als Betroffener und langjähriger (2009–2021) Sprecher der ständigen Kommission für Qualitätsentwicklung in Forschung und Lehre der Arbeitsgemeinschaft Wissenschaftlicher Medizinischer Fachgesellschaften (AWMF) beschäftigt er sich seit 2009 auch mit Aspekten der Evaluation medizinischer Forschungs- und Lehrleistung.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Dimensionen und Qualitätsmaße zur Evaluation medizinischer Forschung – Kontext und Ansätze des Projekts QuaMedFo

Sophie Biesenbender

1 Einleitung und Ausgangslage

Das Verbundprojekt „Qualitätsmaße zur Evaluation medizinischer Forschung“ (QuaMedFo, Projektzeitraum 2019–2022)¹ war der Ausgangspunkt für den vorliegenden Sammelband „Qualitätsmessung als Prisma“, welcher verschiedene Perspektiven auf die Evaluation medizinischer Forschung in Deutschland versammelt. Gegenstand des folgenden Beitrags ist eine Beschreibung des Projekts QuaMedFo und eine Einordnung in den allgemeinen Stand der Forschung sowie die aktuelle Diskussion zum Thema datengestützte Bestandsaufnahmen und Forschungsevaluationen. Am Beispiel der medizinischen Forschung und bestehender Evaluationspraxen im deutschen Wissenschaftssystem werden strukturelle, methodische und praktische Herausforderungen bei Entwicklung adäquater Betrachtungs- und Bewertungsinstrumente skizziert.

Das Kapitel umfasst eine Kontextualisierung und Hinführung zu den im QuaMedFo-Projekt durchgeführten Analysen verschiedener Dimensionen und Datenzugänge für die Evaluation medizinischer Forschung (Aman; Hartstein;

¹ Die Förderung des Verbunds und seiner Teilprojekte erfolgte durch das Bundesministerium für Bildung und Forschung, Förderlinie: „Quantitative Wissenschaftsforschung“, Laufzeit: 01.01.2019–30.06.2022, Förderkennzeichen: 16PU17011A, 16PU17011B, 16PU17011C und 16PU17011D.

S. Biesenbender (✉)

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Berlin, Deutschland

E-Mail: biesenbender@dzhw.eu

© Der/die Autor(en) 2024

S. Biesenbender und J. Hartstein (Hrsg.), *Qualitätsmessung als Prisma*, Higher Education Research and Science Studies,

https://doi.org/10.1007/978-3-658-43683-4_2

Lemke et al.; Traylor und Aman sowie Lippert und Förstner in diesem Band) und setzt den konzeptionellen Ausgangspunkt für die Ableitung von Empfehlungen und Denkanstößen (Peters und Biesenbender in diesem Band). Es beginnt mit einem Literaturüberblick zu den Themen Evaluation in der Wissenschaft (Abschn. 1.1) und Evaluationspraktiken medizinischer Forschung Deutschland (Abschn. 1.2). Der anschließende Abschnitt beinhaltet eine Einführung in das Projekt QuaMedFo (Abschn. 2), stellt den zugrunde liegenden Ansatz für eine systematische und machbare Betrachtung medizinischer Forschung vor (Abschn. 2.1) und leitet über auf die in QuaMedFo durchgeführten und in den nachfolgenden Kapiteln dieses Sammelbands zusammengefassten Analysen (Abschn. 2.2). Es schließt mit einer kurzen Bilanz zu dreieinhalb Jahren Projektarbeit in QuaMedFo und gibt einen Ausblick auf nächste Schritte und das Nachfolgevorhaben QuaMedFo-Transfer (Abschn. 3).

1.1 Kontexte und Strukturen der Evaluation in der Wissenschaft

Forschung – nicht nur im medizinischen Kontext – steht laufend unter Bewertung und Beobachtung durch verschiedene Akteure und zu unterschiedlichen Zwecken bzw. mit unterschiedlichen Zielstellungen. Neben der qualitativen Bewertung durch *Peers* nimmt die quantitative, datengestützte Betrachtung von Forschung eine immer stärkere Rolle ein (Biesenbender & Hornbostel, 2016b; Herwig & Schlattmann, 2016). Dabei werden katalogartige Informationen über die eigentlichen Ergebnispräsentationen („Outputs“) von Forschung – also etwa über Publikationen, Patente, Forschungsdaten oder (erfolgreiche) Drittmittelanträge – als Grundlage herangezogen, um Forschungsprozesse und -leistungen abzubilden. Diese sogenannten Forschungsinformationen werden in unterschiedlichen Kontexten verwendet und verarbeitet; ihre Erhebung und Verarbeitung erfolgt unter Zugrundelegung verschiedener Informationsinfrastrukturen, Systeme und Techniken sowie gewachsener Strukturen, die sich über Wissenschaftssysteme und Forschungseinrichtungen unterscheiden. U. a. unter Zuhilfenahme verschiedener, über die letzten Jahrzehnte etablierter Indikatoren sind sie heute integraler Bestandteil von Evaluationen, Steuerungsinstrumenten und Rechenschaftslegungen über Aktivitäten und zentrale Leistungsprozesse der Forschung (Wissenschaftsrat, 2020; Wissenschaftsrat, 2011).

Neben der direkten Funktion, wissenschaftliche Ergebnisse und Inhalte zu dokumentieren und zu kommunizieren oder für die weitere Forschungsarbeit aufzubereiten, dienen Publikationen und andere Formen wissenschaftlicher

Ergebnispräsentation auch indirekt dazu, Urheberschaft und Reputation zuzuschreiben (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022, S. 10). Vor diesem Hintergrund haben sie sich in den vergangenen Jahrzehnten als integraler Bestandteil von und Ausgangspunkt für qualitative und quantitative output-gestützte Evaluationen in der Wissenschaft (Hornbostel, 2012) etabliert.

Im Kontext des New Public Management als Governance-Form, die auch für öffentlich finanzierte Hochschulen und Forschungseinrichtungen in Deutschland mit wachsenden Anforderungen an Rechenschaftslegung und Transparenz sowie Wettbewerb verbunden ist (Wissenschaftsrat, 2011; Schimank, 2005; Biesenbender & Hornbostel, 2016a; de Rijcke et al., 2016; Hicks, 2012), ist der Bedarf nach Evaluationen sowie der Bereitstellung von Kennzahlen bzw. Indikatoren und Datengrundlagen für die Betrachtung von Forschung kontinuierlich gestiegen. Ursprünglich für andere Bedarfe entwickelte Instrumente – wie etwa der Science Citation Index für Publikationen in wissenschaftlichen Journals, der in den 1950er Jahren als Mittel zur Unterstützung der Literatursuche entwickelt wurde (Hornbostel, 2012) sowie seine Weiterentwicklungen – bildeten so zunächst die Grundlage für empirische Analysen von Forschungspraktiken und die Herausbildung des Felds der Szientometrie auf der Basis von Publikationsdaten (Elkana et al., 1978; Mingers & Leydesdorff, 2015). Seit den 1970er Jahren ist dann eine zunehmende Anwendung und Zweitverwertung in Evaluationen und Berichtslegungen beobachtbar, deren Verfahren und Grad an Institutionalisierung über verschiedene Wissenschaftssysteme variieren (Geuna & Piolatto, 2016; Hicks, 2012; Reborá & Turri, 2013; Thomas et al., 2020).

Etwas zeitversetzt zu den Instrumenten geschah die Entwicklung digitaler kommerzieller Infrastrukturen (Krüger, 2020), die zum einen den gestiegenen Bedarf an Daten bedienen, der sich für Hochschulen und Forschungseinrichtungen aus Evaluationen und Berichtslegung ergibt, und die zum anderen zu der Entstehung weiterer Anwendungsfälle (wie z. B. internationaler Rankings; siehe Rushforth & de Rijcke, 2015; Krüger, 2020) durch ein stetig wachsendes Datenangebot beitragen. Weiterentwicklungen dieser Infrastrukturen greifen den Bedarf nach einfach verfügbaren Tools für individuelle Datenauswertungen und für die Ermittlung bibliometrischer Kennzahlen zur Betrachtung von Forschungseinrichtungen, Personen oder Fächern² für unterschiedliche Evaluations- und Steuerungsanlässe auf und haben die Verwendung von Publikationsdatenbanken und publikationsbasierten Kennzahlen stark vereinfacht sowie für die Anwendung in der Breite geöffnet.

² Wie z. B. bei der Pilotstudie des Wissenschaftsrats zum Forschungsrating (Wissenschaftsrat, 2013b).

Verschiedentlich wird in der Literatur auf die negativen Konsequenzen dieser Entwicklungen hingewiesen. In Bezug auf die strategischen Entscheidungen von Forschungseinrichtungen (de Rijcke et al., 2016), wird beobachtet, dass

- (a) die Abhängigkeit von Dritten im Hinblick auf die Daten, Berechnungen, Maßstäbe und Prozesse, die in kommerziellen Infrastrukturen wie den großen kommerziellen Publikationsdatenbanken hinterlegt sind, sich negativ auf die Weiterentwicklung von Methoden, Instrumenten und Ansätzen auch im Bereich Wissenschaftsevaluation und -steuerung sowie auf eine kritische Auseinandersetzung mit Evaluationsergebnissen auswirkt (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022; Wilsdon et al., 2015),
- (b) die hohen (versunkenen) Anschaffungs- bzw. Nutzungskosten der kommerzialisierten Daten für wissenschaftliche Einrichtungen eine Perpetuierung der Nutzung dieser Datenbanken und damit auch bestehender Evaluationsverfahren und ihrer selektiven Foki sowie die strukturelle Benachteiligung ganzer Fächer (z. B. über die einseitige Betrachtung von über die Datenbanken abgedeckten Journal-Artikeln) zeitigen, sowie dass
- (c) der Rückgriff auf kommerzielle Datenbanken sich ferner negativ auf die Datensouveränität, d. h. den Aufbau eigener Datenbestände und professioneller institutioneller Informationssysteme in Forschungseinrichtungen sowie die Entwicklung von Kompetenzen in diesem Bereich (wie bspw. eine gezielte Personalentwicklung) auswirkt (DINI-AG Forschungsinformationssysteme, 2022).

Vor dem Hintergrund dieser Entwicklungen spielt der Zusammenhang zwischen der Qualität von Indikatoren und ihrer Nutzbarkeit sowie der Ausgestaltung der dahinter liegenden Prozesse und Systeme der institutionellen Datenverarbeitung eine besondere Rolle (Biesenbender, 2019). Entsprechend rücken nichtkommerzielle Standards zunehmend in den Fokus: sowohl für die Forschungsinformationsverarbeitung – wie etwa über den KDSF-Standard (Wissenschaftsrat, 2013a; Wissenschaftsrat, 2016; DINI-AG Forschungsinformationssysteme, 2022) – als auch für die Entwicklung und Anwendung von Bewertungsinstrumenten (Petersohn et al., 2020) und verantwortungsvollen Metriken (Wilsdon et al., 2015; Curry et al., 2020).³

³ Siehe auch das von vielen Akteuren des Wissenschaftssystem unterzeichnete sogenannte „Agreement on Reforming Research Assessment“ der Coalition for Advancing Research Assessment (CoARA) aus dem Jahr 2022: https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf (abgerufen am 22.03.2023), siehe auch <https://coara.eu/agreement/the-commitments/>.

Ein weiterer sehr umfangreicher Strang in der Literatur befasst sich mit den Effekten evaluativer Praktiken und von Indikatoren auf wissenschaftliche Prozesse und Entscheidungen von Forscher:innen. Die Bandbreite von Evaluationen reicht von datengestützten deskriptiven Betrachtungen von Institutionen, Fächern oder Organisationseinheiten (etwa für Berichtslegungen) über die Heranziehung von Forschungsinformationen zum Zwecke der Personalentwicklung oder -auswahl (etwa in Zielvereinbarungsgesprächen oder Berufungsverfahren) bis hin zur leistungsorientierten Mittelvergabe als Grundlage für die Verteilung von Finanzmitteln in Forschungseinrichtungen (Wissenschaftsrat, 2011; Wissenschaftsrat, 2013a). Letztere erhält besonders große Aufmerksamkeit in der wissenschaftlichen Literatur. Neben zahlreichen deskriptiven Überblicksbeiträgen zum Thema, die der international vergleichenden Analyse von leistungsorientierten Mittelvergaben dienen (Geuna & Piolatto, 2016; Hicks, 2012; Rebora & Turri, 2013), gibt es jüngst auch eine Art Meta-Betrachtung, die sich mit der Art und Weise befasst, wie Verfahren der leistungsorientierten Mittelvergabe selbst aus wissenschaftlicher Sicht betrachtet und bewertet werden (Thomas et al., 2020).

Während sich zahlreiche Studien mit der Analyse einzelner Fälle und Instrumente der leistungsorientierten Mittelvergabe auf verschiedenen Ebenen (z. B. national oder organisational) auseinandersetzen (etwa Butler, 2008; Martin & Whitley, 2010), wählen andere Studien eine international vergleichende Perspektive auf das Thema und die Ausgestaltung entsprechender Anreizsysteme (Rebora & Turri, 2013; Geuna & Piolatto, 2016; Hicks, 2010).

Vor diesem Hintergrund lässt sich aus der Literatur ableiten, dass mit der Einrichtung leistungsorientierter Mittelvergaben über Wissenschaftssysteme und Ebenen (Fächer, Institutionen, Leistungseinheiten etc.) hinweg ähnliche Ziele verfolgt werden: die Stärkung wissenschaftlicher Qualität, Exzellenz und Effektivität (Sörlin, 2007; Hicks, 2012). Dabei sind der Wettbewerbsgedanke und Produktivitätserwägungen zentrale einende Komponenten der verschiedenen Instrumente (Thomas et al., 2020). Welche Folgen die Institutionalisierung von leistungsorientierten Mittelvergaben für die Forschung an sich und das Verhalten von Forscher:innen bzw. die Wissensproduktion hat und wie die spezifischen Anreizsysteme auf ihre jeweiligen Zielgrößen wirken, ist Gegenstand eines weiteren Strangs in der Literatur. Während sich einige Studien auf die Effekte auf der Makroebene beziehen, analysieren andere die Folgen und Anpassungsstrategien verschiedener Akteure auf der institutionellen Ebene bzw. in Forschungseinrichtungen (Management sowie Wissenschaftler:innen). Beobachtet werden dabei aufseiten der Forscher:innen neben einem Anstieg in der Anzahl einzelner Outputs etwa durch ‚Salamitaktik‘ (Herrmann-Lingen in diesem Band; Thomas et al.,

2020; Abramo et al., 2011; Mallapaty, 2020) auch der Rückgang von Kreativität und Diversität in den Forschungsthemen, Verwerfungen bzw. Verlust von Synergien in Forschungs- und Lehrprozessen und erhöhte Aufwände der Administration verbunden mit einem Rückgang an Produktivität (Martin, 2011; Whitley et al., 2018).

1.2 Evaluation in der medizinischen Forschung in Deutschland

Im internationalen Vergleich sind Evaluationen in Verbindung mit der leistungsorientierten Vergabe von Forschungsmitteln im föderal organisierten deutschen Wissenschaftssystem wenig etabliert (Biesenbender, 2019). So gibt es bspw. kein zentrales Instrument des übergreifenden Vergleichs von Forschungseinrichtungen oder Fächern (Wissenschaftsrat, 2011). Eine Ausnahme bildet hier die leistungsorientierte Mittelvergabe (LOM) einzelner Bundesländer („Landes-LOM“), über welche jeweils ein Anteil des Landeszuführensbetrags an die Hochschulen für Forschung und Lehre leistungsorientiert verteilt wird. Die institutionelle LOM bzw. Indikatorgestützte Mittelallokation (IMA) (Medizinischer Fakultätentag, 2022) an medizinischen Fakultäten (Herrmann-Lingen in diesem Band; Brähler & Strauß, 2009) hat sich seit 1990er Jahren zunehmend an medizinischen Fakultäten in Deutschland etabliert. Die jeweiligen fakultätsspezifischen Modelle zur Verteilung von Forschungsmitteln basieren in der Regel auf dem Publikationsoutput und eingeworbenen Drittmitteln der Organisationseinheiten der Fakultäten (bspw. Kliniken), sind jedoch nicht umfassend standardisiert und damit nur begrenzt vergleichbar (siehe auch Krupka et al., 2022).

Verschiedene Studien und Bestandsaufnahmen haben sich direkt und indirekt mit den Vor- und Nachteilen solcher Instrumente im deutschen Kontext befasst, intendierte und unintendierte Effekte beleuchtet (Aman & van den Besselaar, 2023; Krempkow et al., 2013; Krempkow & Schulz, 2012) und eine kritische Diskussion der Methoden und Indikatoren initiiert – im übergeordneten Kontext der allgemeinen Debatte in der Wissenschaftsforschung zu Fragen reliabler Maße sowie zu Methoden für die Erfassung von wissenschaftlicher Qualität und Exzellenz (Wilsdon et al., 2015; Bozeman et al., 2001; Aman in diesem Band). Eine kritische Reflektion des Instruments ist auch zunehmend Thema an medizinischen Fakultäten in Deutschland selbst (siehe z. B. Herrmann-Lingen et al., 2014; Medizinischer Fakultätentag, 2022; Traylor & Herrmann-Lingen, 2023; Young et al., 2008; Krupka et al., 2022; siehe auch Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022).

Zentrale Kritik an der institutionellen Forschungs-LOM bezieht sich auf

- (a) die Einführung des Evaluationsinstruments auf wenige Outputs, die dazu führt, dass zentrale Leistungsprozesse und -dimensionen (wie z. B. die Förderung wissenschaftlicher Qualifizierung) (Herrmann-Lingen et al., 2014) keine oder unzureichend Berücksichtigung finden,
- (b) die unzureichende bzw. verzerrende Betrachtung des Impacts berücksichtigter Outputs, etwa durch die fehlende Einbeziehung von Zitationen wissenschaftlicher Publikationen in medizinischen Leitlinien oder an anderen zentralen Publikationsorten (Herrmann-Lingen et al., 2014; Medizinischer Fakultätentag, 2022),
- (c) die Struktur des Instruments und die Einbettung in vorhandene Evaluationsinfrastrukturen, die über einen Fokus auf die in den Datenbanken enthaltenen Publikationen und Publikationstypen sowie verfügbare Impact-Maße bspw. zu einer Bevorzugung bestimmter Publikationstypen und damit einer systematischen Benachteiligung einzelner Fächer bzw. Publikationskulturen führen (Herrmann-Lingen et al., 2014; Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022),
- (d) den Fokus auf einzelne Leistungseinheiten (in diesem Fall Kliniken bzw. Fächer), welcher die Strukturen von übergreifender Forschung (z. B. in übergreifenden interdisziplinären Verbänden oder Netzwerken) nur unzureichend betrachtet,
- (e) den Fokus auf Indikatoren, die es nur ermöglichen, die Rezeption eines Outputs zu messen, aber keinen direkten Bezug zu seiner wissenschaftlichen Qualität herstellen (Medizinischer Fakultätentag, 2022; Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022),
- (f) den damit verbundenen mangelnden Beitrag der LOM zur Qualitätssteigerung (Medizinischer Fakultätentag, 2022),
- (g) das Fehlen eines gelebten Diskurses zu den Methoden bzw. schrittweisen Modifikationen der LOM (Medizinischer Fakultätentag, 2022) sowie
- (h) die mangelnde Zieldefinition (Herrmann-Lingen et al., 2014) bzw. einen mangelnden Konnex zwischen Anreiz und der dem Output zugrunde liegenden Aktivität (Medizinischer Fakultätentag, 2022).

Diese Auseinandersetzung findet statt vor dem Hintergrund einer allgemeinen Diskussion zum Impact medizinischer Forschung, die aus einer weiter gefassten Perspektive die sozialen, medizinischen oder wirtschaftlichen Effekte medizinischer Forschung beleuchtet (z. B. Banzi et al., 2011; Cruz Rivera et al., 2017).

Die in dieser Literatur in Vergleich gesetzten Ansätze eint eine systemische Perspektive auf Forschung, die Impact einrichtungübergreifend – mit Blick auf das Gesundheits- oder Forschungssystem sowie die Gesellschaft – begreift mit dem Ziel, informierte Entscheidungen für die Verteilung knapper Ressourcen zu ermöglichen oder Grundlagen für die Entwicklung von Anreizsystemen zur Steigerung wissenschaftlicher oder medizinischer Leistung bzw. für Fortschritt in der öffentlichen Gesundheit zu schaffen. Dabei bauen einige Ansätze auf komplexen Modellen auf, die (kurzfristige) Inputs, Aktivitäten und Outputs (auf Ebene der Forscher:innen bzw. ihrer Leistungseinheiten oder Forschungseinrichtungen) mit den mittel- bis langfristigen Effekten bzw. Outcome oder Impact verbinden (Weiss, 2007; siehe auch Sarli & Holmes 2009). Strategische Entscheidungen (etwa zu Strukturen und der Bereitstellung von Ressourcen für medizinische Forschung durch Förderorganisationen oder in Fakultäten) werden jedoch auf unterschiedlichen Ebenen und mit verschiedenen zeitlichen Perspektiven getroffen, weswegen die Evaluation der zugrunde liegenden Leistungsprozesse der medizinischen Forschung im Rahmen der LOM allein schon aufgrund des Betrachtungszeitraums nicht an den langfristigen Effekten von Forschung orientiert werden kann. Hier werden kurzfristig verfügbare Kenngrößen benötigt, die sich möglichst direkt den ihnen zugrunde liegenden Prozessen zuordnen lassen und Anhaltspunkte für ihre Qualität bilden.

2 QuaMedFo: Ansatz für eine strukturierte Betrachtung medizinischer Forschung

Im Projekt QuaMedFo wurde die Debatte um die Erfassung von Leistung und die Evaluation medizinischer Forschung aufgegriffen. Dabei orientierte sich das Projekt so eng wie möglich an den Prozessen und Ergebnispräsentationen der Forschung – mittels der Analyse von Outputs im Hinblick auf ihre Eignung für die Beschreibung und Evaluation von Forschung (siehe auch Herrmann-Lingen et al., 2014; Medizinischer Fakultätentag, 2022).

Zentrale Annahme war dabei, dass eine objektive und einfache Messung von Leistung nicht möglich ist, weil a) zugrunde liegende Prozesse der Forschung und Leistungserbringung der Wissenschaftler:innen sehr vielfältig und fachspezifisch sind, b) verschiedene Bewertungskontexte und die damit verbundenen (strategischen) Zielstellungen und Profilen variieren und c) die Auswahl von Indikatoren und Kennzahlen mit ihren jeweiligen Vor- und Nachteilen vor diesem Hintergrund individuell erfolgen muss.

Dementsprechend war es nicht der Anspruch des Projekts, eine Standardisierung bestehender Verfahren der Leistungsorientierten bzw. Indikatorgestützten Mittelvergabe zu initiieren oder Vorschläge für ein allgemeingültiges Evaluierungsverfahren zu entwickeln, sondern eine Reihe verschiedener Datenzugänge, Methoden und Indikatoren zu unterschiedlichen Leistungsprozessen der Forschung zu analysieren, die sich vielfältig einsetzen und kombinieren lassen. Die Analysen sollten auf diese Weise einen Beitrag zur kritischen Bewertung gängiger Bewertungsansätze und Indikatoren, zur laufenden Reflexion bestehender Evaluationsverfahren sowie ihrer Potenziale in der Medizin und damit zur Erhöhung der Bewertungsgerechtigkeit leisten (s. Abschn. 1.2).

Im Rahmen des QuaMedFo-Verbunds wurden in sieben Modulen unterschiedliche Datenzugänge im Hinblick auf ihre Eignung zur Indikatorenbildung sowie zur praktischen Verwendung in Evaluationskontexten beleuchtet. Diese wurden von den vier Projektpartnern teils arbeitsteilig, teils in Kooperation aus verschiedenen methodischen und analytischen Perspektiven bearbeitet. Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) koordinierte den Verbund, entwickelte bibliometrische Indikatoren, analysierte Input-Output-Relationen auf Basis von Daten im KDSF-Standard (siehe auch Wissenschaftsrat, 2013a; Wissenschaftsrat, 2016) und konzipierte sowie implementierte eine Online-Befragung unter Forschenden und Verwaltungspersonal an medizinischen Fakultäten in Deutschland. Das Informationszentrum Lebenswissenschaften (ZB MED) führte Text- und Patentanalysen durch. Der Fokus des Teilprojekts am Leibniz-Informationszentrum Wirtschaft (ZBW) lag auf Social Media und Altmetrics-Daten. Die Universitätsmedizin der Georg-August-Universität Göttingen schließlich brachte die Binnenperspektive der medizinischen Forschung und Praxis in den Verbund ein, organisierte die Kommunikation mit verschiedenen medizinischen Fach-Communities, analysierte Meta-Daten medizinischer Leitlinien und Tagungen sowie die Replikation medizinischer Studien und führte Interviews durch.

Voraussetzung für die erfolgreiche Bearbeitung des Projekts war die enge Kooperation mit drei medizinischen Fakultäten in Deutschland (Pilotfakultäten). Die Pilotfakultäten haben durch die Bereitstellung von Daten zu Fakultätsbeschäftigten, Drittmittelprojekten sowie Publikationen und die Teilnahmen in der Projekt-Befragung, verschiedenen Interviews sowie Projektveranstaltungen einen wesentlichen Anteil am Ergebnis und Erfolg von QuaMedFo. Die folgenden Abschnitte skizzieren den konzeptionellen Ansatz von QuaMedFo und ordnen ihn schlaglichtartig in ausgewählte Ergebnisse der parallel zu den weiteren Analysen

durchgeführten der Online-Befragung unter Forschenden und Verwaltungspersonal an medizinischen Fakultäten in Deutschland ein (Biesenbender et al., 2022a; Biesenbender et al., 2022b).

2.1 Dimensionen medizinischer Forschung und Forschungsoutputs

Forschungsoutputs dienen im Projekt QuaMedFo als Ankerpunkt, um vielfältige Betätigungskontexte der Wissenschaftler:innen innerhalb ihrer Forschungseinrichtung sowie deren jeweilige Anteile am Gesamtoutput abzubilden. Zentral für den Forschungsprozess – nicht nur an medizinischen Fakultäten – ist die Überführung von Ideen in Projektpläne und damit die Konzipierung von Forschungsprojekten, die den Rahmen und Strukturen für wissenschaftliche Aktivitäten und Zusammenarbeit sowie oftmals die Grundlage für weitere Outputs – wie z. B. Publikationen – bilden (Sarli & Holmes, 2009).⁴

Obwohl die erfolgreiche Einwerbung und Verausgabung von Drittmitteln als eine der zentralsten Aufgaben von Forscher:innen gesehen wird – wie die entsprechende in QuaMedFo im Jahr 2021 durchgeführte Befragung unter Forschenden und Verwaltungspersonal an medizinischen Fakultäten in Deutschland (Biesenbender et al., 2022a; Biesenbender et al., 2022b) zeigt –, bringt eine sinnvolle Nutzung in Evaluationskontexten wie der institutionellen LOM Herausforderungen an die Datenhaltung und -verarbeitung mit sich. So bestätigt die Befragung die Befunde aus der Literatur, dass die pauschale Gewichtung von Drittmittelgebern in der LOM Auswirkungen auf die Wahl der Drittmittelgeber im Rahmen von Projektanträgen, aber nicht zwangsläufig auf die Originalität und Qualität der Anträge habe (Medizinischer Fakultätentag, 2022). Gefragt danach, inwieweit Entscheidungen bei der Planung von Drittmittelprojekten auch davon abhängig sind, welche Effekte diese auf das LOM-Ergebnis der eigenen Leistungseinheit haben dürften, geben 36 % der Befragten an, dass die Auswahl der Drittmittelgeber davon abhängig oder stark abhängig ist. Hinsichtlich der Wahl des Forschungsthemas oder der Forschungsmethoden trifft dies auf 30 % bzw. 19 % der Befragten zu.

⁴ Einwerbungen von Drittmittelprojekten lassen sich dabei auch als Inputs in die Forschung begreifen, wenn der Fokus auf die Durchführung des Projekts (Verausgabung der Drittmittel) anstelle der Einwerbung (Bewilligung der Drittmittel) gelegt wird (siehe auch Herrmann-Lingen et al., 2014; s. Hartstein in diesem Band).

Neben der Konzipierung und Durchführung von Forschung besteht eine wesentliche Tätigkeit nicht nur in der medizinischen Forschung in der Kommunikation von Forschungsergebnissen ggf. für verschiedene Adressatenkreise auf der Grundlage von Publikationen (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022). Dies umfasst neben den in gängigen kommerziellen Datenbanken verzeichneten Publikationen in wissenschaftlichen Journals auch Monographien oder Sammelbände sowie andere Publikationstypen und -formate, die unterschiedliche Stadien des zugrunde liegenden Forschungsprozesses abbilden und durch ihre zielgruppenspezifische Aufbereitung einen zentralen Beitrag für eine effektive Kommunikation von wissenschaftlichen Ergebnissen und Grundlagen (wie bspw. Forschungsdaten) liefern (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022, S. 80). Im Bereich der medizinischen Forschung ist insbesondere die Beteiligung bei der Erarbeitung von Leitlinien relevant, die sich als eine Form der Translation im Kontext der klinischen Forschung verorten lässt (vgl. Dembe et al., 2014). Leitlinienkommissionen übernehmen die Erstellung sowie Aktualisierung von Leitlinien auf der Basis von Ergebnissen klinischer Studien (Blümel et al., 2015; Trochim et al., 2011). Während man jedoch davon ausgehen kann, dass jede Forscherin bzw. jeder Forscher nach eigener Maßgabe wissenschaftlich publizieren kann und dann die wissenschaftliche Qualität bspw. im Peer-Review überprüft wird, sind der Umfang und der Bedarf an Leitlinienarbeit stark abhängig von der Fachzugehörigkeit und der Sichtbarkeit bzw. Seniorität der Wissenschaftler:innen. Dieser Aspekt gilt auch für die Möglichkeit, über die Konzipierung und das Anbieten von Fortbildungen einen Beitrag zur Translation von Forschungsergebnissen in die medizinische Praxis zu leisten.

Neben der genuin wissenschaftlichen Arbeit umfassen die Betätigungskontexte medizinischer Forscher:innen auch verschiedene Formen des disziplinären Engagements wie die (aktive) Beteiligung in Fachgesellschaften oder die Ausrichtung von großen Tagungen oder Kongressen als Beitrag zur Vernetzung zu spezifischen Themen auch über die Fachgrenzen hinaus. Der Praxistransfer wissenschaftlicher Ergebnisse schließlich lässt sich für einige Fachgebiete auch über die Anmeldung bzw. Erteilung von Patenten beurteilen.

In der QuaMedFo-Befragung unter Forschenden und Verwaltungspersonal an medizinischen Fakultäten in Deutschland wurden die Teilnehmenden gebeten, die Relevanz dieser unterschiedlichen Betätigungskontexte für die Bewertung von Forschungsleistung von Wissenschaftler:innen bzw. wissenschaftlich-medizinischer Einrichtungen einzuschätzen.

Die Teilnehmenden attestierten den Betätigungskontexten unterschiedliche Wichtigkeit (s. Tab. 1).

Tab. 1 Anteil der Bewertungen der Aktivitäten/Forschungsausputs als „wichtig“ oder „sehr wichtig“ für die Evaluation von Wissenschaftler:innen bzw. Einrichtungen (5er-Skala). (Quelle: Befragung QuaMedFo (Biesenbender et al., 2022b))⁵

Aktivität/ Forschungsausput	(A) Nutzung für Bewertung der Wissenschaftler:in (%)	(B) Nutzung für Bewertung der Einrichtung (%)
Erfolgreiche Einwerbung/ Verausgabung von Drittmitteln	75,35	77,58
Anzahl Publikationen	60,38	61,14
Veröffentlichung von Forschungssoftware	33,15	32,24
Veröffentlichung von Forschungsdaten	64,92	62,53
Mitwirkung an der Verfassung von Leitlinien	57,88	47,99
Anbieten von Fortbildungen	42,98	50,43
Ausrichtung wissenschaftlicher Tagungen	53,58	52,16
Engagement in Fachgesellschaften und wiss. Organisationen	66,19	44,29
Anzahl erteilter Patente	37,74	35,99

So halten 75 bzw. 78 % der Befragungsteilnehmer:innen Einwerbungen von Drittmitteln für wichtig oder sehr wichtig für die Bewertung von Wissenschaftler:innen bzw. Einrichtungen. Beim Publikationsvolumen ergeben sich entsprechende Zustimmungswerte in 60 bzw. 61 % der Rückmeldungen. Bei den Publikationen wird seitens der Wissenschaftler:innen der Veröffentlichung von Forschungsdaten eine besonders große Priorität eingeräumt – ein Aspekt,

⁵ Für verschiedene Aktivitäten und Forschungsausputs wurden jeweils die Einschätzungen (5er-Skala von „Sehr wichtig“ bis „Gar nicht wichtig“) mit den folgenden Fragen ermittelt:

(A) „Für wie wichtig halten Sie persönlich die Aspekte für das Ansehen als Wissenschaftler:in?“ (B) „Bewertung der Einrichtung: Für wie wichtig halten Sie persönlich die Aspekte für die Bewertung wissenschaftlich-medizinischer Einrichtungen?“

der in der Leistungsorientierten Mittelvergabe bislang nur bei einer medizinischen Fakultät in Deutschland Berücksichtigung gefunden hat (Kip et al., 2022; Medizinischer Fakultätentag, 2022), jedoch fundamental für die Replikation von Studien und damit die Qualitätssicherung von Forschungsergebnissen sowie die Nachnutzung von Daten in anderen Kontexten ist (Fabry & Fischer, 2015).

Auch in Bezug auf disziplinäres und praktisches Engagement und dessen Relevanz in der Evaluation medizinischer Forschung ergeben sich hohe Zustimmungswerte, die allerdings differenziert ausfallen je nachdem, welcher Kontext betrachtet wird: Während die Mitwirkung an der Verfassung von Leitlinien von 58 % der Teilnehmer:innen als wichtig oder sehr wichtig erachtet wird für die Bewertung von Wissenschaftler:innen, trifft das im Hinblick auf die Bewertung von Einrichtungen nur auf 48 % der Teilnehmenden zu. In Bezug auf das Engagement in Fachgesellschaften betragen die Zustimmungswerte 66 % bzw. 44 %. An diesen Punkten lässt sich erkennen, dass die entsprechenden Prozesse zwar als wichtig für die Selbstorganisation der medizinischen Wissenschaft erachtet werden, deren Förderung aber möglicherweise nicht im Aufgaben- und Verantwortungsbereich der Forschungseinrichtung gesehen wird (s. Tab. 1).

2.2 Reichweite, Resonanz und Transfer von Forschungsoutputs

Vor dem Hintergrund des übergeordneten Forschungsziels von QuaMedFo – der Entwicklung und Bewertung von Qualitätsmaßen als Grundlage für die Betrachtung und Evaluation von Forschung – fokussierte das Vorhaben auf die Beleuchtung unterschiedlicher Datenzugänge im Hinblick auf das jeweilige Potenzial, daraus sinnvolle und machbare Kenngrößen und Indikatoren zu entwickeln. Neben der grundsätzlichen Eignung eines Datenzugangs für die Nutzung im Rahmen von an medizinischen Fakultäten oder Institutionen durchgeführten Bewertungs- und Berichtskontexten, standen dabei unterschiedliche Aspekte der Machbarkeit im Fokus:

Zunächst wurde für die verschiedenen Forschungsoutputs bzw. Dimensionen die Möglichkeit der Entwicklung von Indikatoren zur Abbildung der Reichweite, der Resonanz und/oder des Transfers in verschiedenen Kontexten bzw. im Hinblick auf verschiedene Adressaten (Wissenschaft, Berufspraxis, Gesellschaft, Wirtschaft) bewertet. Betrachtet wurden dabei auch Aspekte der praktischen Machbarkeit: Über die Zusammenarbeit mit drei Pilotfakultäten wurden zum einen Aspekte der Datenzugänge an medizinischen Fakultäten und zum anderen

Anforderungen an methodische und bibliometrische Kenntnisse für die Anwendung verschiedener Maße und Indikatoren in Bewertungskontexten einbezogen.

Vor diesem Hintergrund analysierten die vier Teilprojekte des QuaMedFo-Verbunds in Zusammenarbeit mit den Pilotfakultäten

- (a) die Erfassbarkeit verschiedener Forschungsoutputs (s. Abschn. 2.1) und die Vollständigkeit der Erfassung,
- (b) die Möglichkeiten und Grenzen, über weitere Parameter und Metadaten zentrale Leistungsprozesse abzubilden und entsprechende Indikatoren anzuwenden sowie
- (c) neue Indikatoren für als relevant bewertete Dimensionen zu entwickeln.

Dabei standen Maße zur Abbildung der Reichweite, der Resonanz und des Transfers an unterschiedliche Adressatenkreise (Wissenschaft, Berufspraxis, Gesellschaft, Wirtschaft) im Vordergrund (Abb. 1), die in den weiteren Beiträgen dieses Sammelbands detailliert beleuchtet werden.

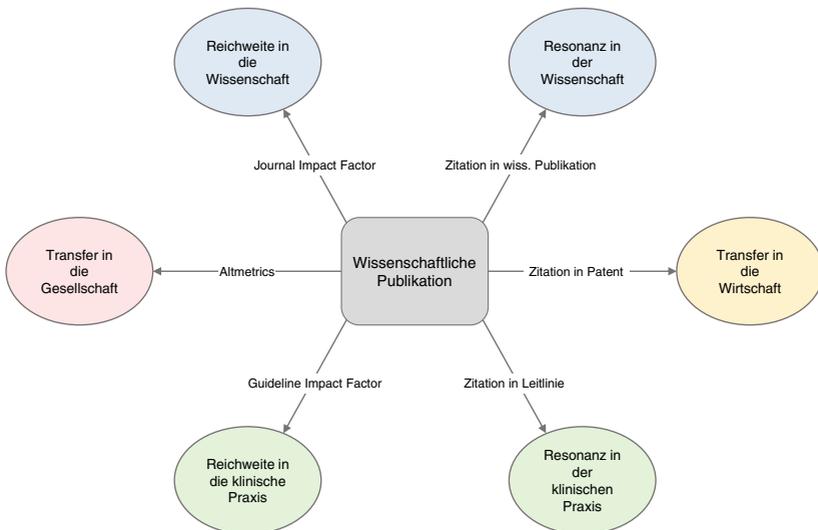


Abb. 1 Reichweite, Resonanz und Transfer von Forschung

Publikationen als zentrales Medium der Kommunikation wissenschaftlicher Ergebnisse und zugrunde liegender Forschungsprozesse (s. Abschn. 2.1) bieten dabei einen Ansatzpunkt, um wissenschaftliche Leistungsprozesse und die Diffusion von Forschungsergebnissen zu rekonstruieren. Vor diesem Hintergrund dienten Publikationen im Projekt QuaMedFo als Ausgangsgröße für die Analyse von Reichweite, Resonanz und Transfer von Forschung in Bezug auf verschiedene Adressatenkreise.

Journalbasierte Metriken wie der Journal Impact Factor (JIF) sind in der Regel Grundlage für die Bewertung von Publikationen in der leistungsorientierten Mittelvergabe an medizinischen Fakultäten in Deutschland, obwohl sie kein verlässliches Maß zur Abbildung der Qualität der in einer Zeitschrift erschienenen Artikel darstellen. Allerdings bieten sie Anhaltspunkte für die (wohlwollende wie kritische) Rezeption bzw. die Reichweite der in ihr erschienenen Artikel bzw. das Prestige der Zeitschrift: Angesehene Journale (mit einem hohen Impact-Faktor) weisen verhältnismäßig hohe Anzahlen von Beitragseinreichungen verbunden mit einer hohen Selektivität auf und haben eine hohe Anziehungskraft auch bei etablierten Wissenschaftler:innen mit hoher Reputation. Darüber hinaus lässt sich davon ausgehen, dass vergleichsweise hohe Zitationszahlen auch ein Effekt der Reputation der Zeitschrift und der damit verbundenen Aufmerksamkeit sind (Wolbring, 2015; siehe auch Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022). Diese als Matthäus-Effekt beschriebene Entwicklung leistet auch einen Beitrag zur Stärkung der Anbieter von Publikationsdatenbanken und der Verfestigung bestehender Infrastrukturen und Bewertungsansätze (Krüger, 2020; s. Abschn. 1.1).

Während journalbasierte Metriken als Qualitätsmaß für wissenschaftliche Publikationen bzw. ihrer zugrunde liegenden Forschungsprozesse also berechnete Kritik erfahren, bieten die Publikationen, die in wissenschaftlichen Zeitschriften erschienen und in weit verbreiteten kommerziellen Datenbanken mit entsprechenden Metadaten verzeichnet sind, dennoch umfassende und weit über den JIF hinausgehende Möglichkeiten zur Ermittlung der Effekte und des Impacts von Forschung.

Aus diesem Grund wurde in QuaMedFo das Potenzial diverser publikationsbasierter Indikatoren als Qualitätsmaße für die medizinische Forschung zur Abbildung von Reichweite, Resonanz und Transfer analysiert, wie in den nachfolgenden Kapiteln dieses Sammelbands zusammenfassend dargestellt wird.

Valeria Aman | Publikations- und Rezeptionsanalyse von drei medizinischen Fakultäten unter Berücksichtigung der LOM

Analysen in QuaMedFo zeigen exemplarisch für die drei Pilotfakultäten des Projekts, welchen Anteil verschiedene Publikationstypen jenseits der klassischen Artikelpublikation (wie bspw. Buchbeiträge und Monographien) am jeweiligen Publikationsaufkommen der Fakultäten haben (Aman in diesem Band) und welche mangels Datenverfügbarkeit (Abdeckung in Zitationsdatenbanken) nicht oder nur untergeordnet in bestehenden LOM-Verfahren honoriert werden können. Während dieser Missstand sich angesichts bestehender Strukturen nicht kurzfristig beheben lässt, sind für den Publikationstyp „Journal-Artikel“ hingegen die Möglichkeiten bereits jetzt groß, über Zitationen in anderen wissenschaftlichen Publikationen die wissenschaftliche Resonanz in differenzierter Form abzubilden (Aman in diesem Band). Vor dem Hintergrund der berechtigten Kritik am Journal Impact Factor (JIF), der in bestehenden Evaluationsverfahren häufig als gängiges Maß zur Bemessung des wissenschaftlichen Werts der in der Zeitschrift erschienenen Artikel sowie des Einflusses auf wissenschaftlichen Fortschritt herangezogen wird (Bornmann et al., 2012; Aksnes et al., 2019), bieten zitationsbasierte Metriken auf der Basis von Publikationen vielfältige und differenzierte Möglichkeiten einer besseren Erfassung der Resonanz von Forschungspublikationen in der Wissenschaft etwa unter Berücksichtigung von fächerspezifischen Publikationskulturen oder über eine Fokussierung auf besonders hochzitierte Publikationen (Aman in diesem Band). Eine Diversifizierung von Indikatoren ermöglicht darüber hinaus den methodischen Schwächen einzelner Maße zu begegnen (Herrmann-Lingen et al., 2014).

Steffen Lemke, Anne Witthake und Isabella Peters | Altmetrics zur Evaluation medizinischer Forschung in Deutschland

Journal-Publikationen mit biomedizinischem Schwerpunkt von Forscher:innen an deutschen Einrichtungen (einschließlich der Pilotfakultäten) wurden in QuaMedFo darüber hinaus einer Analyse mit altmetrischen Daten unterzogen mit dem Ziel, die Eignung von Altmetrics für die Bewertung gesellschaftlicher Resonanz von Forschung zu beleuchten. Anders als bei der Analyse des wissenschaftlichen Impacts ergeben sich zwar Grenzen der Anwendbarkeit altmetrischer Daten aufgrund der geringen Fallzahlen, besonders im Bereich der Policy-Papiere sind die Ergebnisse jedoch vielversprechend und bieten Anhaltspunkte für eine perspektivisch differenzierte Betrachtung von Impact auf hohen Aggregationsebenen (Lemke et al. in diesem Band).

Christoph Traylor und Valeria Aman | Welche Charakteristika weist die in Leitlinien zitierte Literatur auf? – Eine Fallstudie anhand einer S2e- und einer S3-Leitlinie

Ein weiterer Schwerpunkt von QuaMedFo bestand in der Entwicklung neuer Ansätze zur Abbildung der praktischen Reichweite und Resonanz medizinischer Forschung (Traylor und Aman in diesem Band; Traylor & Herrmann-Lingen, 2023) und der Ausweitung des Fokus auf das Thema Leitlinienarbeit (Herrmann-Lingen in diesem Band; Herrmann-Lingen et al., 2014). Als geeignetes Maß hierfür wird in der Literatur die direkte Referenz auf eine bestimmte wissenschaftliche Veröffentlichung mit Empfehlungscharakter innerhalb der Leitlinie gesehen (Dembe et al., 2014). Ergänzend dazu wurde in QuaMedFo der sogenannte Guideline Impact Factor (GLIF) entwickelt, der analog zum JIF als Maß, wie oft Artikel aus einer Zeitschrift in anderen Artikeln zitiert werden, angibt, welche praktische Resonanz Artikel einer wissenschaftlichen Zeitschrift bei der Erarbeitung von Leitlinien finden.

Klaus Lippert und Konrad Förstner | Nutzung von Patentdaten zur Erfassung der wirtschaftlichen Verwertung von Forschung

Der Transfer in die Wirtschaft steht im Vordergrund der Analyse von Patenten der Einrichtungen – ebenfalls über eine Verknüpfung mit den Publikationsdaten (Journal-Artikeln) der Pilotfakultäten. Mit dem Ansatz der Bildung von Publikations-Patent-Paaren und der Bewertung ihrer technischen und methodischen Machbarkeit wurde in QuaMedFo ein Verfahren getestet, bei dem eine Einrichtung (z. B. medizinische Fakultät) ihr inhaltliches Profil in Bezug auf den Transfer von Forschungsergebnissen in die Wirtschaft durch ihre Wissenschaftler:innen selbst ermitteln kann (Lippert und Förstner in diesem Band).

Judith Hartstein | Input–Output-Relationen – Zu den Voraussetzungen einer Effizienzmessung medizinischer Fakultäten anhand von Drittmitteln

Um den publikationszentrierten Fokus (perspektivisch) zu erweitern, umfasste das Projekt QuaMedFo ebenfalls eine fallstudienartige Auseinandersetzung mit der Nutzung von Informationen zu Drittmittelprojekten zur Erfassung von Outputs der bzw. Inputs in die Forschung (Hartstein in diesem Band), welche bereits gängige Praxis in der institutionellen LOM medizinischer Fakultäten in Deutschland ist (Medizinischer Fakultätentag, 2022). Die Analysen in QuaMedFo untersuchen auf der Grundlage realer Drittmitteldaten der drei Pilotfakultäten, wie sich Drittmittelzusagen (geglättet über jährliche Drittmittelleinnahmen) nutzen lassen, um den Forschungsoutput Projekt(-Zusage) abzubilden. Die Analysen verdeutlichen, welche Herausforderungen bei der Betrachtung von Drittmitteln als

Input-Größe bezüglich Modellierung und Datenaufbereitung sowie -verfügbarkeit derzeit noch bestehen – auch in Ermangelung eines Standards für die Verarbeitung und Aggregation (Hartstein in diesem Band).

Isabella Peters und Sophie Biesenbender | Qualitätsmaße zur Evaluierung medizinischer Forschung – Erkenntnisse und Empfehlungen aus dem Projekt QuaMedFo

Der publikationszentrierte Fokus in QuaMedFo ermöglichte – trotz aller Einschränkungen bei der Abdeckung und der Berücksichtigung von Publikationstypen und -formen – die explorative Erkundung verschiedener Dimensionen von Forschung über die Anreicherung mit weiteren Daten zu Leitlinien und Patenten und Altmetrics sowie eine differenzierte Betrachtung von Drittmittelinformationen. Schlussfolgerungen und Empfehlungen aus dem Projekt QuaMedFo werden überblicksartig in einem Kapitel zusammengefasst (Peters und Biesenbender in diesem Band). Der Beitrag benennt Herausforderungen für die Verarbeitung und Anwendung von Forschungsinformationen im Allgemeinen sowie für die Evaluation und das Monitoring von Forschung – nicht nur im medizinischen Kontext. Dabei bestehen Anforderungen sowohl an (vorhandene) Systeme und zugrunde liegende Prozesse der Informationserhebung und -verarbeitung in den Einrichtungen als auch an methodisch-fachliche Kenntnisse des dafür zuständigen Personals und der Leitungen. Dies ist von besonderer Relevanz vor dem Hintergrund heterogener und unvollständiger Datenquellen (wie bspw. Publikationsdatenbanken) und komplexer Indikatorik, die eine an das jeweilige Steuerungsanliegen angepasste differenzierte Kalibrierung von Evaluationsinstrumenten und -verfahren erfordert.

3 Ausblick

In seiner insgesamt dreieinhalbjährigen Projektlaufzeit befasste sich QuaMedFo aus verschiedenen methodischen Perspektiven und Zugängen mit den unterschiedlichsten Facetten der institutionellen Forschungsinformationsverarbeitung, der Entwicklung bzw. Bewertung von Indikatoren und der Nutzung dieser Daten in konkreten Evaluationskontexten.

Das damit verbunden Ziel von QuaMedFo war es, – neben der Stärkung bestehender LOM-Systeme und Verfahren der Personalauswahl – auch Leistungsdimensionen und Datenzugänge für die Nutzung in internen Strategieprozessen jenseits der bestehenden Steuerungssysteme aufzuzeigen (Wissenschaftsrat, 2020; Wissenschaftsrat, 2011). Von Relevanz waren in diesem Zusammenhang Fragen

der Machbarkeit angesichts der vorhandenen Systeme und Datenzugänge und bestehender methodischer Kompetenzen zu Themen wie Evaluationen, Bibliometrie oder Indikatorik in medizinischen Fakultäten. Ziel der Analysen war es, einen Beitrag zum professionellen Umgang mit Forschungsinformationen über geeignete Systeme und Standards zu leisten (Wissenschaftsrat, 2013a; Wissenschaftsrat, 2016; DINI-AG Forschungsinformationssysteme, 2022).

Mit dieser Perspektive wählte QuaMedFo eine erweiterte Sicht, um ausgehend von tatsächlichen Betätigungskontexten von Forscher:innen die Entwicklung adäquater und ausgewogener Betrachtungs- und Bewertungssysteme anzuregen und zu unterstützen. Dabei griff das Projekt die vielfach geäußerte Kritik an bestehenden Evaluationssystemen nicht nur im Bereich der medizinischen Forschung auf (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022; Herrmann-Lingen et al., 2014; Medizinischer Fakultätentag, 2022).

Im Verlauf von QuaMedFo wurde wiederholt sichtbar, dass die Etablierung von innovativen Kenngrößen und Indikatoren zur Erfassung medizinischer Forschung(sleistungen) ein umfassender Prozess ist, der stets in einem Spannungsfeld mit den dafür erforderlichen Anpassungen von Systemen und Verfahren der institutionellen Forschungsinformationsverarbeitung und -nutzung steht und alle Beteiligten – Management, Forscher:innen und datenbereitstellende Organisationen (wie bspw. Datenbankanbieter:innen) – einbezieht. QuaMedFo hat in dieser Hinsicht einen ersten Schritt und ein Beispiel dafür aufgezeigt, dass eine multiperspektivische Befassung in einem Verbundkontext einen vielversprechenden Ansatz darstellt, um der Komplexität des Themas gerecht zu werden (s. auch Abschlusskapitel in diesem Band zu den Erkenntnissen und Empfehlungen aus dem Projekt QuaMedFo; Peters und Biesenbender in diesem Band). Ein Transfer- bzw. Nachfolgeprojekt von QuaMedFo „Abbildung des Impacts medizinischer Forschung durch Leitlinien-Zitationen und -Autorschaften“ (2023–2024)⁶ geht vor diesem Hintergrund die nächsten Schritte. Das Verbundvorhaben der QuaMedFo-Projektpartner Universitätsmedizin der Georg-August-Universität Göttingen und Deutsches Zentrum für Hochschul- und Wissenschaftsforschung widmet sich vertieft insbesondere den Fragen und Herausforderungen der praktischen Machbarkeit, indem es die prototypische Umsetzung und Anwendung entwickelter Indikatoren zur Abbildung von Leitlinien-Engagement und -Relevanz unterstützt. Auf dieser Grundlage wird das Transfer-Projekt weitere wichtige und praxisrelevante Anhaltspunkte liefern zur Bewertung der Möglichkeiten, Grenzen und Modalitäten von Wandlungsprozessen im Bereich der Forschungsevaluation und des institutionellen Forschungsinformationsmanagements.

⁶ https://www.dzhw.eu/forschung/projekt?pr_id=724.

Danksagung Besonderer Dank gilt Prof. a.D. Dr. Stefan Hornbostel und Prof. Dr. Christoph Herrmann-Lingen für die Anbahnung und laufende Begleitung des Projekts QuaMedFo. Weiterer Dank gilt den Pilotfakultäten des Projekt QuaMedFo für die Kooperation über die gesamte Projektlaufzeit und die Bereitstellung von Primärdaten für verschiedene Analysen des Projekts. Das Projektteam dankt ferner Alexander Kratzenberg für die Koordination der Akquise dieser Daten und Prof. Dr. Christoph Herrmann-Lingen für die Entwicklung der für das Projekt maßgeblichen Klassifikation der Fächergruppen und die Erstellung von Konkordanzen zu den Organisationseinheiten und Fächern/Fächergruppen der Pilotfakultäten.

Literatur

- Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). National research assessment exercises: The effects of changing the rules of the game during the game. *Scientometrics*, 88(1), 229–238. <https://doi.org/10.1007/s11192-011-0373-2>.
- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1). <https://doi.org/10.1177/2158244019829575>.
- Aman, V., & van den Besselaar, P. (2023). Effects of performance-Based funding systems on publication behavior – A case study of German Medical Faculties. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4388750>.
- Banzi, R., Moja, L., Pistotti, V., Facchini, A., & Liberati, A. (2011). Conceptual frameworks and empirical approaches used to assess the impact of health research: An overview of reviews. *Health Research Policy and Systems*, 9(1), 26. <https://doi.org/10.1186/1478-4505-9-26>.
- Biesenbender, S. (2019). The governance and standardisation of research information in different science systems: A comparative analysis of Germany and Italy. *Higher Education Quarterly*, 73(1), 116–127. <https://doi.org/10.1111/hequ.12193>.
- Biesenbender, S., Hartstein, J., Herrmann-Lingen, C., Kratzenberg, A., & Traylor, C. (2022). *Methodenbericht zur Befragung von wissenschaftlichen und administrativen Beschäftigten an medizinischen Fakultäten im Projekt QuaMedFo (Qualitätsmaße zur Evaluierung medizinischer Forschung)*. <https://doi.org/10.21249/DZHW:QMF2021-DMR:1.0.0>.
- Biesenbender, S., Hartstein, J., Herrmann-Lingen, C., Kratzenberg, A., Traylor, C., Deutsches Zentrum Für Hochschul- Und Wissenschaftsforschung (DZHW), & Universitätsmedizin Göttingen. (2022). *Befragung von wissenschaftlichen und administrativen Beschäftigten an medizinischen Fakultäten im Projekt QuaMedFo (Qualitätsmaße zur Evaluierung medizinischer Forschung)*. German Centre for Higher Education Research and Science Studies (DZHW). <https://doi.org/10.21249/DZHW:QMF2021:1.0.0>.
- Biesenbender, S., & Hornbostel, S. (2016a). The research core dataset for the German science system: Challenges, processes and principles of a contested standardization project. *Scientometrics*, 106(2), 837–847. <https://doi.org/10.1007/s11192-015-1816-y>.
- Biesenbender, S., & Hornbostel, S. (2016b). The Research Core Dataset for the German science system: Developing standards for an integrated management of research information. *Scientometrics*, 108(1), 401–412. <https://doi.org/10.1007/s11192-016-1909-2>.

- Blümel, C., Gauch, S., Hendriks, B., Krüger, A. K., & Reinhart, M. (2015). *In search of translational research – Report on the Development and Current Understanding of a New Terminology in Medical Research and Practice*. https://www.bihealth.org/fileadmin/publikationen/dateien/iFQ-BIH-Report_2015_web.pdf.
- Bornmann, L., Marx, W., Gasparyan, A. Y., & Kitas, G. D. (2012). Diversity, value and limitations of the journal impact factor and alternative metrics. *Rheumatology International*, 32(7), 1861–1867. <https://doi.org/10.1007/s00296-011-2276-1>.
- Bozeman, B., Dietz, J. S., & Gaughan, M. (2001). Scientific and technical human capital: An alternative model for research evaluation. *International Journal of Technology Management*, 22(7/8), 716. <https://doi.org/10.1504/IJTM.2001.002988>.
- Brähler, E., & Strauß, B. (2009). Leistungsorientierte Mittelvergabe an Medizinischen Fakultäten: Eine aktuelle Übersicht. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 52(9), 910–916. <https://doi.org/10.1007/s00103-009-0918-1>.
- Butler, L. (2008). Using a balanced approach to bibliometrics: Quantitative performance measures in the Australian Research Quality Framework. *Ethics in Science and Environmental Politics*, 8, 83–92. <https://doi.org/10.3354/esep00077>.
- Cruz Rivera, S., Kyte, D. G., Aiyegbusi, O. L., Keeley, T. J., & Calvert, M. J. (2017). Assessing the impact of healthcare research: A systematic review of methodological frameworks. *PLOS Medicine*, 14(8), e1002370. <https://doi.org/10.1371/journal.pmed.1002370>.
- Curry, S., de Rijcke, S., Hatch, A., Pillay, D. (Gansen), van der Weijden, I., & Wilsdon, J. (2020). *The changing role of funders in responsible research assessment: Progress, obstacles and the way ahead* (S. 2449096 Bytes). Research on Research Institute. <https://doi.org/10.6084/M9.FIGSHARE.13227914.V1>.
- de Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—A literature review. *Research Evaluation*, 25(2), 161–169. <https://doi.org/10.1093/reseval/rvv038>.
- Dembe, A. E., Lynch, M. S., Gugiu, P. C., & Jackson, R. D. (2014). The translational research impact scale: Development, construct validity, and reliability testing. *Evaluation & the Health Professions*, 37(1), 50–70. <https://doi.org/10.1177/0163278713506112>.
- Deutsche Forschungsgemeinschaft | AG Publikationswesen. (2022). *Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung*. <https://doi.org/10.5281/ZENODO.6538163>.
- DINI-AG Forschungsinformationssysteme. (2022). *Management von Forschungsinformationen in Hochschulen und Forschungseinrichtungen*. Humboldt-Universität zu Berlin. <https://doi.org/10.18452/25440>.
- Elkana, Y., Lederberg, J., Merton, R. K., Thackray, A., & Zuckerman, H. (Hrsg.). (1978). *Toward a metric of science: The advent of science indicators*. Wiley.
- Fabry, G., & Fischer, M. R. (2015). Replication – The ugly duckling of science? *GMS Zeitschrift Für Medizinische Ausbildung*; 32(5):Doc57. <https://doi.org/10.3205/ZMA000999>.
- Geuna, A., & Piolatto, M. (2016). Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). *Research Policy*, 45(1), 260–271. <https://doi.org/10.1016/j.respol.2015.09.004>.
- Herrmann-Lingen, C., Brunner, E., Hildenbrand, S., Loew, T. H., Raupach, T., Spies, C., Treede, R.-D., Vahl, C.-F., & Wenz, H.-J. (2014). Evaluation of medical research performance – position paper of the Association of the Scientific Medical Societies in Germany

- (AWMF). *GMS German Medical Science; 12:Doc11; ISSN 1612–3174*. <https://doi.org/10.3205/000196>.
- Herwig, S., & Schlattmann, S. (2016). Eine wirtschaftsinformatische Standortbestimmung von Forschungsinformationssystemen. *Lecture Notes in Informatics (LNI)*, 901–914. <https://dl.gi.de/bitstream/handle/20.500.12116/1198/901.pdf?sequence=1&isAllowed=y>.
- Hicks, D. (2010). Overview of models of performance-based research funding systems. In OECD (Hrsg.), *Performance-based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings* (S. 23–52). OECD. https://read.oecd-ilibrary.org/education/performance-based-funding-for-public-research-in-tertiary-education-institutions/overview-of-models-of-performance-based-research-funding-systems_9789264094611-4-en#page1.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251–261. <https://doi.org/10.1016/j.respol.2011.09.007>.
- Hornbostel, S. (2012). Evaluation und Evaluationsforschung. In S. Maasen, M. Kaiser, M. Reinhart, & B. Sutter (Hrsg.), *Handbuch Wissenschaftssoziologie* (S. 277–287). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18918-5_22.
- Kip, M., Bobrov, E., Koenig, S., Riedel, N., nachev, vladislav, & Dirnagl, U. (2022). *Open Data LoM – The introduction of Open Data in the institutional performance-based funding (Leistungsorientierte Mittelvergabe, LoM) at Charité Universitaetsmedizin Berlin*. <https://doi.org/10.17605/OSF.IO/GEHDA>.
- Krempkow, R., Landrock, U., Neufeld, J., & Schulz, P. (2013). *Intendierte und nicht-intendierte Effekte dezentraler Anreizsysteme am Beispiel der fakultätsinternen leistungsorientierten Mittelvergabe in der Medizin*. iFQ – Institut für Forschungsinformation und Qualitätssicherung. <https://www.forschungsinform.de/projekte/GOMED/GOMED-Abschlussbericht.pdf>.
- Krempkow, R., & Schulz, P. (2012). Welche Effekte hat die leistungsorientierte Mittelvergabe? Das Beispiel der medizinischen Fakultäten Deutschlands. *Die Hochschule: Journal für Wissenschaft und Bildung*, 21, 121–141. <https://doi.org/10.25656/01:16302>.
- Krüger, A. K. (2020). Quantification 2.0? Bibliometric Infrastructures in Academic Evaluation. *Politics and Governance*, 8(2), 58–67. <https://doi.org/10.17645/pag.v8i2.2575>.
- Krupka, S., Koenig, S., & Kip, M. (2022). *Responsible Indicators (RespInd) – Investigating the status quo of various models of performance-based allocation of funds (LoM) in research among institutions in the Berlin University Alliance (BUA)*. <https://doi.org/10.17605/OSF.IO/JGMQZ>.
- Mallapaty, S. (2020). China bans cash rewards for publishing papers. *Nature*, 579(7797), 18–18. <https://doi.org/10.1038/d41586-020-00574-8>.
- Martin, B. R. (2011). The Research Excellence Framework and the „Impact Agenda“: Are we Creating a Frankenstein Monster? *Research Evaluation*, 20(3), 247–254. <https://doi.org/10.3152/095820211X131118583635693>.
- Martin, B., & Whitley, R. (2010). The UK Research Assessment Exercise: A Case of Regulatory Capture. In R. Whitley, J. Gläser, & L. Engwall (Hrsg.), *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation* (S. 51–80). Oxford University Press. <https://doi.org/10.1093/acprof:Oso/9780199590193.003.0002>.

- Medizinischer Fakultätentag. (2022). *Impulspapier: Weiterentwicklung der Indikator-gestützten Mittelallokation für die Forschung in der Hochschulmedizin*. Medizinischer Fakultätentag. https://medizinische-fakultaeten.de/wp-content/uploads/2022/10/MFT-Impulspapier_IMA_LOM_2022.pdf.
- Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, 246(1), 1–19. <https://doi.org/10.1016/j.ejor.2015.04.002>.
- Petersohn, S., Biesenbender, S., & Thiedig, C. (2020). Investigating Assessment Standards in the Netherlands, Italy, and the United Kingdom: Challenges for Responsible Research Evaluation. In K. Jakobs (Hrsg.), *Advances in Standardization Research* (S. 54–94). IGI Global. <https://doi.org/10.4018/978-1-7998-2181-6.ch003>.
- Rebora, G., & Turri, M. (2013). The UK and Italian research assessment exercises face to face. *Research Policy*, 42(9), 1657–1666. <https://doi.org/10.1016/j.respol.2013.06.009>.
- Rushforth, A., & de Rijcke, S. (2015). Accounting for Impact? The Journal Impact Factor and the Making of Biomedical Research in the Netherlands. *Minerva*, 53(2), 117–139. <https://doi.org/10.1007/s11024-015-9274-5>.
- Sarli, C., & Holmes. (2009). *The Becker Medical Library Model for Assessment of Research Impact*. The Becker Medical Library Model for Assessment of Research Impact. <https://becker.wustl.edu/impact-assessment>.
- Schimank, U. (2005). ‘New Public Management’ and the Academic Profession: Reflections on the German Situation. *Minerva*, 43(4), 361–376. <https://doi.org/10.1007/s11024-005-2472-9>.
- Sörlin, S. (2007). Funding diversity: Performance-based funding regimes as drivers of differentiation in Higher Education Systems. *Higher Education Policy*, 20(4), 413–440. <https://doi.org/10.1057/palgrave.hep.8300165>.
- Thomas, D. A., Nedeva, M., Tirado, M. M., & Jacob, M. (2020). Changing research on research evaluation: A critical literature review to revisit the agenda. *Research Evaluation*, 29(3), 275–288. <https://doi.org/10.1093/reseval/rvaa008>.
- Traylor, C., & Herrmann-Lingen, C. (2023). Does the journal impact factor reflect the impact of German medical guideline contributions? *Scientometrics*, 128(3), 1951–1962. <https://doi.org/10.1007/s11192-023-04647-z>.
- Trochim, W., Kane, C., Graham, M. J., & Pincus. (2011). Evaluating translational research: A process marker model. *Clinical and Translational Science*, 4(3), 153–162. <https://doi.org/10.1111/j.1752-8062.2011.00291.x>.
- Weiss, A. P. (2007). Measuring the impact of medical research: Moving from outputs to outcomes. *American Journal of Psychiatry*, 164(2), 206–214. <https://doi.org/10.1176/ajp.2007.164.2.206>.
- Whitley, R., Gläser, J., & Laudel, G. (2018). The impact of changing funding and authority relationships on scientific innovations. *Minerva*, 56(1), 109–134. <https://doi.org/10.1007/s11024-018-9343-7>.
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Ker-ridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. <https://doi.org/10.13140/RG.2.1.4929.1363>.

- Wissenschaftsrat. (2011). *Empfehlungen zur Bewertung und Steuerung von Forschungsleistung* (Drs. 1656–11). https://www.wissenschaftsrat.de/download/archiv/1656-11.pdf?__blob=publicationFile&v=3.
- Wissenschaftsrat. (2013a). *Empfehlungen zu einem Kerndatensatz Forschung* (Drs. 2855–13). https://www.wissenschaftsrat.de/download/archiv/2855-13.pdf?__blob=publicationFile&v=1.
- Wissenschaftsrat. (2013b). *Empfehlungen zur Zukunft des Forschungsratings* (Drs. 3409–13). https://www.wissenschaftsrat.de/download/archiv/3409-13.pdf?__blob=publicationFile&v=1.
- Wissenschaftsrat. (2016). *Empfehlungen zur Spezifikation des Kerndatensatz Forschung* (Drs. 5066–16). https://www.wissenschaftsrat.de/download/archiv/5066-16.pdf?__blob=publicationFile&v=1.
- Wissenschaftsrat. (2020). *Stellungnahme zur Einführung des Kerndatensatz Forschung* (Drs. 8652–20). https://www.wissenschaftsrat.de/download/2020/8652-20.pdf?__blob=publicationFile&v=5.
- Wolbring, T. (2015). Anatomie des Journal Impact Faktors. *Soziale Welt*, 66(2), 121–140. <https://doi.org/10.5771/0038-6073-2015-2-121>.
- Young, N. S., Ioannidis, J. P. A., & Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLOS Medicine*, 5(10), e201. <https://doi.org/10.1371/journal.pmed.0050201>.

Biesenbender, Sophie Sophie Biesenbender ist promoviert in Politik- und Verwaltungswissenschaften. Sie forscht und arbeitet seit Oktober 2012 am DZHW (bzw. der Vorgängerorganisation, dem Institut für Forschungsinformation und Qualitätssicherung – iFQ) in Berlin mit einem Fokus auf die Themen Forschungsinformationen und Indikatorik. Von Mitte 2020 bis Anfang 2022 war sie kommissarische Leiterin der Abteilung „Forschungssystem und Wissenschaftsdynamik“ des DZHW. Seit 2022 ist sie Leiterin der Geschäftsstelle der Kommission für Forschungsinformationen in Deutschland (KFiD).

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Publikations- und Rezeptionsanalyse von drei medizinischen Fakultäten unter Berücksichtigung der LOM

Valeria Aman

1 Einleitung

In diesem Beitrag steht die bibliometrische Publikations- und Rezeptionsanalyse von drei medizinischen Fakultäten in Deutschland im Fokus. Die Bibliometrie nutzt primär Publikations- und Zitationsdaten für ihre Analysen, die in Publikations- und Zitationsdatenbanken wie Web of Science oder Scopus enthalten sind. Bibliometrische Indikatoren werden zumeist in quantitative Indikatoren zur Messung der Produktivität eingeteilt und in Leistungsindikatoren, die einen Proxy für die Messung von Qualität darstellen (Joshi, 2014).

Indikatoren liefern entscheidungs- und steuerungsrelevante Informationen und versuchen dabei die Realität des Wissenschaftssystems abzubilden, wobei sie zugleich durch ihren Gebrauch diese Realität miterzeugen (Felt, 1999). Zur Bewertung der medizinischen Forschung hat sich der Journal Impact Factor (JIF) durchgesetzt, da er das bekannteste Rezeptionsmaß für Zeitschriften ist, leicht einsehbar ist und einen Vergleich zwischen Zeitschriften ermöglicht. Es ist in der Literatur wiederum gut belegt, dass der JIF kein geeignetes Maß ist, um die Qualität von Forschung zu evaluieren (Seglen, 1998; PLOS Medicine, 2006).

Der Wettbewerb um akademische Positionen, um finanzielle Ressourcen und Reputation stellt die Forschenden unter den Leistungsdruck möglichst viel in Zeitschriften mit hohem JIF zu publizieren (Hostettler et al., 2015). Diese Anreize

V. Aman (✉)

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Berlin, Deutschland

E-Mail: aman@dzhw.eu

© Der/die Autor(en) 2024

S. Biesenbender und J. Hartstein (Hrsg.), *Qualitätsmessung als Prisma*, Higher Education Research and Science Studies, https://doi.org/10.1007/978-3-658-43683-4_3

führen zu einem wachsenden Publikationsaufkommen und können einen negativen Einfluss auf die Qualität der Forschungsergebnisse mit sich bringen (s. Herrmann-Lingen in diesem Band).

Im Folgenden werden die Publikationen von drei medizinischen Fakultäten in Deutschland analysiert, die im Rahmen der internen Leistungsorientierten Mittelvergabe (LOM) an den drei Fakultäten erfasst wurden. Zwei der medizinischen Fakultäten haben Publikationsdaten aus den Jahren 2005–2018 bereitgestellt, eine weitere aus dem Zeitraum 2007–2018. Für die einheitliche Handhabung des Publikationszeitraums werden Publikationen in dem 12-jährigen Zeitraum 2007–2018 betrachtet.

Die Publikationsanalyse hat das Publikationsaufkommen zum Gegenstand und die verschiedenen Publikationstypen, die an den medizinischen Fakultäten im Rahmen der LOM erfasst werden. Für die Analyse liegt die Klassifikation des Medizinischen Fakultätentags (MFT) zugrunde, um Unterschiede in den Publikationspraktiken zwischen den Fächergruppen berücksichtigen zu können. Die drei medizinischen Fakultäten werden mit Fakultät A, B und C adressiert. An den Fakultäten A und B werden zehn Fächergruppen unterschieden, an der Fakultät C gibt es hingegen acht Fächergruppen.

Die Publikationen der drei Pilotfakultäten wurden den Publikationen in einer lizenzierten Datenbank-Version des Web of Science (WoS) zugeordnet. Bei den zugeordneten Publikationen handelt es sich vorwiegend um Zeitschriftenaufsätze. Die Rezeptionsanalyse umfasst zitationsbasierte Indikatoren, die die Leistungen einzelner Fächergruppen auszuweisen vermögen. Als geeignete Indikatoren haben sich in der Bibliometrie solche etabliert, die vom CWTS im sog. Leiden Ranking verwendet werden.¹ Darunter finden sich normalisierte Indikatoren wie PP(top10 %) und der MNCS, die im untenstehenden Abschnitt zur Rezeptionsanalyse (s. Abschn. 3) erläutert werden. Ferner wird in der Rezeptionsanalyse auf die fächerspezifisch unterschiedlich hoch ausfallenden Zitationsraten eingegangen und auf den Journal Impact Factor (JIF), der zentral in der LOM ist. Ein besonderer Fokus liegt auf dem Zusammenhang der Sprache von Publikationen und dem JIF sowie dem damit verbundenen Rückgang von deutschsprachigen Aufsätzen. Im Fazit werden die wichtigsten Ergebnisse diskutiert und Empfehlungen zum Umgang mit Publikationen in der LOM gegeben.

¹ CWTS (Centre for Science and Technology Studies): <https://www.leidenranking.com/information/indicators>.

2 Publikationsanalyse

Dieses Kapitel unterteilt sich in die Analyse von Zeitschriftenaufsätzen (Abschn. 2.1) und die Analyse von Buchbeiträgen und Monographien (Abschn. 2.2). Bei den Zeitschriftenaufsätzen handelt es sich um solche, die Publikationen in WoS zugeordnet werden konnten. Die Rezeptionsanalyse in Abschn. 3 basiert auf den Zeitschriftenaufsätzen und den Zitationen aus WoS.

2.1 Zeitschriftenaufsätze

Tab. 1 gibt einen Überblick über die Zeitschriftenaufsätze an den drei Fakultäten im Zeitraum 2007–2018, unterteilt nach Fächergruppen. Die Verteilung der Zeitschriftenaufsätze über die Fächergruppen unterscheidet sich an den drei Fakultäten, was ein Indiz für unterschiedliche Schwerpunktsetzungen und Fachspezifika ist.

Tab. 1 Publikationsoutput im Zeitraum 2007–2018, unterteilt nach Fächergruppen

Fächergruppe	Fakultät A	Fakultät B	Fakultät C
Chirurgie	12,9%	10,5%	13,3%
Diagnostische und Strahlenfächer	8,9%	5,9%	14,5%
Innere Medizin	15,0%	15,6%	20,2%
Neuromedizinische Fächer	6,0%	18,5%	7,4%
Psychomedizinische Fächer	10,8%	9,6%	
Theoretische Fächer	18,9%	13,3%	7,4%
Vorklinische Fächer	7,8%	8,3%	8,7%
Weitere klinische Fächer	12,1%	11,6%	4,8%
Weitere Klinisch-theoretische Fächer	6,4%	4,6%	23,7%
Zahnmedizin	1,0%	2,1%	
Gesamtzahl an Zeitschriftenaufsätzen in 2007-2018	20.916	12.589	11.927

Die tabellarische Übersicht zeigt, dass an der Fakultät A mehr als 20.000 Zeitschriftenaufsätze zwischen 2007 und 2018 erschienen sind. Fakultät B und Fakultät C bewegen sich mit ungefähr 12.000 Zeitschriftenaufsätzen in der gleichen Größenordnung. An der Fakultät A stellen die Theoretischen Fächer die meisten Zeitschriftenaufsätze, gefolgt von der Inneren Medizin. An der Fakultät B dominieren die Zeitschriftenaufsätze in den Neuromedizinischen Fächern, ebenso gefolgt von der Inneren Medizin. Hingegen sind an der Fakultät C die Weiteren klinisch-theoretischen Fächer führend mit nahezu einem Viertel des gesamten Publikationsoutputs. Wie an Fakultät A und B, belegt die Innere Medizin an der Fakultät C den zweiten Rang.

Die folgenden drei Abbildungen bieten eine Übersicht über das Publikationswachstum an den jeweiligen Fakultäten A, B und C in den Jahren 2007–2018, unterteilt nach Fächergruppen. Aus Abb. 1 lässt sich für Fakultät A entnehmen, dass die Theoretischen Fächer, die Innere Medizin und die Weiteren klinischen Fächer führend beim Publikationsoutput sind und insbesondere seit 2012 einen starken Anstieg erfahren haben. Die Zahnmedizin hat den geringsten Publikationsoutput.

Abb. 2 zeigt das jährliche Publikationswachstum in den zehn unterschiedenen Fächergruppen an der Fakultät B. Anders als an Fakultät A ist der Publikationsoutput geringer, sodass die y-Achse nur von 0 bis 300 Zeitschriftenaufsätze reicht. An der Fakultät B haben die Neuromedizinischen Fächer trotz Auf und

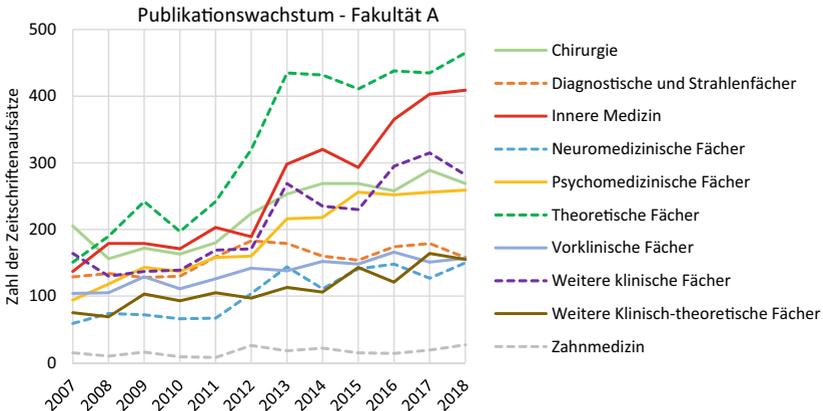


Abb. 1 Publikationswachstum an der Fakultät A in den Jahren 2007–2018, unterteilt nach Fächergruppen

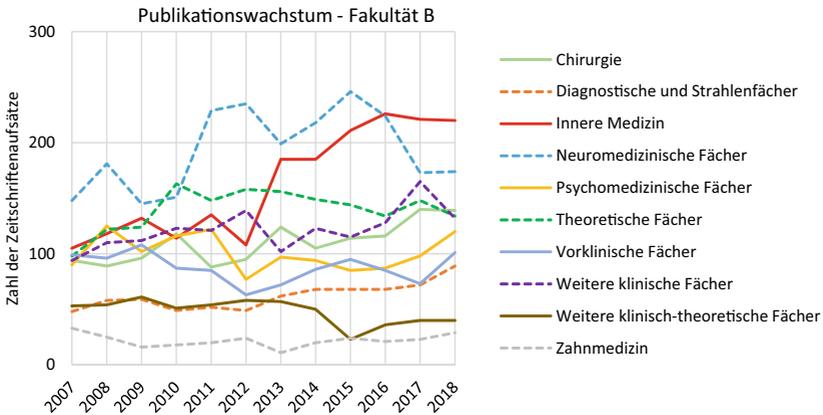


Abb. 2 Publikationswachstum an der Fakultät B in den Jahren 2007–2018, unterteilt nach Fächergruppen

Ab im Laufe des untersuchten Zeitraums den höchsten Publikationsoutput. Ähnlich wie an Fakultät A, hat die Innere Medizin einen rasanten Anstieg seit 2012 erfahren. Die Zahnmedizin bewegt sich auch hier im unteren Bereich.

Abb. 3 verdeutlicht, dass auch an der Fakultät C die Innere Medizin führend ist. Die meisten Fächergruppen sind gekennzeichnet durch Schwankungen in ihrem jährlichen Output. Die Vorklinischen Fächer haben zwar den geringsten Publikationsoutput, haben jedoch eine kontinuierliche Zunahme erfahren.

2.2 Buchbeiträge und Monographien

An den Fakultäten A und B werden nicht nur Zeitschriftenaufsätze erfasst, sondern auch Buchbeiträge und Monographien. An der Fakultät C hingegen werden nur Zeitschriftenaufsätze im Rahmen der LOM berücksichtigt. Im LOM-System der Fakultät A wird keine Differenzierung zwischen Monographien und Buchbeiträgen vorgenommen, jedoch zeigt die Analyse der Seitenangaben in den von den Fakultäten bereitgestellten Daten, dass es sich überwiegend um Buchbeiträge handelt. In dem untersuchten Zeitraum 2007–2018 gibt es an der Fakultät A insgesamt 1831 Buchbeiträge, was einem Anteil von 6,6 % entspricht (bezogen auf 27.653 Publikationen aller Dokumententypen in 2007–2018). An der Fakultät

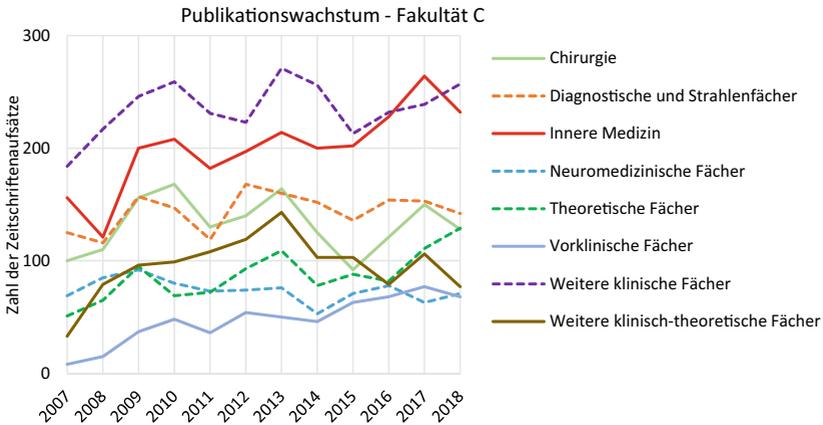


Abb. 3 Publikationswachstum an der Fakultät C in den Jahren 2007–2018, unterteilt nach Fächergruppen

B wird neben dem Publikationstyp Zeitschriftenaufsatz, auch zwischen Buchbeiträgen und Monographien unterschieden. Im Zeitraum 2007–2018 wurden an der Fakultät B 128 Monographien und 1796 Buchbeiträge erfasst. Diese Publikationstypen wurden für nachfolgende Analysen zusammengenommen (1924) und machen 11,3 % des gesamten Publikationsaufkommens der Fakultät B aus (16.988 in 2007–2018). Tab. 2 zeigt die Verteilung der Buchbeiträge und Monographien über die verschiedenen Fächergruppen an den Fakultäten A und B.

Es ist erkennbar, dass an beiden Fakultäten die Psychomedizinischen Fächer zu einem großen Teil ihre Forschungsergebnisse in Form von Buchbeiträgen veröffentlichen. An beiden Fakultäten finden sich die geringsten Prozentsätze in den Diagnostischen und Strahlenfächern, der Inneren Medizin und den Neuromedinischen Fächern. An der Fakultät A schwankt der Anteil der Buchbeiträge als Forschungsoutput zwischen 1,3 % und 15,7 %. An der Fakultät B liegt der Anteil der Buchbeiträge zwischen 4,6 % und 21 % und suggeriert, dass Buchbeiträge an Fakultät B einen höheren Stellenwert haben. Wie sich die Zahl der Buchbeiträge an den Fakultäten A und B über die Jahre entwickelt hat, gilt es anhand Abb. 4 zu illustrieren.

Aus der Abbildung geht hervor, dass der Anteil an Buchbeiträgen an der Fakultät A sich von 8,5 % in 2009 auf 6,5 % in 2018 verringert hat. An der Fakultät B wird der Publikationstyp Buchbeiträge erst seit 2009 erfasst. Der Anteil der

Tab.2 Überblick über den Anteil von Buchbeiträgen und Monographien am gesamten Publikationsaufkommen an den Fakultäten A und B im Zeitraum 2007–2018, unterteilt nach Fächergruppen

Fächergruppe	Fakultät A	Fakultät B
Chirurgie	1,8%	9,3%
Diagnostische und Strahlenfächer	1,3%	5,4%
Innere Medizin	2,6%	4,6%
Neuromedizinische Fächer	4,4%	5,5%
Psychomedizinische Fächer	15,7%	21,0%
Theoretische Fächer	7,8%	16,8%
Vorklinische Fächer	11,5%	11,2%
Weitere klinische Fächer	3,8%	16,8%
Weitere klinisch-theoretische Fächer	6,6%	12,5%
Zahnmedizin	3,7%	6,7%
Gesamtzahl an Buchbeiträgen/Monographien	1831	1924

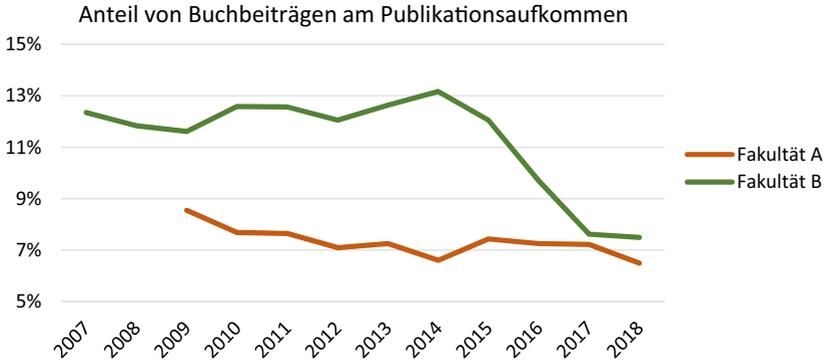


Abb. 4 Anteil der Buchbeiträge am Publikationsaufkommen an den Fakultäten A und B im Zeitraum 2007–2018, unter Verwendung von gleitenden 3-Jahres-Mittel

Buchbeiträge lag im Zeitraum 2007–2014 relativ konstant zwischen 11,6 % und 13,2 %, jedoch gab es seit 2015 einen deutlichen Abwärtstrend.

3 Rezeptionsanalyse

Die Publikationsanalyse hat gezeigt, dass jährliche Publikationszahlen in einigen Fächergruppen gering ausfallen und Schwankungen aufweisen. Da Zitationszahlen noch stärkeren Schwankungen unterliegen können, wurden in der nun folgenden Rezeptionsanalyse die zwölf Publikationsjahre 2007–2018 zu sechs Blöcken mit je zwei Jahren aggregiert. Die Publikationen, die in der Rezeptionsanalyse berücksichtigt werden, entsprechen den in der Tab. 1 aufgeführten Zeitschriftenaufsätzen (s. Abschn. 2.1). Das Zitationsfenster beträgt drei Jahre und reicht somit für Publikationen aus dem Jahr 2018 von 2018 bis 2020.

3.1 Zitationen pro Publikation

Ein einfaches Rezeptionsmaß ist die erzielte Zitationszahl innerhalb von drei Jahren nach Erscheinen der Publikation, kurz CPP (Citations Per Paper). In Abb. 5 sind die durchschnittlichen Zitationszahlen an den drei Fakultäten aufgeführt.

Aus der Abbildung wird ersichtlich, dass Zitationszahlen fachspezifisch sind. An Fakultät A weist die Innere Medizin mit Abstand die höchsten CPP-Werte auf. Die niedrigsten CPP-Werte sind in der Zahnmedizin zu finden, gefolgt von den Psychomedizinischen Fächern.

An der Fakultät B weisen die Innere Medizin, die Neuromedizinischen Fächer, die Psychomedizinischen Fächer und die Vorklinischen Fächer vergleichsweise hohe CPP-Werte auf. Eine Ausnahme stellt erneut die Zahnmedizin dar. An der Fakultät C haben die Innere Medizin und die Weiteren klinischen Fächer hohe CPP-Werte. Auffällig ist der Anstieg des CPP der Publikationen in den Weiteren klinisch-theoretischen Fächern in 2017–2018 und der Anstieg der Zitationszahlen für Publikationen aus 2015–2016 in den Weiteren klinischen Fächern, gefolgt von einem Absinken für Publikationen aus 2017–2018.

3.2 Prozentsatz unzitierter Publikationen

Der Prozentsatz unzitierter Publikationen gibt Auskunft darüber, welcher Anteil an Publikationen drei Jahre nach Veröffentlichung nicht in anderen Publikationen

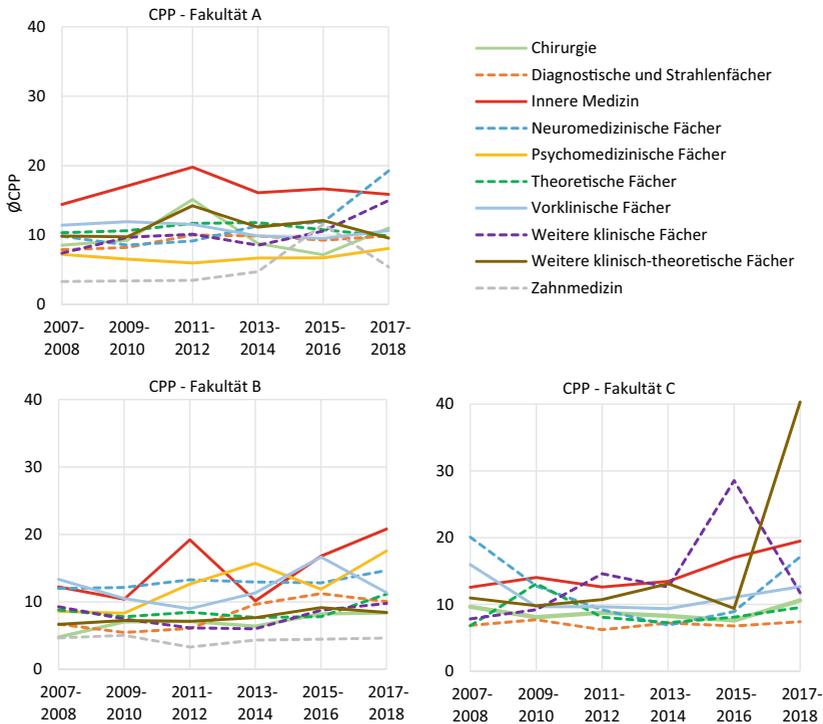


Abb. 5 Überblick über den Indikator CPP (Citations Per Paper) an den Fakultäten A, B und C, unterteilt nach Fächergruppen

aufgegriffen wurde, was auf eine geringe Relevanz für weitere Forschung hindeuten kann. Abb. 6 zeigt für alle drei Fakultäten den Prozentsatz an Publikationen, die im Jahr der Veröffentlichung und in den folgenden zwei Jahren nicht zitiert wurden.

Aus der Abbildung kann man entnehmen, dass an allen Fakultäten A, B und C die Vorklinischen Fächer den geringsten Prozentsatz an unzitierten Publikationen in WoS haben. An der Fakultät A blieben im Durchschnitt 6 % der publizierten Aufsätze in den Vorklinischen Fächern drei Jahre nach ihrer Publikation unzitiert. Der Anteil an unzitierten Publikationen in der Zahnmedizin mag hoch sein, jedoch ist die Zahl aller Publikationen aus dem Zeitraum gering. An der Fakultät B sind es neben den Vorklinischen Fächern die Neuromedizinischen Fächer,

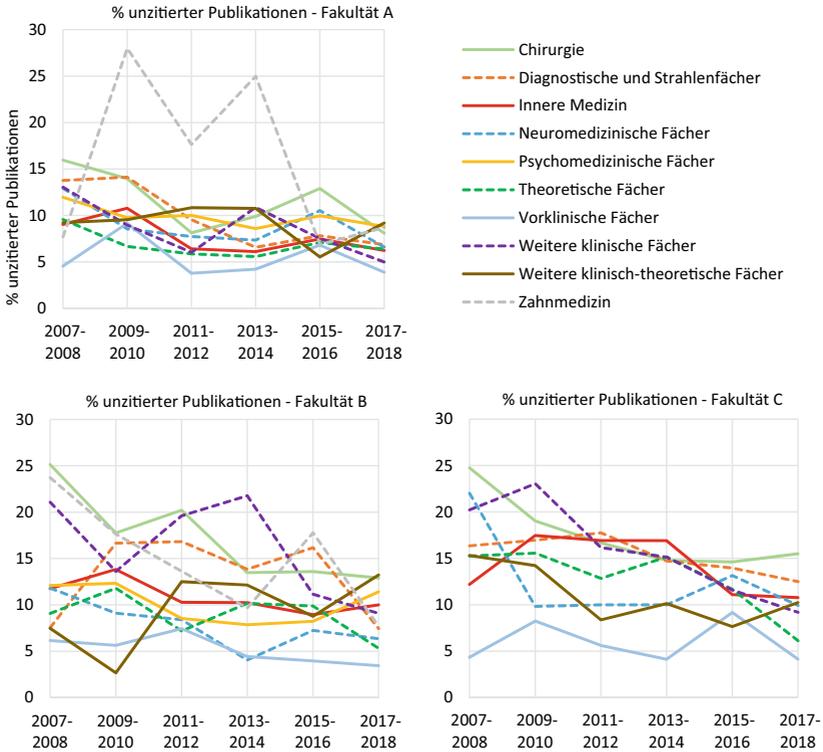


Abb. 6 Überblick über den Prozentsatz unzitierter Publikationen an den Fakultäten A, B und C, unterteilt nach Fächergruppen

die den geringsten Prozentsatz an unziterten Publikationen in WoS aufweisen. An der Fakultät C kann man in allen Fächergruppen einen leichten Abwärtstrend des Prozentsatzes unzitierter Publikationen feststellen. Somit lässt sich festhalten, dass die gestiegenen Publikationszahlen an allen drei Fakultäten nicht vermehrt zu Veröffentlichungen geführt haben, die für die Forschung von geringer Relevanz sind und unzitirt blieben.

3.3 Feldnormalisierte Zitationsrate MNCS

Die mittlere normalisierte Zitationsrate (MNCS – Mean Normalized Citation Score) zeigt, ob Publikationen einer Fakultät über- oder unterdurchschnittlich oft zitiert werden, indem tatsächliche Publikationszahlen mit erwartbaren Zitationszahlen in Beziehung gesetzt werden. Die Normalisierung erfolgt unter Berücksichtigung des gleichen Publikationsjahres, des gleichen Publikationstyps und der Zuordnung zur gleichen WoS-internen Fächerklassifikation. Bei der Berechnung des MNCS wird zunächst das Verhältnis zwischen den tatsächlichen und den durchschnittlich erwartbaren Zitationszahlen für jede einzelne Publikation erfasst. Erst dann wird der Durchschnitt der Verhältnisse einzelner Publikationen berechnet. Die Berechnung des MNCS ist komplex, informiert aber gut über die tatsächliche Leistung der Fakultäten im internationalen Vergleich unter Berücksichtigung von Fächerunterschieden. In Abb. 7 sind die mittleren normalisierten Zitationsraten der Fakultäten A, B und C dargestellt. Der sog. Expected Citation Score (ECS) fungiert als Benchmark und entspricht 1,0.

Abb. 7 zeigt, dass die meisten Fächergruppen über alle drei Fakultäten hinweg eine exzellente Leistung im globalen Vergleich aufweisen. Der MNCS liegt nahezu flächendeckend über dem durchschnittlich zu erwartenden Wert 1. An der Fakultät A weist die Innere Medizin einen Peak von MNCS > 5 für Publikationen der Jahre 2011–12 auf. Das bedeutet, dass diese Publikationen im globalen Vergleich fünfmal so viele Zitierungen erhalten haben, wie erwartbar gewesen wäre. Was die überdurchschnittliche Rezeption von Publikationen in der Inneren Medizin betrifft, sind die MNCS-Werte im Einklang mit den CPP-Werten.

An der Fakultät B hat die Innere Medizin den höchsten MNCS, während die übrigen Fächergruppen eine konstant gute Leistung in dem untersuchten Zeitraum aufweisen. Lediglich die Zahnmedizin hat MNCS-Werte kleiner eins und wird somit im globalen Vergleich seltener zitiert. An der Fakultät C schneiden Publikationen aller Fächergruppen überdurchschnittlich gut ab. In der Inneren Medizin wurden Publikationen ungefähr dreimal so oft zitiert wie vergleichbare Publikationen. Einen steilen Anstieg am MNCS haben Publikationen in den Weiteren klinisch-theoretischen Fächern erfahren, was wiederum im Einklang mit einem gestiegenen CPP für 2017–2018 ist (vgl. Abb. 5).

3.4 PP(top10 %)-Indikator

Wie der MNCS, so ist auch der PP(top10 %)-Indikator feldnormalisiert und kann somit als valider Leistungsindikator fungieren. Dabei steht das PP für Proportion

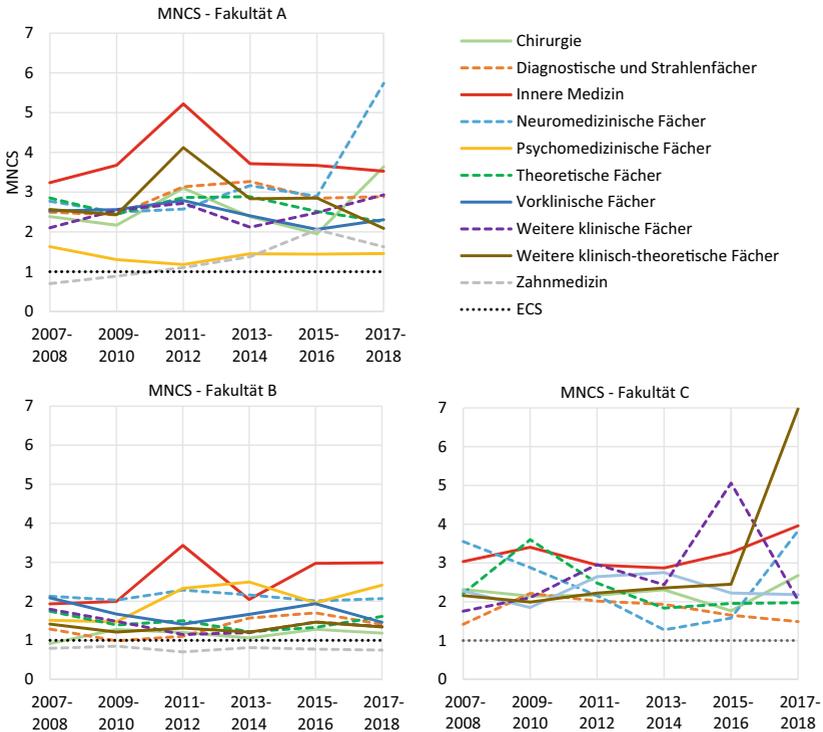


Abb. 7 Überblick über den MNCR an den Fakultäten A, B und C, unterteilt nach Fächergruppen

of Publications und die 10 % bezieht sich auf die 10 % bestzitierten Publikationen (geläufig sind auch top5 % und top1 %). Der PP(top10 %)-Indikator gibt den Prozentsatz der Publikationen aus einer Publikationsmenge an, die im internationalen Vergleich mit anderen WoS-Publikationen gleichen Dokumenttyps, Publikationsjahres und Themenfeldes zu den 10 % am häufigsten zitierten gehören. Es ist zunächst von der Erwartung auszugehen, dass 10 % der Publikationen einer Fächergruppe zu den 10 % meistzitierten eines Feldes gehören. Wenn der Prozentsatz der Publikationen einer Fächergruppe höher ausfällt, so schneiden diese besser als erwartet ab, während ein PP(top10 %) von unter 10 % eine mindere Leistung als die zu erwartende aufzeigt. Abb. 8 illustriert, dass die Publikationen

der Fakultäten A, B und C im internationalen Vergleich überdurchschnittlich gut abschneiden.

An allen drei Fakultäten gibt es in der Inneren Medizin den höchsten Prozentsatz an Publikationen, die zu den meistzitierten ihres Feldes gehören. An der Fakultät A liegen die Werte im gesamten Zeitraum bei ca. 25 %, d. h. jede vierte Publikation in der Inneren Medizin gehört zu den 10 % am meisten zitierten. Nur in den Psychomedizinischen Fächern und der Zahnmedizin ist der PP(top10 %)-Indikator zeitweise unter der 10 %-Marke, d. h. die Publikationen der Fakultät A werden in diesen zwei Fächern unterdurchschnittlich oft zitiert. An Fakultät B sind es die Neuromedizinischen Fächer, die im untersuchten Zeitraum einen konstanten PP(top10 %)-Wert von ca. 24 % aufweisen. Die übrigen Fächergruppen

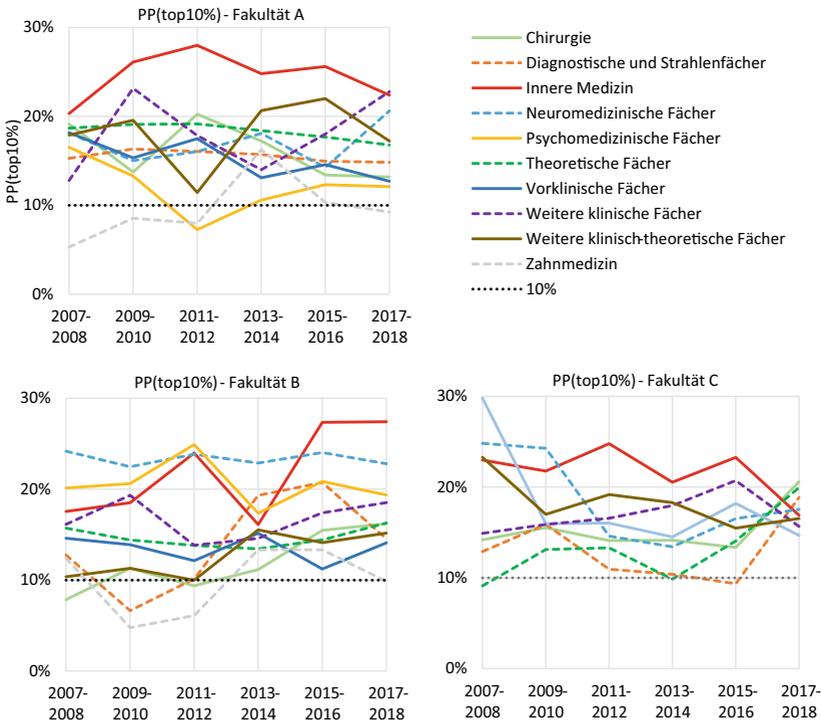


Abb. 8 Überblick über den PP(top10 %)-Indikator an den Fakultäten A, B und C, unterteilt nach Fächergruppen

weisen starke Schwankungen auf. In den Diagnostischen und Strahlenfächern und der Zahnmedizin ist der PP(top10 %)-Indikator kurzzeitig unter der 10 %-Marke. An der Fakultät C werden die meisten Fächergruppen überdurchschnittlich oft zitiert und zeigen verschiedene Trends über die Jahre.

3.5 Journal Impact Factor (JIF)

Die LOM-Systeme an den meisten medizinischen Fakultäten in Deutschland basieren auf dem Journal Impact Factor. Dieser wird jahresweise für wissenschaftliche Zeitschriften berechnet und gibt an, wie oft Beiträge einer Zeitschrift aus den vorherigen Jahren im darauffolgenden Jahr zitiert wurden. Das Zitationsfenster ist – anders als beim Indikator CPP – nur ein Jahr lang.

Abb. 9 zeigt die durchschnittlichen Impact-Faktoren der Zeitschriften, in denen Autor*innen aus den Fakultäten A, B und C publiziert haben. Autor*innen der Inneren Medizin an Fakultät A publizierten in Zeitschriften, die den höchsten durchschnittlichen JIF aufweisen.

Die niedrigsten durchschnittlichen Impact-Faktoren gibt es in der Zahnmedizin und den Psychomedizinischen Fächern. An der Fakultät B sind es die Publikationen der Vorklinischen Fächer und der Inneren Medizin, die in Zeitschriften mit einem hohen durchschnittlichen JIF erschienen sind. An der Fakultät C sind es die Innere Medizin, die Vorklinischen und die Weiteren klinischen-theoretischen Fächer, die in Zeitschriften mit hohen Impact-Faktoren publizierten.

3.6 Publikationssprache und Impact

In diesem Abschnitt gilt es aufzuzeigen, inwiefern die Sprache der Publikationen mit dem Zitations-Impact zusammenhängt und welcher Trend sich in den letzten Jahren abzeichnet. In WoS werden nicht nur englischsprachige Zeitschriften erfasst, sondern auch deutschsprachige Zeitschriften, sofern die Aufsätze einen Abstract in englischer Sprache zur Verfügung stellen. Es gibt jedoch auch deutschsprachige Zeitschriften, die englischsprachige Aufsätze publizieren. Die Identifizierung von deutschsprachigen Aufsätzen basiert auf der Angabe in WoS über die Sprache eines Aufsatzes. Absolut gesehen hat Fakultät A mit 1.471 deutschsprachigen Publikationen zwischen 2007 und 2018 die meisten vorzuweisen, während Fakultäten B und C sich mit ca. 1.000 Publikationen in der gleichen Größenordnung bewegen. In Tab. 3 sind die prozentualen Anteile der deutschsprachigen Publikationen nach Fächergruppe gelistet.

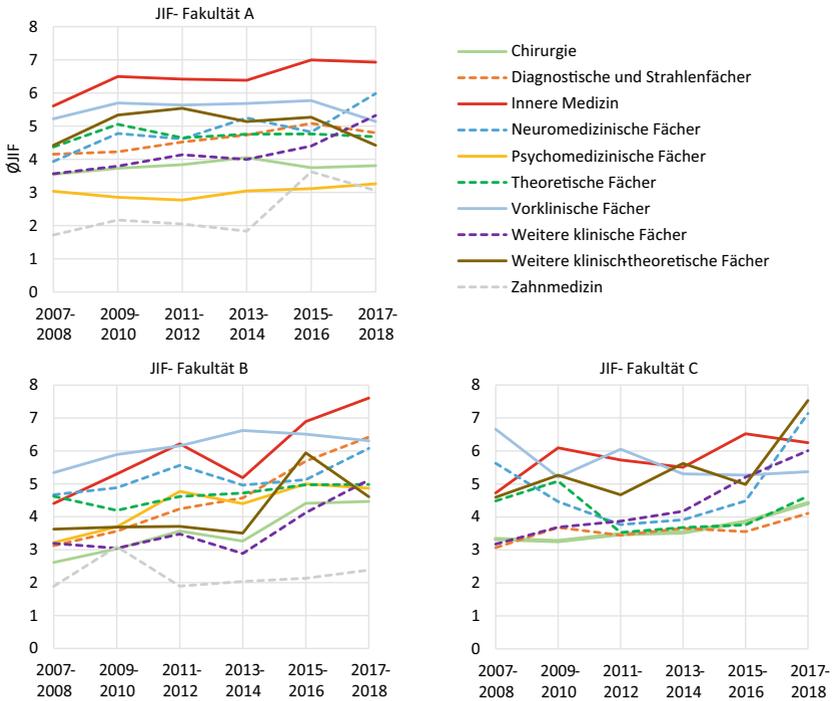


Abb. 9 Überblick über den JIF (Journal Impact Factor) an den Fakultäten A, B und C, unterteilt nach Fächergruppen

Aus Tab. 3 geht hervor, dass jede der drei untersuchten Fakultäten in einer anderen Fächergruppe den höchsten Prozentsatz an deutschsprachigen Publikationen aufweist. Fakultät A hat die meisten deutschsprachigen Publikationen in den Psychomedizinischen Fächern, Fakultät B in den Weiteren klinischen Fächern und Fakultät C in der Chirurgie. Gemeinsam ist allen drei Fakultäten, dass der Prozentsatz an deutschsprachigen Publikationen am geringsten in der Zahnmedizin und den Vorklinischen Fächern ist, gefolgt von den Neuromedizinischen und den Theoretischen Fächern. Während Tab. 3 die Verteilung für den gesamten Zeitraum 2007–2018 wiedergibt, ist in Abb. 10 die Entwicklung des Anteils deutschsprachiger Publikationen aufgeführt.

Die Abbildung verdeutlicht, dass es an allen drei Fakultäten einen Rückgang an deutschsprachigen Publikationen gab. An der Fakultät A hat sich der Anteil

Tab. 3 Überblick über den Anteil an deutschsprachigen Zeitschriftenaufsätzen im Zeitraum 2007–2018, unterteilt nach Fächergruppen

Fächergruppe	Fakultät A	Fakultät B	Fakultät C
Chirurgie	6%	9%	14%
Diagnostische und Strahlenfächer	5%	7%	9%
Innere Medizin	5%	6%	6%
Neuromedizinische Fächer	4%	3%	5%
Psychomedizinische Fächer	19%	12%	
Theoretische Fächer	5%	4%	5%
Vorklinische Fächer	3%	0%	1%
Weitere klinisch-theoretische Fächer	5%	6%	5%
Weitere klinische Fächer	10%	26%	12%
Zahnmedizin	2%	2%	
Gesamtzahl deutschsprachiger Publikationen	1471	1004	1012

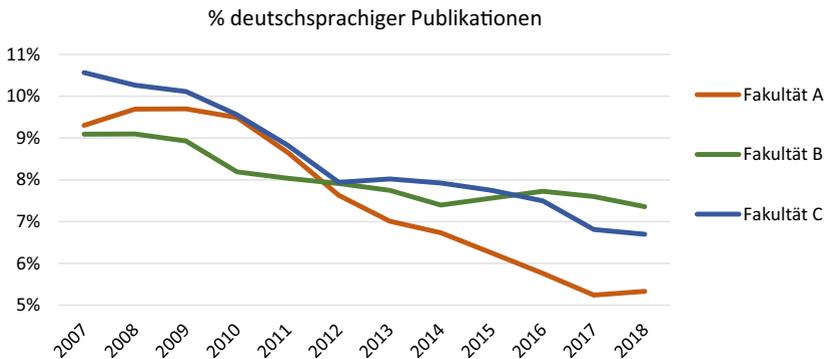


Abb. 10 Entwicklung des Prozentsatzes an deutschsprachigen Publikationen im Zeitraum 2007–2018 an den Fakultäten A, B und C, unter Verwendung von gleitenden 3-Jahres-Mittel

von 9,7 % in 2008 auf 5,4 % in 2018 halbiert. An der Fakultät B ist der Rückgang nicht ganz so stark. Waren es im Jahr 2007 noch 9,1 %, so sind es im Jahr 2018 noch 7,4 %. Fakultät C wiederum hatte in 2007 einen Anteil von 10,5 % an deutschsprachigen Publikationen, jedoch ist der Anteil auf knapp 6,7 % in 2018 gesunken.

Eine Erklärung für den sinkenden Anteil an deutschsprachigen Publikationen könnten die niedrigeren Impact-Faktoren deutschsprachiger Zeitschriften im Vergleich zu englischsprachigen sein, falls Autor*innen die Publikationsorte aufgrund der LOM vorrangig nach dem JIF auswählen. Dies wird in Abb. 11 deutlich, die die durchschnittlichen Impact-Faktoren von deutschsprachigen und englischsprachigen Publikationen an den drei Fakultäten gegenüberstellt.

Der durchschnittliche JIF der Zeitschriften, in denen deutschsprachige Aufsätze erschienen sind, bewegt sich an allen drei Fakultäten zwischen 0,5 und

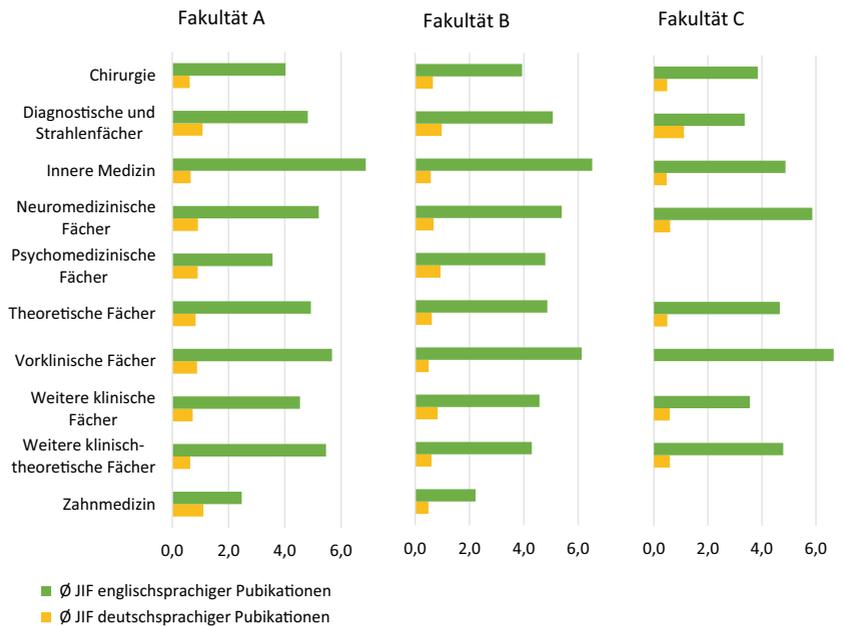


Abb. 11 Durchschnittlicher JIF in Abhängigkeit von der Sprache der Publikationen an den Fakultäten A, B und C für Publikationen in 2007–2018, unterteilt nach Fächergruppen

1,1, für englischsprachige Zeitschriften hingegen zwischen 2,2 und 6,7. Ein unintendierter Effekt der LOM, welche über ein Punktesystem die Publikation in hochzitierten Zeitschriften belohnt, könnte darin liegen, dass Autor*innen animiert werden in englischsprachigen Zeitschriften zu publizieren, da diese mehr JIF-Punkte bringen als deutschsprachige. Unter den deutschsprachigen Zeitschriften finden sich u. a. Die Anaesthesiologie/Der Anästhesist, Die Dermatologie/Der Hautarzt, HNO, Der Nervenarzt, Psychiatrische Praxis, Schmerz, Strahlentherapie und Onkologie und Der Urologe/Die Urologie.

4 Fazit

Dieser Beitrag hatte die Publikationen von drei medizinischen Fakultäten in Deutschland und deren Impact im Fokus. Die Publikationsanalyse hat gezeigt, dass es an allen drei Fakultäten eine Zunahme an Zeitschriftenaufsätzen im untersuchten Zeitraum 2007–2018 gab. Die Datenlage zeigt, dass nur die Fakultäten A und B Monographien und Beiträge in Lehr- und Handbüchern erfassen und in der LOM berücksichtigen. Jedoch ist der Anteil an Buchbeiträgen an den Fakultäten A und B über die Jahre gesunken, was ein unintendierter Effekt der LOM sein könnte, da Buchbeiträge verglichen mit Aufsätzen in hochzitierten internationalen Zeitschriften weniger Impact-Faktor-Punkte einbringen. Zum anderen entziehen sich Monographien der bibliometrischen Evaluation, allein aufgrund ihres über Jahre dauernden Rezeptionszeitraums (Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022).

Der CPP-Indikator in der Rezeptionsanalyse hat gezeigt, dass sich die absoluten Zitationszahlen in den Fächern unterscheiden, über die Fakultäten hinweg jedoch in der gleichen Größenordnung bewegen. Der Prozentsatz unzitierter Publikationen ist für alle drei Fakultäten über die Jahre gesunken. Dem MNCS und dem PP(top10 %)-Indikator nach, weisen die meisten Fächergruppen der drei Fakultäten im internationalen Vergleich einen exzellenten Impact auf. Diesen vier Impactmaßen zufolge erfahren Publikationen der Inneren Medizin die höchste Rezeption, während Publikationen der Zahnmedizin unterdurchschnittlich rezipiert werden und zu einem großen Teil unzitiert bleiben. Ebenso unterscheidet sich der durchschnittliche Journal Impact Factor (JIF) der Zeitschriften, in denen publiziert wird, innerhalb der Fakultäten.

Die Ergebnisse der Rezeptionsanalyse haben verdeutlicht, wie sehr sich das Zitationsverhalten unter den Fächergruppen unterscheidet und wie ungerecht der Gebrauch des gleichen Indikators für Fächergruppen mit unterschiedlicher Publikations- und Zitationskultur ist. Während Ergebnisse der Grundlagenfächer

in international hochzitierten Zeitschriften erscheinen, ist anwendungsorientiertes klinisches Wissen kulturkreisspezifisch und wird vermehrt in nationalen Zeitschriften publiziert, die seltener zitiert werden (Meenen, 1997).

Die Analyse der im Web of Science (WoS) verzeichneten deutschsprachigen Aufsätze zeigt einen Rückgang an allen drei Fakultäten im Zeitraum 2007–2018. Ein unintendierter Effekt der LOM könnte somit der Rückgang deutschsprachiger Aufsätze zugunsten von englischsprachigen Aufsätzen sein, da letztere höhere JIF-Werte aufweisen. In einigen Fächergruppen wie der Chirurgie, den psychomedizinischen und den klinischen Fächern sind Fachzeitschriften in deutscher Sprache für die Umsetzung relevanter Forschungsergebnisse und klinischer Daten unverzichtbar (Kaltenborn & Kuhn, 2003). Daher wäre zu wünschen, dass LOM-Systeme dahingehend umgestaltet werden, dass Lehr- und Handbücher als auch deutschsprachige Aufsätze eine Aufwertung erfahren und nicht unter dem Diktat des Journal Impact Factors einen Rückgang erfahren. Überhaupt ist die klinische Medizin eine nationale Medizin, in der deutschsprachige Publikationen ihre Berechtigung haben (Vahl 2008). Für die informationelle Versorgung der Bevölkerung mit medizinischem Wissen sollte dieses auch in der Muttersprache zur Verfügung stehen (Kaltenborn & Kuhn, 2006).

Bereits 1997 wurde bei der Delegiertenkonferenz der AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften) von der Kommission „Bibliometrie“ die Anwendung des JIF bei der LOM kritisch hinterfragt und Empfehlungen herausgegeben. Diese sahen unter anderem vor bei deutschsprachigen Zeitschriften den Impact-Faktor doppelt zu gewichten und Originalarbeiten aus Zeitschriften, die nicht in Web of Science gelistet sind und somit keinen JIF aufweisen, mit einem „äquivalenten Impact-Faktor“ von 0,2 zu werten. Nur wenige medizinische Fakultäten haben diese Empfehlung befolgt. Hingegen haben die von der Senatskommission für klinische Forschung der Deutschen Forschungsgemeinschaft im Jahr 2004 herausgegebenen Empfehlungen für eine leistungsorientierte Mittelvergabe zu einer nachhaltigen Zementierung des Impact-Faktor-Systems geführt (Brähler & Strauß, 2009).

Literatur

Brähler, E., & Strauß, B. (2009). Leistungsorientierte Mittelvergabe an Medizinischen Fakultäten: Eine aktuelle Übersicht. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 52(9), 910–916. <https://doi.org/10.1007/s00103-009-0918-1>.

- Deutsche Forschungsgemeinschaft | AG Publikationswesen. (2022). *Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung*. <https://doi.org/10.5281/ZENODO.6538163>.
- Felt, U. (1999). Evaluation im wissenschaftspolitischen Kontext. In Qualitätsförderung durch Evaluation? Ziele, Aufgaben und Verfahren von Forschungsbewertungen im Wandel. Dokumentation des Workshops vom 20. und 21. Mai 1999 (S. 11–31).
- Hostettler, S., Kraft, E., & Bosshard, C. (2015). Medizinische Forschung – Qualität vor Quantität. *Schweizerische Ärztezeitung*, 96(49), 1794–1799. <https://doi.org/10.4414/saez.2015.04171>.
- Joshi, M. A. (2014). Bibliometric indicators for evaluating the quality of scientific publications. *The Journal of Contemporary Dental Practice*, 15(2), 258–262. <https://doi.org/10.5005/jp-journals-10024-1525>.
- Kaltenborn, K.-F., & Kuhn, K. (2003). Der Impact-Faktor als Parameter zur Evaluation von Forscherinnen/Forschern und Forschung. *Klinische Neuroradiologie*, 13(4), 173–193. <https://doi.org/10.1007/s00062-003-4364-4>.
- Meenen, N. M. (1997). Der Impact-Faktor—Ein zuverlässiger scientometrischer Parameter? *Unfallchirurgie*, 23(4), 128–134. <https://doi.org/10.1007/BF02630217>.
- PLOS Medicine. (2006). *The Impact Factor Game* (PLOS Medicine). <https://doi.org/10.1371/journal.pmed.0030291>.
- Seglen, P. O. (1998). Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthopaedica Scandinavica*, 69(3), 224–229. <https://doi.org/10.3109/17453679809000920>.
- Vahl, Ch.-Fr. (2008). Forschungsförderung durch leistungsorientierte Mittelvergabe (LOM): Argumente für eine medizinische Wissenschaftskultur jenseits der Impact-Punkte. *Zeitschrift für Herz-, Thorax- und Gefäßchirurgie*, 22(2), 94–97. <https://doi.org/10.1007/s00398-008-0630-8>.

Aman, Valeria Valeria Aman studierte Bibliotheks- und Informationswissenschaft an der Humboldt-Universität zu Berlin und an der Royal School of Library and Information Science in Kopenhagen. Seit 2013 beteiligt sie sich am Kompetenznetzwerk Bibliometrie an der Entwicklung einer robusten Infrastruktur für bibliometrische Anwendungen. Nach dem Wechsel an das DZHW in 2016 hat sie diverse bibliometrische Studien und Evaluationen durchgeführt und beendet aktuell ihre Promotion zur Entwicklung bibliometrischer Methoden zum Nachweis von Wissenstransfer.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Altmetrics zur Evaluation medizinischer Forschung in Deutschland

Steffen Lemke, Anne Witthake und Isabella Peters

1 Einleitung

Mit der zunehmenden Verbreitung wissenschaftlicher Publikationen in elektronischen Formaten sowie der immer stärkeren Durchsetzung des Berufsalltags vieler Forschender mit webbasierten Tools und Plattformen gehen rasant anwachsende Möglichkeiten einher, Interaktionen mit wissenschaftlichen Erzeugnissen in Online-Umgebungen zu messen. Aus den verschiedenen Versuchen, einen solche Messungen zusammenfassenden Begriff zu etablieren, hat sich das Kofferwort *Altmetrics* – für *alternative metrics* – weitgehend durchgesetzt (Priem et al., 2010). Hinsichtlich einer exakten und allgemeingültigen Definition befinden sich Altmetrics noch immer in einer Findungsphase (Glänzel & Gorraiz, 2015; Haustein, 2016). Doch einem weiten Teilen der einschlägigen Literatur folgendes Verständnis nach umschließen Altmetrics nicht nur Messungen der Rezeption wissenschaftlicher Veröffentlichungen auf Plattformen des Social Webs, also beispielsweise in sozialen Netzwerken, Blogs oder Wikis, sondern

S. Lemke (✉) · A. Witthake · I. Peters
ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel, Deutschland
E-Mail: sle@informatik.uni-kiel.de

A. Witthake
E-Mail: a.witthake@zbw-online.de

I. Peters
E-Mail: ipe@informatik.uni-kiel.de

S. Lemke · I. Peters
Christian-Albrechts-Universität zu Kiel, Kiel, Deutschland

auch in anderen nicht-wissenschaftlichen Kontexten, wie beispielsweise in journalistischen Medien oder Policy-Dokumenten (siehe auch Sugimoto et al., 2017). Ursprünglich von Priem et al. (2010) in erster Linie als hilfreiche Filter im Umgang mit dem stetig anwachsenden akademischen Publikationsaufkommen propagiert, ist das wissenschaftliche Interesse an Altmetrics häufig auch durch die Hoffnung motiviert, diese neue Familie von Indikatoren für den Einfluss von Publikationen könnte auch in Evaluationen wissenschaftlicher Forschung nützlich sein. Insbesondere die Aussicht, durch Altmetrics in evaluativen Szenarien vielfältigere Formen wissenschaftlicher Leistungen sowie deren Einfluss in nicht-akademischen Sphären besser abbilden zu können (Tahamtan & Bornmann, 2020), als ein nahezu ausschließlich auf zitationsbasierte Indikatoren ausgelegtes Bewertungssystem dies vermag, lässt Altmetrics als vielversprechendes Komplement zu bestehenden Methoden quantitativer Wissenschaftsbewertung erscheinen (Wouters & Costas, 2012).

Für die medizinische Forschung im Besonderen erscheint eine genauere Untersuchung des Potenzials von Altmetrics als komplementierende Bewertungsmaße aus mehreren Gründen naheliegend. Zum einen regt die tiefgreifende Kritik an der prominenten Stellung des Journal Impact Factors im Kontext der leistungsorientierten Mittelvergabe (s. auch die Beiträge von Aman und Herrmann-Lingen in diesem Band) die Suche nach ergänzenden Metriken an – insbesondere im Lichte des in vergangenen Analysen der Beziehung zwischen bibliometrischen Indikatoren und Forschungsqualität gezogenen Schlusses, dass Qualität in Evaluationen als multidimensionales Konzept begriffen werden sollte, welches sich unmöglich durch eine einzelne Metrik adäquat abbilden lässt (Wilsdon et al., 2015). Darüber hinaus zeigen Fallstudien, dass biomedizinische Forschung eine vergleichsweise hohe Präsenz auf verschiedenen altmetrischen Plattformen wie zum Beispiel sozialen Netzwerken aufweist (Costas et al., 2015; Haustein et al., 2015), eine günstige Datengrundlage also gegeben ist.

Ein altmetrischer Indikator scheint für den Anwendungsfall der Medizin besonders interessant – Erwähnungen wissenschaftlicher Artikel in Policy-Dokumenten. Verglichen mit vielen anderen wissenschaftlichen Disziplinen weist die Medizin eine hohe Anzahl von Policy-Dokumenten auf, in denen publizierte Studien direkte Verwendung finden – beispielsweise in Form sogenannter Leitlinien (s. Traylor und Aman in diesem Band). Solche Beispiele für Implementierungen wissenschaftlicher Forschungsergebnisse in konkreter Gesundheitspolitik können als Nachweise für die erfolgreiche Translation medizinischer Erkenntnisse von der Theorie in die Praxis betrachtet werden. Zitationen wissenschaftlicher Studien innerhalb solcher Policy-Dokumente könnten daher ein geeigneteres Maß

zur Erfassung der praktischen Relevanz von Forschungsergebnissen sein, als akademische Zitationen innerhalb anderer Forschungspublikationen es sind.

Die in diesem Kapitel vorgestellte Analyse verfolgt zwei primäre Ziele. Das erste Ziel ist die übergreifende Darstellung von Abdeckung und Ausprägung verschiedener populärer Altmetrics in Hinblick auf den medizinischen Forschungsoutput deutscher Autor:innen. Diese Darstellung soll bei der Beurteilung der prinzipiellen Eignung altmetrischer Indikatoren für evaluative Anwendungen im Kontext deutscher medizinischer Forschung helfen, indem ein Überblick über die zur Verfügung stehende Datengrundlage gegeben wird. Das zweite Ziel dieses Beitrags ist die detaillierte Betrachtung von Policy-Dokument-Zitationen als altmetrischem Indikator. Hierfür soll insbesondere betrachtet werden, welche Typen medizinischer Forschung durch diesen Indikator besonders gut abgebildet werden, um anschließend bestehende Chancen und Herausforderungen erörtern zu können, die sich für eine Nutzung von Policy-Dokument-Zitationen zur Beurteilung medizinischer Forschung in Deutschland ergeben.

Sämtliche in dieser Studie erhobenen altmetrischen Daten stammen vom Anbieter *Altmetric.com*. Selbstverständlich existieren andere Anbieter und Quellen zur Erhebung altmetrischer Daten mit individuellen Stärken und Schwächen (siehe Zahedi & Costas, 2018) für eine vergleichende Diskussion vier prominenter Datenanbieter) – bezüglich Policy-Dokument-Zitationen sei insbesondere auf die vergleichsweise junge Plattform *Overton*¹ hingewiesen (Szomszor & Adie, 2022). Für diese Studie fiel die Wahl der Quelle auch deshalb auf *Altmetric.com*, da sie einen technisch verhältnismäßig unkomplizierten, gut skalierbaren und damit pragmatischen Ansatz zur Beschaffung altmetrischer Daten darstellt und sich aus diesem Grund auch für eventuelle praktische Umsetzungen altmetrischer Evaluationen in der Medizin anböte.

Zur Erreichung ihrer vorgenannten Ziele widmet sich die folgende Analyse der Beantwortung zweier zentraler Forschungsfragen: Wie ist die in vergangenen Jahren unter deutscher Beteiligung entstandene medizinische Forschung in verschiedenen altmetrischen Datenquellen repräsentiert? Welche Journale und Themenfelder wissenschaftlicher Artikel sind besonders häufig mit Zitationen in Policy-Dokumenten assoziiert? Mit der Beantwortung dieser Fragen kumuliert und erweitert der vorliegende Beitrag Ergebnisse und Befunde, die in zwei vorbereitenden Studien veröffentlicht wurden – siehe auch Fraser et al. (2021) und Lemke et al. (2022).

¹ <https://www.overton.io/>

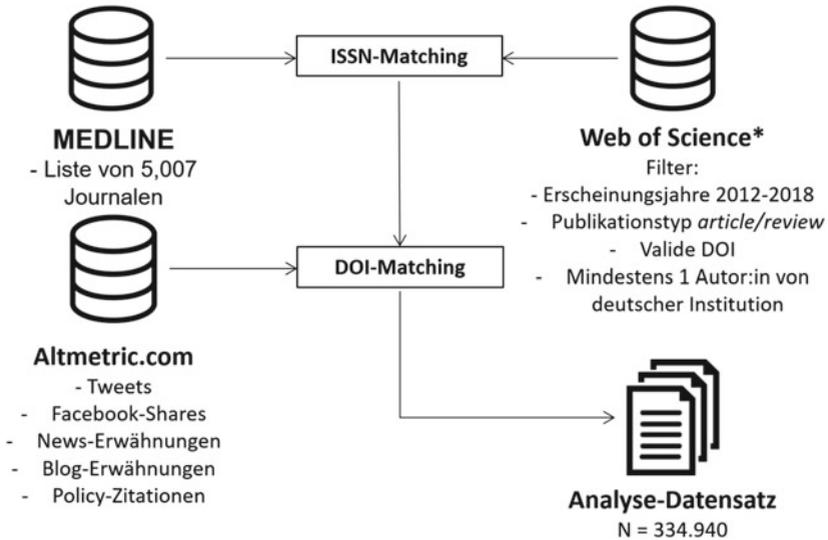


Abb. 1 Schematische Darstellung der Datensatzerstellung

2 Datensatz

Zur Erstellung eines Datensatzes unter deutscher Mitarbeit entstandener medizinischer Forschung beginnen wir mit der Liste in MEDLINE² indizierter Journale – der Ablauf ist auch in Abb. 1 dargestellt. Zum Zeitpunkt der Erhebung führte die von der US National Library of Medicine unterhaltene bibliografische Datenbank 5007 Journale auf, in denen, einem weitgefassten Verständnis der Disziplin folgend, für Forschende biomedizinischer Fächer relevante Literatur publiziert wird. Diese Journale wurden im nächsten Schritt, basierend auf ihren Titeln und ISSN-Nummern, mit dem *Web of Science* abgeglichen. Dabei wurden 16 Journale aus der Web-of-Science-Kategorie „multidisciplinary sciences“ (z. B. *Scientific Reports*, *PLOS ONE*) aufgrund ihres nicht biomedizinischen Schwerpunkts aus der Menge entfernt. Insgesamt ließen sich auf diesem Wege 4442 MEDLINE-Journale Web-of-Science-Einträgen zuordnen, welche die Basis für die weiteren Auswertungen bilden.

² https://www.nlm.nih.gov/medline/medline_overview.htm.

Im nächsten Schritt wurden aus Web of Science³ Metadaten aller Publikationen vom Typ *article* oder *review* heruntergeladen, für die eine valide DOI bekannt und an deren Erstellung mindestens ein:e Autor:in von einer deutschen Institution beteiligt war. Um die Aktualität der analysierten Forschung und gleichzeitig ausreichend große Zitationszeiträume zu gewährleisten, wurde die Ergebnismenge außerdem auf die Erscheinungsjahre 2012 bis 2018 beschränkt. Diese Filter ergaben einen finalen Datensatz aus 334.940 Artikeln.

Die Artikel betreffende altmetrische Kennzahlen wurden über die Altmetric.com-API⁴ abgerufen. Dabei wurden fünf populäre Metriken erfasst: Nennungen der Artikel auf Facebook, auf Twitter, in Blogs, auf Nachrichtenseiten sowie in Policy-Dokumenten. Da Altmetric.com nur Einträge für Artikel enthält, die mindestens eine Nennung auf mindestens einer der beobachteten Plattformen erhalten haben, wurden für alle Artikel ohne Altmetric.com-Eintrag 0 Erwähnungen auf allen Plattformen angenommen.

Um präzisere Eindrücke davon zu gewinnen, inwiefern die Abdeckungsgrade verschiedener Metriken zwischen konkreten medizinischen Forschungseinrichtungen variieren können, analysieren wir neben dem oben beschriebenen Hauptdatensatz auch die im gleichen Zeitraum von bei drei anonymen deutschen medizinischen Fakultäten ansässigen Autor:innen veröffentlichten Artikel ($N_A = 5831$, $N_B = 5581$, $N_C = 9024$) gesondert. Die Metadaten dieser Publikationen wurden uns direkt von den betreffenden Fakultäten zu Zwecken dieses Projekts zur Verfügung gestellt und unter Anwendung der gleichen Filter wie unser Hauptdatensatz mit Web of Science abgeglichen. Auch für diese Publikationen ermitteln wir altmetrische Kennzahlen über die API von Altmetric.com.

3 Methoden der Auswertung

Um das Potenzial verschiedener altmetrischer Indikatoren für das vorliegende Anwendungsgebiet deutscher medizinischer Forschung zu umreißen, ermitteln wir deskriptive Statistiken zu indikatorspezifischem Abdeckungsgrad (d. h. dem Anteil wissenschaftlicher Artikel, die mindestens eine entsprechende Nennung erhalten haben) und Ausprägung (d. h. der durchschnittlichen Häufigkeiten, mit denen einzelne Artikel Nennungen erhalten haben; Haustein et al., 2015) für die Artikel unseres Datensatzes sowie für die Publikationen der drei anonymen

³ In dieser Studie verwendete Web of Science-Daten wurden vom deutschen *Kompetenznetzwerk Bibliometrie* bezogen (<https://www.bibliometrie.info/>).

⁴ <https://www.altmetric.com/products/altmetric-api/>

medizinischen „Pilotfakultäten“. Zur darüber hinausgehenden visuellen Analyse von Indikator-Abdeckungen und -Ausprägungen nutzen wir die Software VOSviewer.⁵

Zur genaueren Bestimmung der typischen Eigenschaften der Artikel aus unserem Datensatz, welche in mindestens einem Policy-Dokument zitiert wurden, beschreiben wir diese Teilmenge ($n = 7838$) anschließend gesondert. Um quantitativ festzustellen, welche Art medizinischer Studien Policy-Dokumente vorwiegend zitieren, verwenden wir ein Machine-Learning-basiertes Verfahren. Dazu nutzen wir einen Datensatz annotierter Publikationsmetadaten von Major et al. (2018) nach.⁶ Auf Basis dieser händisch klassifizierten MEDLINE-indexierten Journalartikel trainieren wir ein Text-Klassifikationsmodell mittels Word2Vec, welches wir anschließend verwenden, um sämtliche Artikel unseres Datensatzes jeweils entweder der Translationsstufe 1, 2 oder 3 zuzuordnen. Je höher die einem Artikel zugewiesene Translationsstufe, desto höher die Wahrscheinlichkeit, dass es sich dabei um die Beschreibung einer klinischen Studie handelt; niedrige Translationsstufen deuten dagegen auf Artikel aus dem Bereich medizinischer Grundlagenforschung hin.

Training und Anwendung des Text-Klassifikationsmodells erfolgen in Python, alle übrigen statistischen Analysen werden in R durchgeführt (R Core Team, 2020).

4 Ergebnisse und Diskussion

4.1 Abdeckung und Ausprägung deutscher medizinischer Forschung in Altmetrics-Quellen

Abb. 2 zeigt die prozentualen Anteile der Publikationen des Hauptdatensatzes und der drei beispielhaft untersuchten Pilotfakultäten in den fünf hier untersuchten altmetrischen Datenquellen. Abb. 3 zeigt auf analoge Weise die durchschnittliche Anzahl von Nennungen, die Artikel aus den vier Gruppen auf den jeweiligen Plattformen erhalten haben. Sowohl hinsichtlich Abdeckung als auch Ausprägung zeigen sich erhebliche Variationen zwischen den altmetrischen Indikatoren. Die Social-Media-Plattform Twitter, auf der etwa 45 % der Artikel der hier betrachteten Gruppen mindestens eine und im Durchschnitt je Artikel etwa 4

⁵ <https://www.vosviewer.com/>

⁶ Siehe auch https://github.com/vincentmajor/ctsa_prediction/tree/master/data/labeled.

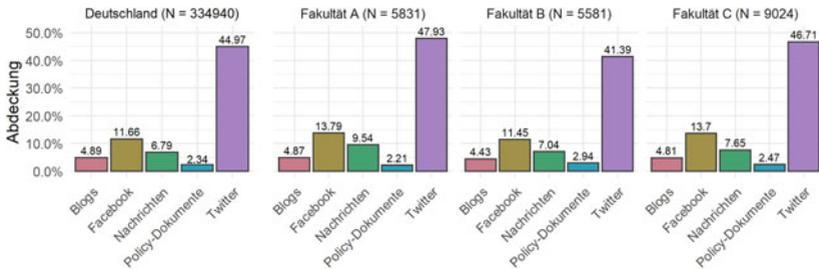


Abb. 2 Abdeckung der Publikationen des Hauptdatensatzes sowie der drei Pilotfakultäten in den fünf betrachteten altmetrischen Datenquellen.

Nennungen erhalten, übertrifft hinsichtlich dieser beiden Dimensionen alle anderen betrachteten Indikatoren deutlich. Die prozentualen Artikel-Abdeckungsraten auf den anderen vier Plattformen liegen ausnahmslos im einstelligen oder niedrigen zweistelligen Bereich, mit durchschnittlichen Anzahlen von Nennungen je Artikel durchweg deutlich unter 1.

Zwischen den Publikationen der Fakultäten werden ebenfalls Unterschiede sichtbar, wenn auch subtilerer Art: während Publikationen der Fakultät A auf Twitter (47,93 % Abdeckung; 4,81 Tweets pro Publikation) und in Nachrichten (9,54 % Abdeckung; 0,46 Nachrichtenerwähnungen pro Publikation) eine etwas weitere Wahrnehmung als die Publikationen der anderen Fakultäten erreichen, weisen die Publikationen von Fakultät B die im Vergleich höchste Abdeckung in Policy-Dokumenten auf (2,94 %).

Insgesamt liegen die in unserem Datensatz festgestellten Abdeckungen und Ausprägungen etwas höher als die, die Haustein et al. (2015) für die Gesamtheit 2012 erschienener biomedizinischer und lebenswissenschaftlicher Publikationen in Web of Science ermittelt haben. Mögliche Gründe für diese Unterschiede sind die in unserer Studie längeren Zitationszeiträume, eine eventuell gestiegene allgemeine Nutzung der Social-Media-Dienste, die dreien der hier betrachteten Metriken zugrunde liegen (Blogs, Facebook, Twitter) sowie die Möglichkeit, dass sich die bei Erzeugung unseres Datensatzes angewandten Filter (beispielsweise die Beschränkung auf Publikationen mit mindestens einer deutschen Autorin oder einem deutschen Autor) auf die durchschnittliche Qualität der enthaltenen Publikationen ausgewirkt haben (Abb. 3).

Um eine Orientierung über die thematische Zusammensetzung unseres Datensatzes zu bieten, zeigt Abb. 4 ein Ko-Okkurrenz-Netzwerk, basierend auf den um Stoppwörter bereinigten Termen aus Titeln und Abstracts von 100.000

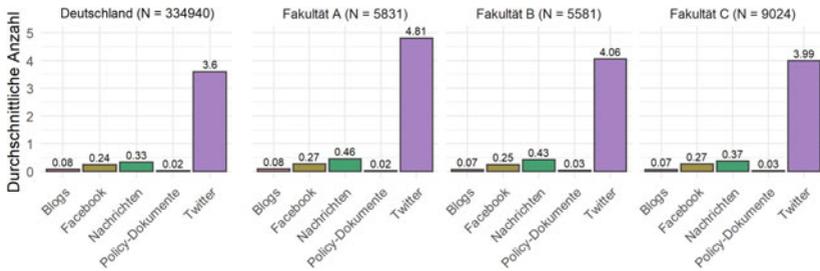


Abb. 3 Ausprägung von Nennungen der Publikationen des Hauptdatensatzes sowie der drei Profilkategorien in den fünf betrachteten altmetrischen Datenquellen.

zufällig ausgewählten Artikeln des Datensatzes (einzelne Knoten entsprechen individuellen Termen, die Knotengröße spiegelt die Häufigkeit des Terms im Datensatz wider, die Nähe der Knoten zueinander entspricht gemeinsamer Auftrittshäufigkeit). Einfärbungen basieren auf Term-Ko-Okkurrenzen und geben damit Hinweise auf grobe inhaltliche Cluster – so enthalten beispielsweise grüne Knoten vorwiegend Terme aus dem Bereich der Zellbiologie (z. B. *cell*, *protein*, *kinase*), während der rot eingefärbte Teil des Netzwerks Begriffe aus der Biochemie enthält (z. B. *compound*, *acid*, *oxidation*). Die erkennbare Unterteilung des Netzwerks in zwei große Partitionen scheint weitgehend der Unterscheidung zwischen typischerweise in medizinischer Grundlagenforschung auftretenden Termen (linke Hälfte des Netzwerks; Beispiel-Terme siehe oben) und für klinische Studien typischeren Begriffen (rechte Hälfte des Netzwerks; z. B. *patient*, *surgery*, *symptom*, *mortality*, *outcome*) zu entsprechen.

Um zu bestimmen, ob unterschiedliche thematische Gebiete unseres Datensatzes hinsichtlich ihrer auf bestimmten altmetrischen Plattformen erhaltenen Aufmerksamkeit variieren, projizieren wir im nächsten Schritt die Ausprägungen einzelner altmetrischer Indikatoren auf das in Abb. 4 dargestellte Ko-Okkurrenz-Netzwerk. In Abb. 5 zeigen rote Bereiche an, dass Artikel mit entsprechenden Termen in Titel oder Abstract überdurchschnittlich viele Erwähnungen auf den entsprechenden Plattformen erhalten haben.

Aus Abb. 5 wird deutlich, dass die Indikatoren hinsichtlich der durch sie abgedeckten Forschung variieren. Besonders auffällig erscheint die Konzentration von Nennungen in Policy-Dokumenten auf die rechte Hälfte des Netzwerkes – welche, wie zuvor gesehen, vorwiegend Inhalte klinischer Studien zu repräsentieren scheint. Während sich in Abb. 5 auch bei den anderen altmetrischen Indikatoren ein leichtes Übergewicht auf der jeweils rechten Netzwerkhälfte andeutet, scheint

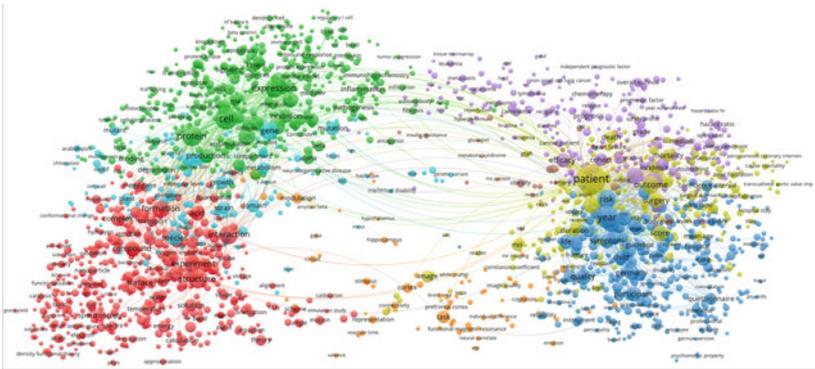


Abb. 4 Ko-Okkurrenz-Netzwerk der Terme aus 100.000 zufälligen Abstracts und Titeln des Datensatzes

die Verteilung der Aufmerksamkeit dort ausgewogener – insbesondere Nennungen in Blogs und in Nachrichten scheinen die Gesamtheit des Publikationskorpus relativ homogen abzudecken.

Als letzten Schritt der Term-Ko-Okkurrenz-Analyse nutzen wir die in Abb. 4 eingeführte Netzwerkansicht unseres Datensatzes, um eine grobe Charakterisierung der Publikationsprofile der drei in dieser Studie als Fallbeispiele betrachteten Pilotfakultäten vorzunehmen. Abb. 6 zeigt dafür Netzwerke, in denen rote Bereiche höhere relative Anteile von Publikationen mit Vorkommen der entsprechenden Terme (in Abstract oder Titel) unter dem Gesamtpublikationsaufkommen der jeweiligen Fakultät anzeigen.

Die in Abb. 6 zu sehenden Netzwerke offenbaren Unterschiede in den Publikationsprofilen der drei Fakultäten. Die zuvor in Abb. 4 gemachten Beobachtungen hinzuziehend deutet Abb. 6 darauf hin, dass Fakultät A das ausgewogenste Verhältnis aus Grundlagen- und klinischer Forschung publiziert, Fakultät C den deutlichsten Schwerpunkt auf klinische Forschung vorweist, während Fakultät B in dieser Hinsicht zwischen den anderen beiden Fakultäten steht. Die in Abb. 5 und 6 gezeigten Unterschiede verdeutlichen, wie relevant in biblio- oder altmetrischen Vergleichen auf Fakultätsebene die eingehende Betrachtung sowohl der thematischen Publikationsprofile der zu beurteilenden Institute als auch der themenabhängigen Verteilung in der Evaluation inkludierter Indikatoren ist. In dem vorliegenden Fall wäre beispielsweise anzunehmen, dass ein offenbar die klinische Forschung betonender Indikator (wie z. B. Policy-Dokument-Zitationen) besser dazu geeignet wäre, Nutzungsgrade bzw. Relevanz der bei Fakultät C

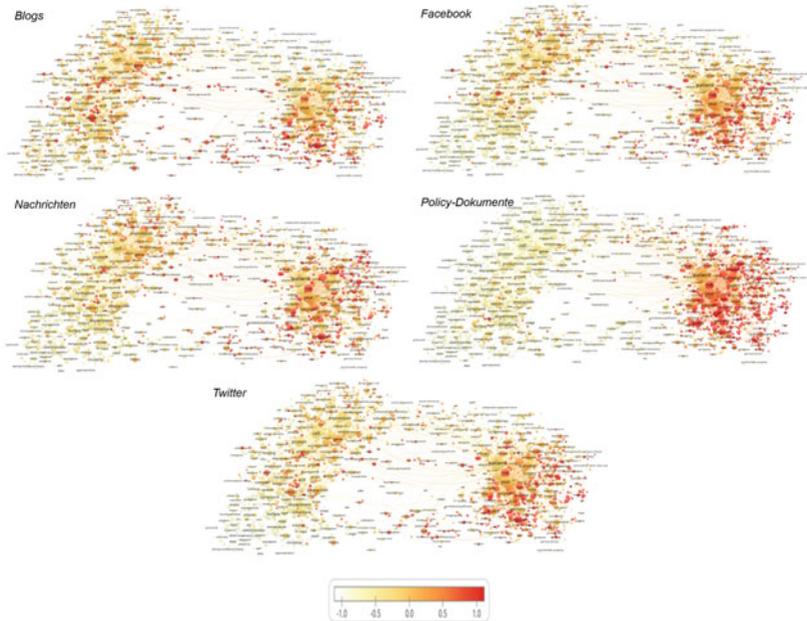


Abb. 5 Term-Ko-Okkurrenz-Netzwerk, eingefärbt entsprechend der Ausprägung altmetrischer Indikatoren bei jeweiligen Artikeln (o. l. Nennungen in Blogs, o. r. Nennungen auf Facebook, m. l. Nennungen in Nachrichten, m. r. Nennungen in Policy-Dokumenten, u. Nennungen auf Twitter). Die Farbvariationen drücken die relativen Häufigkeiten aus, mit denen jeweilige Terme (in Abstract und/oder Titel) enthaltene Publikationen in entsprechenden Umgebungen referenziert werden, je Indikator normalisiert auf eine Skala von -1 bis $+1$

stark vertretenen Publikationstypen darzustellen, als die Relevanz der Publikationen von Fakultät A – denn Fakultät A veröffentlicht mit größerem Anteil in medizinischen Bereichen, die vom Indikator Policy-Dokument-Zitationen kaum abgedeckt werden.

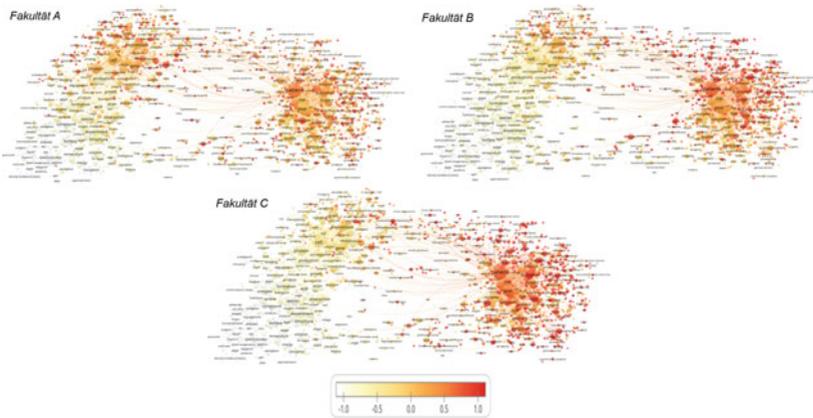


Abb. 6 Term-Ko-Okkurrenz-Netzwerk, eingefärbt entsprechend Ausprägung von Publikationen der drei Pilotfakultäten mit Vorkommen der jeweiligen Terme (o. l. Fakultät A, o. r. Fakultät B, u. Fakultät C). Die Farbvariationen drücken die relativen Häufigkeiten aus, mit denen jeweilige Terme in Abstracts und/oder Titeln der analysierten Publikationen der jeweiligen Fakultäten vorkommen, je Fakultät normalisiert auf eine Skala von -1 bis $+1$

4.2 Eigenschaften in Policy-Dokumenten zitierter Artikel

Ihre in Abb. 5 beobachtete starke Abdeckung klinischer Forschung ergänzt die in der Einleitung begonnene Argumentation, aufgrund derer Policy-Dokument-Zitationen für den Bereich biomedizinischer Forschung als besonders interessanter Relevanz-Indikator erscheinen: so gelangten vergangene Fallstudien bezüglich *akademischer* Zitationen (welche derzeit die maßgebliche Grundlage für quantitative Evaluationen medizinischer Forschung darstellen) wiederholt zu dem Schluss, dass diese die Grundlagenforschung überbetonen und zur Reflektion der Relevanz klinischer Studien daher wenig geeignet sind (Donner & Schmoch, 2020; Ke, 2020; van Eck et al., 2013). Aus diesem Grund würde der in Abb. 5 visuell festgestellte Fokus von Policy-Dokument-Zitationen auf klinische Forschung diese zu einem besonders interessanten Komplement zu bestehenden Indikatoren auf Basis akademischer Zitationen machen. Im nächsten Schritt überprüfen wir daher Existenz und Grad dieses Fokus zusätzlich mit einem quantitativen Verfahren.

Als ein solches quantitatives Verfahren nutzen wir annotierte Publikationsmetadaten von Major et al. (2018) zum Training eines Word2Vec-basierten

Text-Klassifikationsmodells, mit dem wir für die Publikationen unseres Datensatzes anschließend auf Basis ihrer Abstracts und Titel je eine Translationsstufe zwischen 1 (Artikel behandelt wahrscheinlich Grundlagenforschung) und 3 (Artikel behandelt wahrscheinlich klinische Forschung) vorhersagen. Anschließend vergleichen wir die Anteile der drei vorhergesagten Translationsstufen unter Publikationen unseres Gesamtdatensatzes ($n = 334.940$) und der Teilmenge von Publikationen mit mindestens einer gemessenen Policy-Dokument-Zitation ($n = 7838$). Geringen Anteilen der Artikel beider Gruppen (je 1,88 % der Gesamtpopulation und der Policy-zitierten Gruppe) wurde auf diese Weise keine Translationsstufe zugewiesen, da für diese Artikel keine oder aber nur sehr kurze Abstracts (weniger als 50 Wörter) ermittelt werden konnten.

Abb. 7 zeigt die relativen Anteile der drei Translationsstufen an den Publikationen beider Gruppen. Zwischen vorhergesagter Translationsstufe und Gruppenzugehörigkeit besteht ein statistisch signifikanter Zusammenhang ($\chi^2 = 6300,6$, $df = 3$, $p < ,001$). Der erheblich höhere Anteil als klinisch eingestufte Studien in der Gruppe Policy-Dokument-zitierter Artikel (44,49 %) gegenüber dem entsprechenden Anteil in der Gesamtpopulation (18,29 %) erhärtet den visuellen Befund aus Abb. 5, dass Policy-Dokument-Zitationen in besonderem Maße klinische Studien betreffen.

Zur genaueren Beschreibung der Eigenschaften in Policy-Dokumenten zitierter Artikel betrachten wir im nächsten Schritt die in dieser Gruppe häufigsten Journale sowie deren thematische Zuordnungen (anhand von Web of Science-Klassifikationen) und vergleichen deren jeweilige relative Anteile innerhalb der Policy-zitierten Gruppe ($n = 7838$) und der Gesamtpopulation ($n = 334.940$).

In Abb. 8 sind die in der Gruppe Policy-zitierter Artikel häufigsten zehn Journale und ihre jeweiligen relativen Anteile an beiden Gruppen dargestellt. Als unter Policy-zitierten Artikeln stark überrepräsentiert fallen besonders zwei renommierte lebenswissenschaftliche Journale auf, das *New England Journal of Medicine* (NEJM, stellt 2,14 % aller Policy-zitierten Artikel, gegenüber nur 0,11 % der Artikel der Gesamtpopulation) und *Lancet* (1,82 % aller Policy-zitierten, 0,09 % der Artikel aus der Gesamtpopulation). Angemerkt sei, dass die Berechnung der Spearman-Rangkorrelation zwischen dem Journal Impact Factor der die Artikel veröffentlichenden Journale und ihren Policy-Dokument-Zitationen zwar einen signifikant positiven, aber lediglich schwachen Zusammenhang ($r = 0,10$, $p < ,001$) zwischen beiden Maßen anzeigt. Obwohl sich an der Spitze der Policy-zitierten Journale solche befinden, die auch hinsichtlich ihres (zitationsbasierten) Journal Impact Factor herausstechen, ist also anzunehmen, dass die beiden Indikatoren durchaus erheblich voneinander abweichende Formen von Einfluss messen

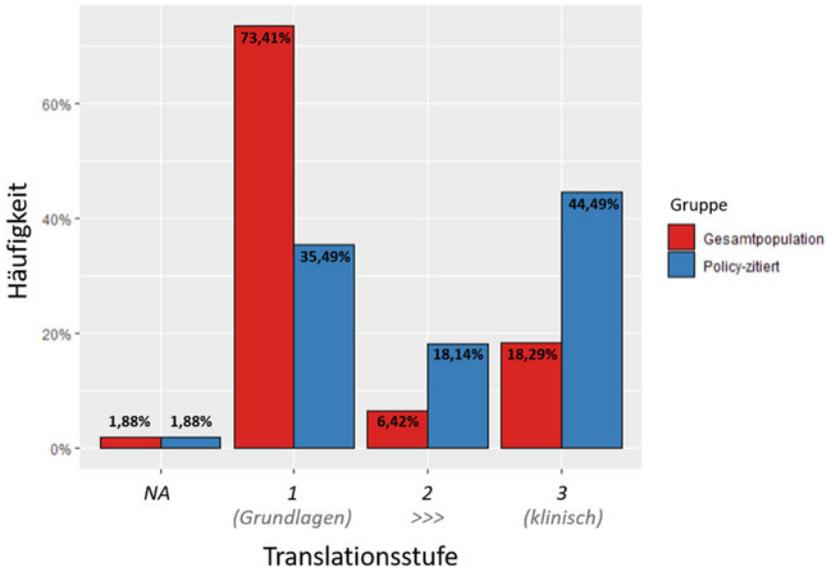


Abb. 7 Anteile per Text-Klassifikation bestimmter Translationsstufen an Artikeln der Gesamtpopulation (rot) und der Artikelgruppe mit je mindestens einer Policy-Dokument-Zitation (blau)

(s. auch Herrmann-Lingen in diesem Band), die sich möglicherweise nur hinsichtlich einiger ausgewählter Spitzenpositionen (wie in den Fällen von NEJM und Lancet) überschneiden.

Ein weiteres auffälliges Merkmal unter den in Policy-Dokumenten besonders häufig zitierten Journalen ist ein thematischer Schwerpunkt im Bereich der Onkologie, mit allein vier Journalen unter den ersten zehn, welche sich ihrem Titel nach explizit auf dieses Thema beziehen (*Journal of Clinical Oncology*, *Annals of Oncology*, *Lancet Oncology* und *European Journal of Cancer*).

Abb. 9 stellt die zehn bei Artikeln mit mindestens einer Policy-Dokument-Zitation häufigsten Web of Science-Klassifikationen dar (mehrere Klassifikationen je Journal sind möglich). Auch hier führen gewisse Themen mit erheblich höherer Repräsentation unter Policy-zitierten Artikeln als in der Gesamtpopulation das Feld an, allen voran *Public, Environmental & Occupational Health* (5,73 % in der Gruppe mit Policy-Dokument-Zitationen, gegenüber 1,34 % in der Gesamtpopulation) und *Oncology* (5,11 % in Policy-zitierten Artikeln, 2,77 % in der Gesamtpopulation). Diese Funde bestätigen die zuvor im Zusammenhang

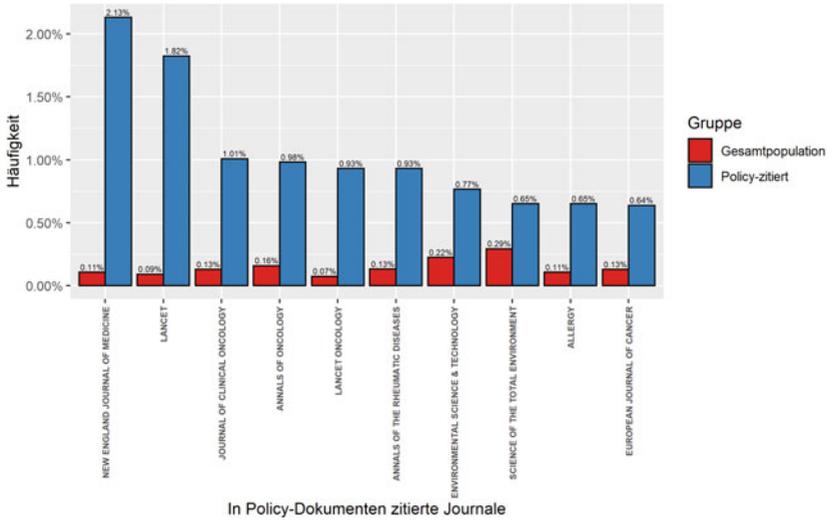


Abb. 8 Die zehn häufigsten Journale unter Artikeln mit mindestens einer Policy-Dokument-Zitation sowie ihre relativen Anteile in beiden untersuchten Gruppen

mit Journalen (s. Abb. 8) gemachten Beobachtungen, dass Policy-Dokument-Zitationen als Indikator deutlich zur Betonung bestimmter, klar umreißbarer Themenfelder zu neigen scheinen.

5 Konklusion

Dieser Beitrag hat die prinzipielle Eignung altmetrischer Indikatoren für das Anwendungsgebiet medizinischer Forschung in Deutschland anhand einer umfangreichen Stichprobe unter deutscher Beteiligung entstandener biomedizinischer Literatur untersucht. Analysen von Abdeckung und Ausprägung altmetrischer Indikatoren für Publikationen der Stichprobe machten dabei eine, allen Tendenzen steigender elektronischer Verbreitung wissenschaftlicher Veröffentlichungen zum Trotz anhaltende, Herausforderung für die praktische Nutzbarmachung altmetrischer Daten deutlich: in den meisten altmetrischen Quellen befinden sich die prozentualen Abdeckungsraten wissenschaftlicher Publikationen im einstelligen Bereich und die Ausprägungen unterhalb eines Wertes von 1 (mit Ausnahme von Erwähnungen auf Twitter). Diese hier für den spezifischen Fall

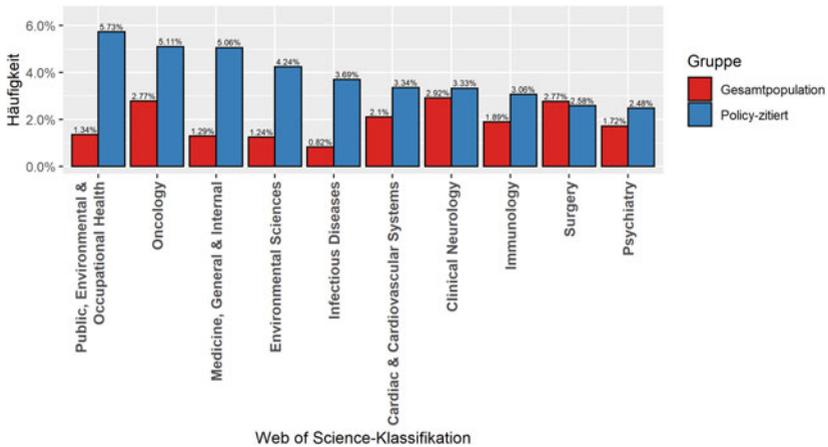


Abb. 9 Die zehn häufigsten Web of Science-Klassifikationen unter Artikeln mit mindestens einer Policy-Dokument-Zitation sowie ihre relativen Anteile in beiden untersuchten Gruppen

aktueller deutscher Medizinforschung gemachten Befunde stimmen mit einem erheblichen Bestand vergangener Fallstudien zu Altmetrics-Abdeckungsraten überein (Erdt et al., 2016; Work et al., 2015).

Diese Spärlichkeit altmetrischer Daten hat Implikationen für ihre Anwendung in evaluativen Kontexten. Praktische Anforderungen, die sich aus der Nutzung der extrem schief verteilten altmetrischen Daten für vergleichende Evaluationen ergeben, sind beispielsweise vorrangig robuste Schätzverfahren zu verwenden, sowie eine erhöhte Notwendigkeit Signifikanzen identifizierter Unterschiede zwischen untersuchten Gruppen mittels geeigneter statistischer Methoden nachzuweisen (Thelwall, 2017). Die szientometrische Forschung hat verschiedene konkrete feld- und zeitnormalisierte Indikatoren entwickelt, die diesen verteilungsbezogenen Eigenheiten altmetrischer Daten Rechnung tragen und trotz ihnen aussagekräftige Vergleiche zwischen Gruppen ermöglichen sollen (siehe beispielsweise Haunschild & Bornmann, 2018; Thelwall, 2017). Allgemein lässt sich jedoch festhalten, dass der hohe Anteil von Publikationen ohne messbaren altmetrischen Impact auch für den vorliegenden Anwendungsfall der deutschen Medizin zur Folge hat, dass Evaluationen unter Rückgriff auf Altmetrics vorrangig in (Makro-) Analysen, die sich auf große Publikationsmengen oder ausgedehnte Beobachtungszeiträume stützen können, zu informativen Ergebnissen führen werden und zu Bewertungen kleiner Publikationskorpora dagegen weniger geeignet sind.

Dieses Fazit zur grundsätzlichen Eignung von Altmetrics für Evaluationen im Bereich Medizin abschließend sei darauf hingewiesen, dass sich ein Großteil der in Bezug auf bibliometrische Forschungsevaluationen im Rahmen des Leidener Manifests (Hicks et al., 2015) formulierten Empfehlungen und Prinzipien auch auf altmetrische Evaluationen übertragen lässt und in entsprechenden Vorhaben berücksichtigt werden sollte (Bornmann & Haunschild, 2016; Lemke, 2022). Besonders betont sei in diesem Zusammenhang das erste Prinzip des Leidener Manifests: quantitative Evaluationen sollten qualitative Expertenurteile lediglich *unterstützen*. Auch wenn Altmetrics verschiedene Schwächen rein qualitativer Bewertungsverfahren wie auch bibliometrischer Indikatoren kompensieren können (Lemke et al., 2020; Weller, 2015), sollten sie – entgegen ihrer wörtlichen Bezeichnung – als *Komplement* zu diesen Formen der Wissenschaftsevaluation begriffen werden, nicht als diese ersetzende Alternativen.

Die ebenfalls in diesem Beitrag vorgenommene Untersuchung von Policy-Dokument-Zitationen und den durch sie betroffenen Artikeln der Stichprobe verstärkt den Eindruck, dass es sich hierbei um einen altmetrischen Indikator mit besonderem Potenzial (siehe auch Tahamtan & Bornmann, 2020), aber auch erheblichem verbleibendem Forschungsbedarf handelt. Unsere Analysen weisen auf besonders hohe Abdeckungen klinischer Forschung durch Policy-Dokument-Zitationen hin. Diese Eignung, relevante klinische Studien anzuzeigen und ihren praktischen Einfluss zu verdeutlichen, könnte in Evaluationsverfahren der Neigung etablierter bibliometrischer Indikatoren zur relativen Überrepräsentation von Grundlagenforschung entgegenwirken (Donner & Schmoch, 2020; Ke, 2020; van Eck et al., 2013). Die statistischen Deskriptionen in Policy-Dokumenten zitierter Artikel dieses Beitrags geben darüber hinaus weitere Hinweise darauf, welche Arten von Veröffentlichungen dieser Indikator seinerseits besonders betont – so deuten sich insbesondere Zusammenhänge zwischen Policy-Zitationen und bestimmten Themenfeldern wie Onkologie oder öffentlicher Gesundheit an. Die Belastbarkeit und Generalisierbarkeit dieser explorativen Befunde sollte im Rahmen tiefergehender Analysen unter Rückgriff auf komplexere statistische Verfahren weiter überprüft werden.

Weiterer Forschungsbedarf besteht auch hinsichtlich der besseren theoretischen Fundierung der Interpretation von Policy-Dokument-Zitationen. In dieser Studie wurden Policy-Dokument-Zitationen durch den Datenanbieter Altmetric.com ermittelt, welcher eine praktikable und leicht skalierbare Quelle für derartige Daten darstellt. Zwischen den von Altmetric.com aggregierten Anzahlen von Policy-Dokument-Zitationen fand jedoch keine weitere inhaltliche Differenzierung statt. Altmetric.com bezieht Policy-Dokument-Zitationen

zum Zeitpunkt des Schreibens von etwa 470 unterschiedlichen Quellen,⁷ darunter viele für den vorliegenden Anwendungsfall zweifellos relevante Plattformen, wie nationale und internationale Repositorien medizinischer Leitlinien. Gleichermaßen enthalten die von Altmetric.com vorgenommenen Zählungen aber auch in hohem Maße heterogene Signale wie beispielsweise Nennungen in Richtlinien von Regierungsbehörden, in Publikationen unabhängiger Think Tanks oder in Berichten internationaler Entwicklungsorganisationen.⁸ Die individuelle Relevanz dieser verschiedenen Datenquellen wird in vielen Fällen vom spezifischen Evaluationskontext abhängen. Für die Altmetrics-Forschung ergibt sich daraus ein dringender Bedarf, konkrete Konzepte und theoretische Frameworks zu einer differenzierteren Handhabung von Policy-Dokument-Zitationen zu entwickeln, deren Umsetzung auf technischer Ebene von einem jeweiligen Datenaggregator unterstützt werden muss. Den Bereich der Aggregatoren von Policy-Dokument-Zitationen hat zudem mit Overton in jüngerer Zeit noch ein weiterer Wettbewerber betreten, dessen Datenangebot in zukünftigen Studien vermehrt mit dem von Altmetric.com verglichen werden sollte (siehe beispielsweise Maleki & Holmberg, 2022).

Es ist die Entwicklung solcher Konzepte, von der die zukünftige Validität, Transparenz und damit auch Nützlichkeit von Policy-Dokument-Zitationen als Indikator in Evaluationen direkt abhängt. Der derzeitige Reifegrad der theoretischen Fundierung dieses speziellen altmetrischen Indikators wie auch zuvor genannte, aus geringen Abdeckungsraten resultierende Herausforderungen lassen ihren Nutzen zur Bewertung medizinischer Forschung zum jetzigen Zeitpunkt stark limitiert erscheinen. Doch langfristig ist das Policy-Dokument-Zitationen innewohnende Potenzial, evaluative Praktiken im Bereich der Biomedizin zu bereichern, groß – sowohl durch eine bessere Berücksichtigung von Leistungen im klinischen Bereich als auch durch eine allgemein vielfältigere Darstellung der Wirkungen akademischer Forschung auf politische Gestaltung und Gesellschaft.

Danksagung Wir danken Nicholas Fraser für seine Beiträge hinsichtlich Erhebung, Bereinigung und Anreicherung der Daten, deren Auswertung in diesem Beitrag beschrieben wurde. Darüber hinaus danken wir dem deutschen Kompetenznetzwerk Bibliometrie sowie Altmetric.com für die uns gewährten Datenzugänge. Weiterer Dank gilt dem deutschen Bundesministerium für Bildung und Forschung für die Förderung dieses Forschungsprojektes (Förderkennzeichen 16PU17011D).

⁷ <https://www.altmetric.com/blog/how-to-view-the-sources-altmetric-is-tracking/>

⁸ <https://help.altmetric.com/support/solutions/articles/6000236695-policy-documents>.

Literatur

- Bornmann, L., & Haunschild, R. (2016). To what extent does the Leiden manifesto also apply to altmetrics? A discussion of the manifesto against the background of research into altmetrics. *Online Information Review*, 40(4), 529–543. <https://doi.org/10.1108/OIR-09-2015-0314>.
- Costas, R., Zahedi, Z., & Wouters, P. (2015). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019. <https://doi.org/10.1002/asi.23309>.
- Donner, P., & Schmoch, U. (2020). The implicit preference of bibliometrics for basic research. *Scientometrics*, 124(2), 1411–1419. <https://doi.org/10.1007/s11192-020-03516-3>.
- Erdt, M., Nagarajan, A., Sin, S.-C.J., & Theng, Y.-L. (2016). Altmetrics: An analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics*, 109(2), 1117–1166. <https://doi.org/10.1007/s11192-016-2077-0>.
- Fraser, N., Bräuer, P., & Peters, I. (2021). Altmetrics for evaluation of medical research in Germany. *Proceedings of the 18th Conference of the International Society for Scientometrics and Informetrics*, 1471–1472. <http://hdl.handle.net/11108/504>.
- Glänzel, W., & Gorraiz, J. (2015). Usage metrics versus altmetrics: Confusing terminology? *Scientometrics*, 102(3), 2161–2164. <https://doi.org/10.1007/s11192-014-1472-7>.
- Haunschild, R., & Bornmann, L. (2018). Field- and time-normalization of data with many zeros: An empirical analysis using citation and Twitter data. *Scientometrics*, 116(2), 997–1012. <https://doi.org/10.1007/s11192-018-2771-1>.
- Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics*, 108(1), 413–423. <https://doi.org/10.1007/s11192-016-1910-9>.
- Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLoS ONE*, 10(3), e0120495. <https://doi.org/10.1371/journal.pone.0120495>.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, 520(7548), 429. <https://doi.org/10.1038/520429a>.
- Ke, Q. (2020). The citation disadvantage of clinical research. *Journal of Informetrics*, 14(1), 100998. <https://doi.org/10.1016/j.joi.2019.100998>.
- Lemke, S. (2022). An Assessment of Impact Metrics' Potential as Research Indicators Based on Their Perception, Usage, and Dependencies from External Science Communication [Doctoral thesis, CAU Kiel University]. https://macau.uni-kiel.de/receive/macau_mods_00003250.
- Lemke, S., Witthake, A., & Peters, I. (2022). Altmetrics for German medical research: What leads to research articles achieving policy impact? *Proceedings of the 26th International Conference on Science, Technology and Innovation Indicators*. STI 2022, Granada, Spain. <https://doi.org/10.5281/zenodo.6645109>.
- Lemke, S., Zagovora, O., Weller, K., Orth, A., Beucke, D., Stropel, J., & Peters, I. (2020). *metrics—Recommendations from the DFG *metrics project for „Measuring the Reliability and Perceptions of Indicators for Interactions with Scientific Products“. <https://doi.org/10.18452/22242.2>.

- Major, V., Surkis, A., & Aphinyanaphongs, Y. (2018). Utility of General and Specific Word Embeddings for Classifying Translational Stages of Research. *AMIA Annual Symposium Proceedings, 2018*, 1405–1414.
- Maleki, A., & Holmberg, K. (2022). Comparing coverage of policy citations to scientific publications in Overton and Altmetric.com: Case study of Finnish research organizations in Social Science. *Informaatiotutkimus, 41*(2–3), Art. 2–3. <https://doi.org/10.23978/inf.122592>.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010, Oktober 26). *Altmetrics: A manifesto* [Altmetrics]. <http://altmetrics.org/manifesto/>.
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology, 68*(9), 2037–2062. <https://doi.org/10.1002/asi.23833>.
- Szomszor, M., & Adie, E. (2022). Overton—A bibliometric database of policy document citations. *arXiv:2201.07643 [cs]*. <http://arxiv.org/abs/2201.07643>.
- Tahamtan, I., & Bornmann, L. (2020). Altmetrics and societal impact measurements: Match or mismatch? A literature review. *El profesional de la información, 29*(1). <https://doi.org/10.3145/epi.2020.ene.02>.
- Thelwall, M. (2017). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics, 11*(1), 128–151. <https://doi.org/10.1016/j.joi.2016.12.002>.
- van Eck, N. J., Waltman, L., van Raan, A. F. J., Klautz, R. J. M., & Peul, W. C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLoS ONE, 8*(4), e62395. <https://doi.org/10.1371/journal.pone.0062395>.
- Weller, K. (2015). Social media and altmetrics: An overview of current alternative approaches to measuring scholarly impact. In I. M. Welpé, J. Wollersheim, S. Ringelhan, & M. Osterloh (Hrsg.), *Incentives and performance: Governance of research organizations* (S. 261–276). Springer International Publishing. https://doi.org/10.1007/978-3-319-09785-5_16.
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Keridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. <https://doi.org/10.13140/RG.2.1.4929.1363>.
- Work, S., Haustein, S., Bowman, T. D., & Larivière, V. (2015). Social media in scholarly communication. A review of the literature and empirical analysis of Twitter use by SSHRC Doctoral Award Recipients | Canada Research Chair on the Transformations of Scholarly Communication. <http://crc.ebsi.umontreal.ca/en/publications/social-media-in-scholarly-communication-sshr/>.
- Wouters, P., & Costas, R. (2012). *Users, Narcissism and Control: Tracking the Impact of Scholarly Publications in the 21st Century*. Stichting Surf. <http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf>.
- Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PLoS ONE, 13*(5), e0197326. <https://doi.org/10.1371/journal.pone.0197326>.

Lemke, Steffen Steffen Lemke ist wissenschaftlicher Mitarbeiter und Postdoc in der Forschungsgruppe Web Science unter Leitung von Prof. Dr. Isabella Peters am Institut für Informatik der Christian-Albrechts-Universität zu Kiel. Steffen forscht in verschiedenen Projekten zu den Themenfeldern Szientometrie, Wissenschaftskommunikation und Altmetrics.

Witthake, Anne Anne Witthake arbeitet in der Abteilung Open Science Transfer der ZBW im Bereich Senior Application Development und war für das Projekt QuaMedFo teamübergreifend als Datenanalytikerin tätig.

Peters, Isabella Isabella Peters ist Professorin für Web Science an ZBW – Leibniz-Informationszentrum Wirtschaft und der Christian-Albrechts-Universität zu Kiel. Sie erforscht den Einsatz von Social Media in der wissenschaftlichen Kommunikation, z.B. Altmetrics, und in Open Science.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Welche Charakteristika weist die in Leitlinien zitierte Literatur auf? – Eine Fallstudie anhand einer S2e- und einer S3-Leitlinie

Christopher Traylor und Valeria Aman

1 Einleitung

Medizinische Leitlinien sind ein Instrument zur Information des medizinischen Personals und zur Überbrückung der Kluft zwischen wissenschaftlichen Erkenntnissen und der täglichen Praxis (Woolf et al., 1999). Leitlinien geben Empfehlungen dazu, wie eine bestimmte Krankheit erkannt und behandelt werden sollte und richten sich primär an Ärzt*innen, Pflegekräfte und Fachpersonal im Gesundheitswesen. Erarbeitete Leitlinien sollen die Patientenversorgung verbessern, indem sie Nutzen und Schaden von Untersuchungen und Behandlungen bewerten und auf dieser Basis konkrete Empfehlungen zu Diagnose, Therapie und Nachsorge geben. Leitlinien fassen dabei das aktuelle medizinische Wissen zusammen, belegen dies mit Referenzen auf publizierte Studien und müssen daher regelmäßig aktualisiert werden (Thelwall & Maflahi, 2016).

Medizinische Forschung, die in die Entwicklung von Leitlinien mündet, ist von Relevanz, da sie die klinische Praxis beeinflusst und zur besseren Gesundheitsversorgung beiträgt (Zhang & Wu, 2021). Damit bildet die Zitation von

C. Traylor (✉)

Klinik für Psychosomatische Medizin und Psychotherapie, Universitätsmedizin
Göttingen – Georg-August-Universität, Göttingen, Deutschland

E-Mail: christopher.traylor@med.uni-goettingen.de

V. Aman

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Berlin,
Deutschland

E-Mail: aman@dzhw.eu

© Der/die Autor(en) 2024

S. Biesenbender und J. Hartstein (Hrsg.), *Qualitätsmessung als Prisma*, Higher
Education Research and Science Studies,

https://doi.org/10.1007/978-3-658-43683-4_5

wissenschaftlicher Literatur in medizinischen Leitlinien deren gesellschaftlichen Impact ab (Kryl et al., 2012).

Traditionell wird der Einfluss wissenschaftlicher Arbeiten mit dem Journal Impact Factor (JIF) gemessen (Garfield, 2006). Der Journal Impact Factor ist jedoch in erster Linie ein Indikator für das Prestige einer Zeitschrift (Thelwall & Kousha, 2016) und eignet sich weniger für die Bewertung der in Leitlinien zitierten Literatur (Traylor & Herrmann-Lingen, 2023). In einer früheren Studie wurde jedoch gezeigt, dass qualitativ hochwertigere Leitlinien auf Artikel verweisen, die hohe Zitationszahlen aufweisen (Andersen, 2013).

Angelehnt an die Arbeit von Thelwall und Maflahi (2016) sind wir zum einen daran interessiert, die Artikel, die in Leitlinien zitiert werden, mit ähnlichen Artikeln zu vergleichen, die nicht in derselben Leitlinie zitiert werden, um herauszufinden, ob medizinische Leitlinien aus Deutschland in ähnlicher Weise hochzitierte Artikel referenzieren. Anstatt jedoch wie Thelwall und Maflahi die Vergleichsmenge auf denselben Jahrgang und dieselbe Ausgabe einzugrenzen, vergleichen wir einen Artikel, der in einer Leitlinie zitiert wird, mit allen anderen Artikeln, die in derselben Zeitschrift und demselben Jahrgang erschienen sind. Durch die verbesserte Methodik senken wir den Einfluss von potenziellen Unterschieden, die zwischen Zeitschriftenbänden und -ausgaben bestehen können und vergrößern den Umfang des zu vergleichenden Datensatzes, was wiederum die Aussagekraft unserer Studie erhöht. Außerdem untersuchen wir die Zahl der Referenzen, die in einer Leitlinie zitiert werden und die Publikationsjahre der in den Leitlinien zitierten Artikel. Dabei stellen wir das Publikationsjahr der zitierten Referenzen den prozentualen Zitierrängen der Leitlinien-Referenzen gegenüber. Diese Untersuchung kann Aufschluss darüber geben, ob Artikel in Leitlinien referenziert werden, da sie aufgrund ihres Alters eine höhere Sichtbarkeit und Verbreitung haben als jüngere Artikel, die weniger Zeit hatten Zitationen zu akkumulieren.

Zum anderen sind wir daran interessiert herauszufinden, inwiefern frühere Arbeiten von Leitlinien-Autor*innen die Wahl von Artikeln, die in einer Leitlinie referenziert werden, beeinflussen. Dafür wurden alle Personen, die an der Erstellung einer Leitlinie beteiligt waren, in der Literaturdatenbank Scopus auffindig gemacht und die von ihnen in früheren Arbeiten zitierten Artikel mit denen in der entsprechenden Leitlinie zitierten verglichen.

Um die Charakteristika von Leitlinien-Referenzen zu untersuchen, haben wir eine S2e- und eine S3-Leitlinie ausgewählt. Diese zwei Leitlinien-Arten unterscheiden sich u. a. hinsichtlich der Zusammensetzung der Leitliniengruppe, ihrer Erarbeitung, ihrer Konsensfindung und Qualität. Somit vermuten wir, dass sich

bei den zwei untersuchten Leitlinien unterschiedliche Muster in den von uns untersuchten Charakteristika zeigen.

2 Methoden

Die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) besteht aus 180 wissenschaftlichen medizinischen Fachgesellschaften und koordiniert die Entwicklung von qualitätsgesicherten medizinischen Leitlinien für Diagnostik und Therapie. Zwei medizinische Leitlinien, die in dem Portal der AWMF verfügbar sind, wurden für die Analyse ausgewählt. Die S2e-Leitlinie *Neue Thrombozytenaggregationshemmer: Verwendung in der Hausarztpraxis* (053–014) ist der Allgemeinmedizin zuzuordnen und wurde federführend von der Deutschen Gesellschaft für Allgemeinmedizin und Familienmedizin e. V. (DEGAM) herausgegeben. Die S3-Leitlinie, *Diagnose, Therapie und Nachsorge des Nierenzellkarzinoms* (043–017) ist der Onkologie zuzuordnen und wurde federführend von der Deutschen Gesellschaft für Urologie (DGU) und der Deutschen Gesellschaft für Hämatologie und Onkologie (DGHO) verfasst. Beide Leitlinien wurden im Jahr 2017 herausgegeben.

Abgesehen von der unterschiedlichen Entwicklungsstufe, unterscheiden sich die untersuchten Leitlinien primär hinsichtlich ihrer fachlichen Ausrichtung: Allgemeinmedizin vs. Onkologie. Die S2e-Leitlinie von der DEGAM richtet sich speziell an Hausärzte und soll für Hausärzte richtungweisend sein, um eine Über-, Unter- oder Fehlversorgung mit Thrombozytenaggregationshemmern zu vermeiden. Die DEGAM-Leitlinie befasst sich mit bestimmten Medikamenten, die bei weit verbreiteten Krankheiten wie Vorhofflimmern eingesetzt werden. Das Besondere an dieser Leitlinie ist, dass darin eine relativ neue Substanzgruppe thematisiert wird, zu der es noch wenige Erfahrungen und somit wenige publizierte Studien gibt, auf die in der Leitlinie verwiesen werden kann. Die S3-Leitlinie hat das Nierenzellkarzinom zum Gegenstand und behandelt ein zwar selteneres, jedoch bedrohliches Krankheitsbild, das seit längerem in der Medizin bekannt ist und zu dem seit mehreren Jahren intensiv geforscht wird. Anders als die S2e-Leitlinie thematisiert die S3-Leitlinie nicht nur die Therapie, sondern auch die Diagnostik und Nachsorge.

Die Referenzen dieser zwei Leitlinien wurden zunächst aus den PDF-Dokumenten extrahiert und maschinenlesbar gemacht. Die Recherche der anderen Artikel, die mit einem in der Leitlinie referenzierten Artikel zu vergleichen sind, erfolgte durch die Suche im Web of Science (WoS). Die Suchergebnisse wurden

nach dem Erscheinungsjahr der Referenz, dem Zeitschriftentitel und dem Dokumenttyp „Artikel“ gefiltert, wobei sich unter Artikeln auch publizierte Reviews finden. Die Suchergebnisse wurden heruntergeladen und für die Auswertung in einer eigens konzipierten Datenbank zusammengeführt. Abb. 1 verdeutlicht den Erhebungsprozess der Artikel in WoS. Jede Leitlinien-Referenz wurde in WoS recherchiert, wie z. B. eine Referenz auf die Zeitschrift Nature im Jahr 2005. Dann wurden alle Artikel, die in Nature im Jahr 2005 publiziert wurden, abgefragt, samt ihrer Zitationsdaten bis zum Oktober 2022.

In der ersten Analyse wurde zu jeder Referenz einer Leitlinie ihr Zitierang unter allen Artikeln aus der Zeitschrift der Referenz und dem Publikationsjahr der Referenz bestimmt. Zur Vergleichbarkeit verschiedener Referenzen wurde daraus der prozentuale Rang berechnet.

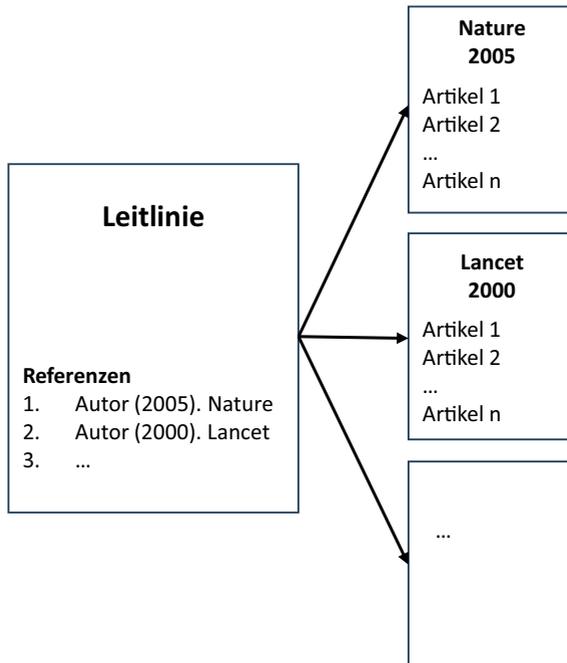


Abb. 1 Der Datensatz besteht aus Zeitschriftenartikeln, die in denselben Jahrgängen derselben Zeitschriften veröffentlicht wurden wie die Artikel, auf die in der entsprechenden Leitlinie verwiesen wird

Für die anschließende Analyse der Artikel, die in Leitlinien zitiert werden und womöglich bereits in früheren Arbeiten der an einer Leitlinie beteiligten Personen zitiert wurden, wurden die Leitlinien-Autor*innen in der Publikations- und Zitationsdatenbank Scopus recherchiert, um ihre weiteren Publikationen aufzufinden.¹ Die so erhobenen Publikationen wurden nach dem Kriterium analysiert, ob sie die in den Leitlinien zitierte Literatur bereits früher referenziert haben. Die Referenzen der Leitlinien wurden dazu sowohl über die DOI als auch die PubMed-ID (PMID) ebenfalls in Scopus zugeordnet.

3 Ergebnisse

Die S3-Leitlinie (043–017) weist insgesamt 702 Referenzen auf, während die S2e-Leitlinie (053–041) insgesamt 147 Referenzen hat. Wir konnten 514 von 702 (73 %) Referenzen der 043–017-Leitlinie in WoS ausfindig machen und 97 von 147 (66 %) Referenzen der 053–041-Leitlinie. Bei den nicht zugeordneten Referenzen handelt es sich um Verweise auf andere Leitlinien oder Bücher sowie um Zeitschriften, die nicht in WoS indexiert sind. In der S3-Leitlinie (043–017) wurden Referenzen aus 109 verschiedenen Fachzeitschriften aus den Jahren 1964–2017 verwendet. Die S2e-Leitlinie (053–041) zitiert Artikel aus 27 verschiedenen Zeitschriften, die in den Jahren 1983 bis 2015 veröffentlicht wurden. In Abb. 2 sind die kumulativen Verteilungsfunktionen der prozentualen Zitierränge der Leitlinien-Referenzen dargestellt. Der durchschnittliche prozentuale Zitierrang beträgt für die in der S2e-Leitlinie zitierten Artikel 80,4 % ($\pm 4,8$ %, $p = 0,05$) und für die in der S3-Leitlinie zitierten Artikel 74,4 % ($\pm 2,1$ %, $p = 0,05$). Somit weisen Artikel, die in den untersuchten Leitlinien zitiert werden, signifikant höhere Zitationszahlen auf als andere Publikationen gleicher Zeitschriften und Jahrgänge.

Im Folgenden wurde geschaut aus welchen Jahren die zitierten Referenzen stammen und wie häufig sie im Vergleich zu Artikeln aus dem gleichen Jahr und der gleichen Zeitschrift zitiert wurden. Abb. 3 zeigt, dass eine Häufung in der Nähe des Veröffentlichungsdatums der Leitlinie zu beobachten ist. Für beide Leitlinien lässt sich sagen, dass es nur wenige Artikel gibt, die vergleichsweise wenige Zitierungen erhalten haben.

¹ Die Scopus Author ID ermöglicht es den gesamten Publikationsoutput einer Person zu einer Autoren-ID in Scopus zuzuordnen. In Web of Science gibt es hingegen keine automatisierte Zuordnung von Artikeln zu Autoren-IDs.

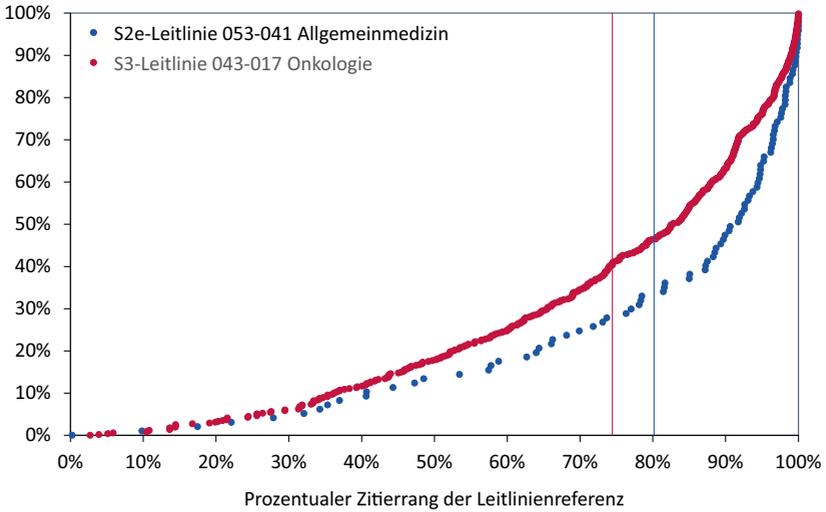


Abb. 2 Vergleich der kumulativen Verteilungsfunktionen der prozentualen Zitieränge der Referenzen der zwei Leitlinien. Der Zitierang wurde bestimmt unter allen Artikeln derselben Zeitschrift und desselben Jahrgangs. Die senkrechten Linien stellen die durchschnittlichen Zitieränge dar

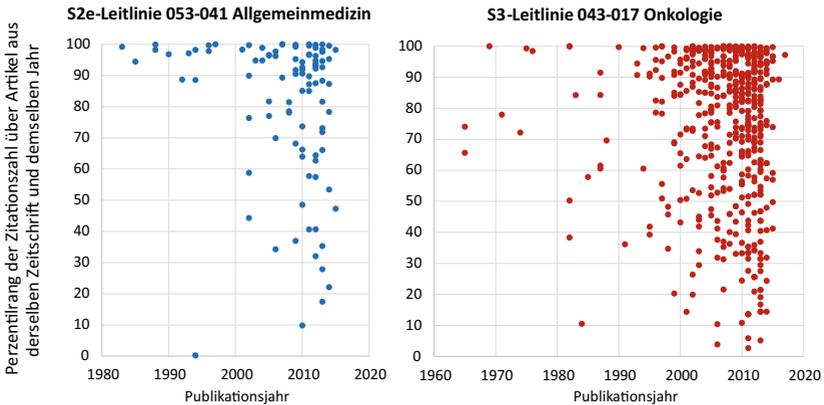


Abb. 3 Vergleich der zwei Leitlinien hinsichtlich der prozentualen Zitieränge ihrer Referenzen und der Publikationsjahre der Referenzen

Während neuere Artikel sowohl in den unteren als auch in den oberen Rängen zu finden sind, weisen ältere Artikel überwiegend höhere Ränge auf, was darauf hindeutet, dass sie trotz ihres lange zurückliegenden Erscheinungsjahres auf die aktuelle medizinische Forschung Einfluss nehmen. Zugleich könnte es ein Indiz dafür sein, dass diese Artikel Eingang in Leitlinien finden, gerade weil sie hochzitiert sind und dadurch eine höhere Sichtbarkeit und Auffindbarkeit haben. Die Zeitspanne der zitierten Referenzen der S2e-Leitlinie ist mit 32 Jahren kürzer als die der S3-Leitlinie mit 52 Jahren. Das durchschnittliche Publikationsjahr der referenzierten Literatur in der S2e-Leitlinie ist 2007,3 (Median: 2010), während die Literatur in der S3-Leitlinie etwas älter ist mit 2006,9 (Median: 2009).

Eine weitere Analyse galt den an den Leitlinien beteiligten Personen und insbesondere der von ihnen in früheren Arbeiten zitierten Literatur. Die S2e-Leitlinie wurde von zwei Autoren erstellt, die der Fachgesellschaft DEGAM angehören und beide in Bremen affiliert sind. Die S3-Leitlinie weist hingegen 63 Autor*innen auf, denen verschiedene Rollen bei der Leitlinienerstellung zuteilkommen. So gibt es zwei Personen, die die Koordination und die Redaktion innehatten, drei Personen, die für die Qualitätsindikatoren zuständig waren, sieben Expert*innen und schließlich 51 Autor*innen und Mandatsträger*innen. Die an der Leitlinie beteiligten Personen gehören verschiedenen Fachgesellschaften wie der DGU (Deutsche Gesellschaft für Urologie) oder DEGRO (Deutsche Gesellschaft für Radioonkologie) an und sind nicht nur zu Einrichtungen in Deutschland, sondern auch im Ausland (u. a. Schweiz oder Niederlande) affiliert. Im Hinblick auf Leitlinien-Referenzen, die im Fokus dieser Studie stehen, wurde untersucht, inwiefern die zitierte Literatur in den Leitlinien von den an der Leitlinie beteiligten Personen in früheren Artikeln zitiert wurden. Da beide Leitlinien im Jahr 2017 veröffentlicht wurden, wurden nur solche Artikel analysiert, die bis einschließlich 2016 erschienen sind. Für die S2e-Leitlinie konnten 108 der insgesamt 147 Referenzen in der Scopus-Datenbank nachgewiesen werden. Von diesen 108 Referenzen wurden insgesamt 28 Referenzen vor Erscheinen der Leitlinie von mindestens einem der beiden Leitlinien-Autoren zitiert (26 %). Einer der S2e-Leitlinien-Autoren hat alle dieser 28 Referenzen bereits in Artikeln vor dem Jahr 2017 genutzt, während der Co-Autor der Leitlinie nur einen Artikel in einer früheren Studie zitiert hat. Für die S3-Leitlinie konnten 604 der insgesamt 702 Artikel in Scopus anhand ihrer DOI oder PubMed-ID auffindig gemacht werden. Insgesamt wurden 421 dieser 604 Referenzen (70 %) in früheren Arbeiten von den Leitlinien-Autor*innen zitiert. Von den 63 an der Leitlinie beteiligten Personen, haben 43 mindestens einen in der Leitlinie zitierten Artikel in früheren Arbeiten zitiert. Die Spannweite reicht von einer Person mit einem einzigen Artikel hin zu einer Person, die 188 der in der Leitlinie zitierten Referenzen bereits in

Artikeln vor 2017 zitiert hat. (Der Mittelwert beträgt 46 Referenzen pro Person). Diese Zahlen können ein Indiz dafür sein, dass die zitierten Artikel nicht nur aufgrund ihrer Qualität und ihrer Sichtbarkeit Eingang in die Leitlinien finden, sondern aus dem Grund, dass sie den beteiligten Personen aus früheren Studien bekannt sind, in denen zu ähnlichen Themen geforscht wurde und zwangsläufig dieselbe Literatur zitiert wurde. Zugleich zeigen diese Ergebnisse, dass die Arbeit an Leitlinien eine systematische Literaturrecherche erfordert, bei der möglichst viele existierende Studien ausgewertet werden. Dabei finden auch Publikationen Eingang, die in früheren Arbeiten der Leitlinien-Autor*innen unentdeckt blieben oder als nicht zitierenswert erachtet wurden.

4 Diskussion

Um möglichst verschiedene Muster bei der in Leitlinien referenzierten Literatur aufzeigen zu können, haben wir in dieser Fallstudie eine S2e- und eine S3-Leitlinie analysiert. S2e- und S3-Leitlinien unterscheiden sich hinsichtlich ihrer Entstehung und Qualität. Wie S3-Leitlinien erfordern auch S2e-Leitlinien eine systematische Recherche, Auswahl und Bewertung von wissenschaftlichen Studien, jedoch gibt es bei unterschiedlichen Auffassungen keine strukturierte Konsensfindung. S3-Leitlinien hingegen erfordern eine Literaturrecherche, die nach a priori festgelegten Kriterien erfolgt und die ausgewählte Evidenz wird hinsichtlich ihrer methodischen Qualität bewertet. Die Kommission ist im Hinblick auf die beteiligten Berufsgruppen und der Zielpopulation (z. B. Patient*innen/Bevölkerung) repräsentativ besetzt und in einer strukturierten Konsensfindung werden Empfehlungsgrade festgelegt. S3-Leitlinien sind am verlässlichsten, aber auch aufwendig in ihrer Erarbeitung. Daher war es interessant zu untersuchen, ob sich die verschiedenen Leitlinien-Arten hinsichtlich der Zahl der Referenzen, der Zitationszahl der Referenzen, des Alters der Referenzen und der Bekanntheit der Referenzen unter den Leitlinien-Autor*innen unterscheiden.

Im Hinblick auf die Referenzmenge gibt es weitere Unterschiede: so weist die S2e-Leitlinie 147 Referenzen auf, wohingegen die S3-Leitlinie 702 Referenzen hat. Gründe für die deutlich längere Referenzliste der S3-Leitlinie wurden oben genannt und umfassen, u. a. das länger bekannte Krankheitsbild, mehr vorhandene Literatur zu dem Thema und die breitere Abhandlung der Krankheit in der Leitlinie. Die Auswertung hat gezeigt, dass die Zeitspanne der zitierten Referenzen in der S2e-Leitlinie mit 32 Jahren kürzer ist als in der S3-Leitlinie mit

52 Jahren. Die Unterschiede hinsichtlich des durchschnittlichen Alters der referenzierten Literatur sind jedoch mit etwas mehr (S3-Leitlinie) bzw. etwas weniger (S2e-Leitlinie) als 10 Jahren (Median) marginal.

Eine weitere Ursache für die vielen Referenzen und deren breitere Streuung über Impact und Jahre im Falle der S3-Leitlinie könnte in der größeren Zahl der beteiligten Personen liegen. Fast 70 % der zitierten Literatur in der S3-Leitlinie ist mindestens einer Person aus der Leitliniengruppe bekannt gewesen, da diese in früheren Arbeiten zitiert wurde. Durch eine große Leitliniengruppe können Artikel in die erarbeitete Leitlinie einfließen, die bei einer systematischen Recherche unentdeckt blieben. Dies würde eine Erklärung dafür liefern, weshalb auch weniger hoch-gerankte Artikel in Leitlinien Eingang finden. Die Autor*innen einer Leitliniengruppe sind schließlich Expert*innen in ihrem Fach und vertrauen eben nicht nur auf eine systematische Literatursuche, sondern stützen sich auf ihr früheres Wissen zu der behandelten Thematik.

Ein wesentlicher Befund unserer Studie ist, dass Artikel, die in den untersuchten Leitlinien zitiert werden, signifikant höhere Zitationszahlen aufweisen als andere Publikationen gleicher Zeitschriften und Jahrgänge. Das durchschnittliche Perzentil beträgt für die S2e-Leitlinie 80,4 % und für die S3-Leitlinie 74,4 %. Somit decken sich unsere Ergebnisse mit der Studie von Thelwall und Mafahi (2016), die gezeigt haben, dass Leitlinien-Referenzen höhere mittlere Zitationszahlen aufweisen als vergleichbare Artikel, die im selben Jahr, in derselben Zeitschrift und in derselben Ausgabe veröffentlicht wurden. Laut Thelwall und Mafahi (2016) werden die in den Leitlinien zitierten Referenzen ohnehin höher zitiert als vergleichbare Artikel, sodass eine zusätzliche Betrachtung ihres Leitlinien-Impacts sich erübrigt und diese Leitlinien-Referenzen nicht durch ein gesondertes Maß noch jenseits ihrer hohen Zitationszahlen belohnt werden sollten. Dies gilt jedoch nicht für neu erschienene Artikel, die aufgrund ihrer kurzen Sichtbarkeit weniger Zitationen akkumulieren können als ältere Artikel.

Anhand der untersuchten Leitlinien zeigt sich jedoch auch, dass es Referenzen gibt, die zwar wenig zitiert wurden, aber dennoch relevant für die klinische Praxis sind. Solange es sich nicht ausschließlich um Selbstzitationen der Leitlinien-Autor*innen handelt, sollten auch unterdurchschnittlich oft zitierte Artikel in der Forschungsevaluation eine entsprechende Anerkennung erhalten.

Eine mögliche Honorierung dieser Artikel könnte durch die Einbeziehung des sogenannten Leitlinien-Impact-Faktors von Zeitschriften in die Evaluation medizinischer Fakultäten erfolgen. Zeitschriften, die in Leitlinien zitiert werden, würden gemäß ihrer Zitationshäufigkeit einen entsprechenden Impact-Wert (analog zum JIF) erhalten. Somit würden Artikel, die in Zeitschriften erscheinen und einen Leitlinien-Impact-Faktor aufweisen, eine Anerkennung bekommen, die

komplementär zu der Anerkennung auf Basis von Zitierungen in Fachzeitschriften wäre.

Zusammenfassend lässt sich festhalten, dass unsere Studie erste wichtige Erkenntnisse darüber liefert, welche Charakteristika die in Leitlinien referenzierte Literatur innehat. Während in dieser kleinen Studie nur eine S2e- und eine S3-Leitlinie ausgewertet wurden, könnte eine umfangreichere Analyse einen besseren Einblick in die Muster der Referenzen hinsichtlich ihrer Anzahl, ihres Alters und ihrer Bekanntheit unter den Leitlinien*Autorinnen aus früheren Arbeiten liefern. Eine Folgestudie mit mehr untersuchten Leitlinien könnte eine größere Aussagekraft besitzen, ob Referenzen in Leitlinien im Durchschnitt häufiger zitiert werden als vergleichbare Artikel desselben Jahrgangs und derselben Zeitschrift.

Danksagung Die Autor*innen danken Herrn Professor Dr. Herrmann-Lingen, Direktor der Klinik für Psychosomatische Medizin und Psychotherapie der Universitätsmedizin Göttingen, für seine Anregungen bei der Durchsicht dieses Manuskripts und die wertvollen Verbesserungsvorschläge.

Literatur

- Andersen, J. P. (2013). Association between quality of clinical practice guidelines and citations given to their references. <http://arxiv.org/pdf/1301.5782v1>.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA*, 295(1), 90–93. <https://doi.org/10.1001/jama.295.1.90>.
- Kryl, D., Allen, L., Dolby, K., Sherbon, B., & Viney, I. (2012). Tracking the impact of research on policy and practice: Investigating the feasibility of using citations in clinical guidelines for research evaluation. *British Medical Journal Open*, 2(2), e000897. <https://doi.org/10.1136/bmjopen-2012-000897>.
- Thelwall, M. & Kousha, K. (2016). Are citations from clinical trials evidence of higher impact research? An analysis of ClinicalTrials.gov. *Scientometrics*, 109(2), 1341–1351. <https://doi.org/10.1007/s11192-016-2112-1>.
- Thelwall, M., & Maflahi, N. (2016). Guideline references and academic citations as evidence of the clinical value of health research. *Journal of the Association for Information Science and Technology*, 67(4), 960–966. <https://doi.org/10.1002/asi.23432>.
- Traylor, C., & Herrmann-Lingen, C. (2023). Does the journal impact factor reflect the impact of German medical guideline contributions? *Scientometrics*, 128(3), 1951–1962. <https://doi.org/10.1007/s11192-023-04647-z>.
- Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: Potential benefits, limitations, and harms of clinical guidelines. *BMJ (Clinical research ed.)*, 318(7182), 527–530. <https://doi.org/10.1136/bmj.318.7182.527>.

Zhang, F., & Wu, S. (2021). Measuring academic entities' impact by content-based citation analysis in a heterogeneous academic network. *Scientometrics*, 126(8), 7197–7222. <https://doi.org/10.1007/s11192-021-04063-1>.

Traylor, Christopher Christopher Traylor ist Psychologe, Wissenschaftlicher Mitarbeiter und Doktorand der Klinik für Psychosomatische Medizin und Psychotherapie an der Universitätsmedizin Göttingen. Im Projekt QuaMedFo führt er Replikations-, Tagungs- und Leitlinienanalysen durch und befasst sich mit der Analyse erwarteter Verhaltenskonsequenzen aus der Einführung neuer Indikatoren.

Aman, Valeria Valeria Aman studierte Bibliotheks- und Informationswissenschaft an der Humboldt-Universität zu Berlin und an der Royal School of Library and Information Science in Kopenhagen. Seit 2013 beteiligt sie sich am Kompetenzzentrum Bibliometrie an der Entwicklung einer robusten Infrastruktur für bibliometrische Anwendungen. Nach dem Wechsel an das DZHW in 2016 hat sie diverse bibliometrische Studien und Evaluationen durchgeführt und beendet aktuell ihre Promotion zur Entwicklung bibliometrischer Methoden zum Nachweis von Wissenstransfer.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Nutzung von Patentdaten zur Erfassung der wirtschaftlichen Verwertung von Forschung

Klaus Lippert und Konrad U. Förstner

1 Einleitung und Einordnung

Die Leistung medizinischer Forschungseinrichtungen lässt sich nicht nur aus dem Blickwinkel des Publikationsoutputs, sondern auch aus dem Blickwinkel der wirtschaftlichen Verwertbarkeit betrachten und bewerten. Um die wirtschaftliche Verwertbarkeit von Forschung zu messen können insbesondere daraus hervorgegangene Patente betrachtet werden.

Diese Verknüpfung von wissenschaftlichen Publikationen und Patenten ist eine Möglichkeit die Verwertung des wissenschaftlichen Outputs und damit den Wissenstransfer aus der Forschung in die Wirtschaft sichtbar zu machen. Hierbei gilt jedoch zu beachten, dass eine Patentanmeldung keine Garantie für einen wirtschaftlichen Erfolg ist, noch spiegelt eine Patentanmeldung zwingend die Qualität der wissenschaftlichen Arbeit wider. So sind es vorwiegend technische Neuerungen oder Verfahren, welche patentiert werden können und nicht der wissenschaftliche Erkenntnisgewinn. Auch die Fachrichtung und speziell die Forschungsfrage spielen eine entscheidende Rolle bei der Patentierbarkeit von (Teil-)Ergebnissen. Das Ziel des Verbundprojekts QuaMedFo war die ganzheitliche Betrachtung von Forschungsleistung aus verschiedensten Blickwinkeln. Vor

K. Lippert (✉) · K. U. Förstner (✉)

ZB MED – Informationszentrum Lebenswissenschaften, Köln, Deutschland

E-Mail: lippert@zbmed.de

K. U. Förstner

E-Mail: foerstner@zbmed.de

K. U. Förstner

Institut für Informationswissenschaft, TH Köln, Köln, Deutschland

© Der/die Autor(en) 2024

S. Biesenbender und J. Hartstein (Hrsg.), *Qualitätsmessung als Prisma*, Higher Education Research and Science Studies,

https://doi.org/10.1007/978-3-658-43683-4_6

diesem Hintergrund beschäftigt sich der vorliegende Artikel mit der Weiterentwicklung von bestehenden Methoden zur Verknüpfung von Forschung und wirtschaftlicher Verwertung durch Betrachtung von Patentaktivitäten.

Der gängige Weg die beiden Textklassen Publikation und Patent in Beziehung zu bringen sind übereinstimmende Vor- und Nachnamen der Autorinnen und Autoren einerseits und der Erfinderinnen und Erfinder andererseits. Da die Vornamen in beiden Textklassen nicht zwingend ausgeschrieben werden müssen, bleibt als kleinster gemeinsamer Nenner nur das Schema *Nachname, Initialen des/der Vornamen*. Dadurch wird allerdings die Mehrdeutigkeit der Zuordnungen noch einmal deutlich erhöht. Vor allem bei nur einem oder zwei übereinstimmenden Namen gibt es deutlich mehr falsche als richtige Zuordnungen. Dieses Problem der Mehrdeutigkeit wurde bisher wie folgt angegangen:

1. Händische Verarbeitung und Prüfung: Ausgehend von Beschäftigtenlisten der einzelnen Einrichtungen (z. B. van Dongen et al., 2014) bis hin zur Erstellung von Patentdatenbanken mit Verknüpfung zu wissenschaftlichem Personal (z. B. KEINS Datenbank (Lissoni et al., 2008) werden die Publikations-Patent-Verknüpfungen händisch geprüft, z. B. mit Online-Befragungen oder persönlichem Kontakt (Dahlborg et al., 2013).
2. Maschinelle Verarbeitung: Bei großskaligen Untersuchungen müssen verschiedene weitere Unterscheidungsmerkmale aus den Metadaten der beiden Textklassen hinzugezogen werden. So verwenden Dornbusch und Neuhäusler (2015) z. B. einen bestimmten Zeitraum, in dem eine Publikation einem Patent zugeordnet werden darf und zusätzlich eine maximale geographische Distanz von 30 km zwischen wissenschaftlicher Einrichtung und Wohnort des Erfinders als Bestätigung eines gefundenen Publikations-Patent-Paares.

Im Rahmen des QuaMedFo-Projektes wurden zwei weitere Werkzeuge zur Verifizierung der Publikations-Patent-Paare entwickelt bzw. geprüft: Die Untersuchung der gemeinsamen Referenzen in beiden Textarten ist bereits in einem anderen Kontext erfolgreich angewendet worden (Gurulingappa et al., 2010) und konnte mit Erfolg auf diese Aufgabenstellung übertragen werden. Das zweite entwickelte Werkzeug ist der inhaltliche Vergleich der beiden Textarten. Die Untersuchungen zu den beiden Verifizierungsarten und Details zur Implementierung finden sich bei Lippert und Förstner (in Erarbeitung). Der vorliegende Artikel befasst sich nur mit einer der beiden Verifizierungsarten: dem inhaltlichen Vergleich der Texte und der Anwendung auf die drei Pilotfakultäten aus dem QuaMedFo-Projekt.

Der Artikel beginnt mit der Vorstellung der zugrunde liegenden Daten, gefolgt von der Beschreibung der Vorgehensweise sowie dem Ergebnis für die Pilotfakultäten und schließt mit einem Ausblick ab.

Die im Anschluss beschriebene methodische Vorgehensweise lässt sich zusammenfassen als:

1. verschiedene Entwicklungsstufen eines Patenten zu einer Patentfamilie zusammenfassen
2. Publikationen aus den Pilotfakultäten in der medizinischen Literaturdatenbank PubMed matchen
3. Mögliche Paare (Roh-Paare) von Publikationen und Patenten durch gleiche Autoren- bzw. Erfindernamen ermitteln
4. die in den Roh-Paaren enthaltenen Dokumente mit standardisiertem medizinischen Fachvokabular (genauer: MeSH-Ontologie) verschlagworten und die Inhalte der Dokumente anschließend in einen Vektorraum überführen.
5. Verifizierung und Ranking der gefundenen Roh-Paare mittels sogenannter Kosinus-Ähnlichkeit im Vektorraum
6. Filtern der gefundenen Paare mittels internationaler Patentklassen (IPC). Es werden nur Patente mit einer bestimmten Patentklasse anerkannt. Diese wurden statistisch ermittelt.

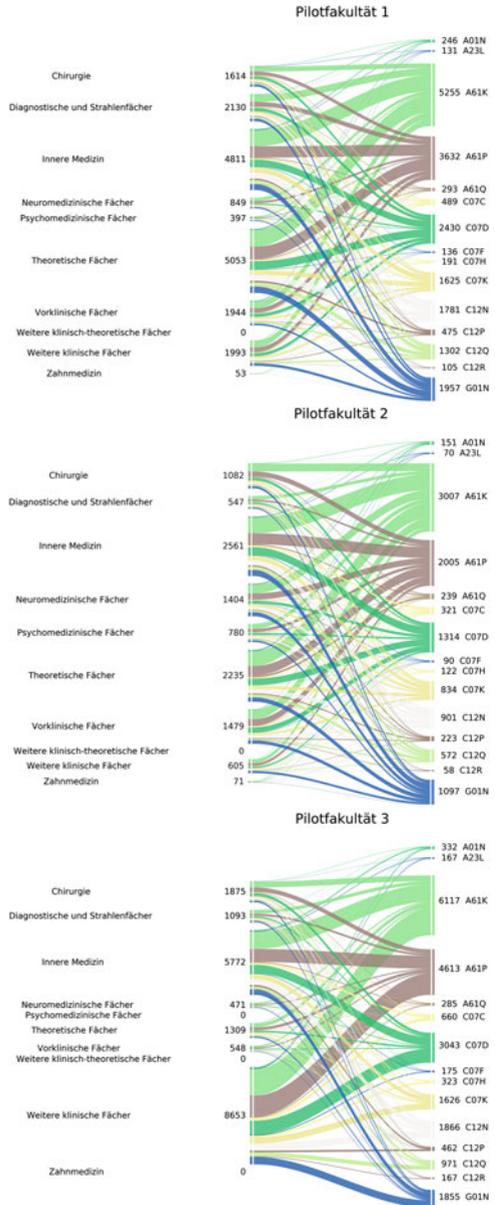
2 Datengrundlagen und -beschreibung

Als Datengrundlage dienen Publikationsdaten von drei Pilotfakultäten des Projektes aus den Jahren 2005 bis 2018. Die einzelnen Einrichtungen sind medizinischen Fachgruppen zugeordnet worden. Diese Fachgruppen sind in Abb. 1 jeweils links aufgelistet. Die gelieferten Daten sind durch die Pilotfakultäten zum Teil mehreren Einrichtungen (und somit mehreren Fachgruppen) zugeordnet, weil es sich entweder um Forschungsk Kooperationen verschiedener Einrichtungen handelt oder weil die Zuordnung der Finanzierung der Autoren nicht eindeutig ist. Die dadurch entstehende Mehrdeutigkeit ist nicht auflösbar und muss bei der Interpretation der Ergebnisse berücksichtigt werden.

Mittels des eindeutigen Bezeichners DOI (Digital Object Identifier) wurden die Publikationen mit dem PubMed-Datensatz verknüpft und somit stehen MeSH-Mainheadings für jeden Artikel zur Verfügung. Diese werden in Abschn. 3 erläutert.

Abb. 1

Patent-Einreichungen der Fachgruppen in die TOP 10 der Patentklassen für die drei Pilotfakultäten



Die Datengrundlage der Patente sind die frei verfügbaren Daten des europäischen Patentamts (EPO)¹ in der xml-Repräsentation² aus den Jahren 2007 bis 2020. Neben den Metadaten sind dort auch vollständige Beschreibungstexte in einer oder mehreren Sprachen (Englisch, Deutsch, Französisch) vorhanden.

Patentdokumente lassen sich grob in A- und B-Patente einteilen. Wobei A-Patente die Patentanträge sind und B-Patente bewilligte Patente. Eine auf diesen Buchstaben folgende Zahl legt den zeitlichen Ablauf der Einreichungen fest. Das erstmalig eingereichte Patent A1 muss zum Einreichungsdatum (engl. filing date) die sogenannte „Neuerungs-Klausel“ erfüllen, d. h. der zu patentierende Inhalt darf vorher nicht veröffentlicht worden sein. Daraus ergibt sich bei der Verknüpfung Publikation-Patent: Das Einreichungsdatum des Patents muss vor dem Veröffentlichungsdatum der Publikation liegen. An dieser Stelle wird der Argumentation von Dornbusch und Neuhäusler (2015) gefolgt und die Publikation darf zusätzlich maximal 1,5 Jahre nach der Patenteinreichung erfolgen. Die verschiedenen A- und B-Patente haben die gleichen Metadaten und können so als eine Patent-Familie zusammengefasst werden. Die enthaltenden Volltexte der Patentbeschreibung werden in diesen Familien zu einem Text zusammengefasst.

Das europäische Patentamt teilt Patente in die internationalen Patentklassen (IPC) ein, welche durch die Straßburg Vereinbarung von 1975 (Wikipedia, 2022) weltweit standardisiert sind. Ein Patent wird normalerweise mehreren (gleichwertigen) Patentklassen zugeordnet. Die dadurch entstehende Mehrdeutigkeit kann nicht aufgelöst werden und muss bei der Interpretation der Ergebnisse ebenfalls berücksichtigt werden.^{3,4} Dies geschieht durch die abschließende Filterung nach Patentklassen.

3 Verknüpfung von Patenten und Publikationen

Für die Verknüpfung der Patente und Publikationen wird nach gleichen Namen der Autoren und Erfinder gesucht. Die einzelnen Namen werden dazu, wie in der obigen Einleitung bereits erwähnt, auf die Form *Nachname, Initialen des/*

¹ www.epo.org

² <https://data.epo.org/publication-server/rest/v1.2/publication-dates>

³ Siehe z.B. hier: <https://at.espacenet.com/help?locale=de-AT&method=handleHelpTopic&topic=ipc>

⁴ Eine durchsuchbare Beschreibung der einzelnen Klassen und Unterklassen findet sich z.B. bei der World Intellectual Property Organization (WIPO): <https://ipcpub.wipo.int>.

der Vornamen vereinheitlicht. Untersuchungen von Milojevic (2013) zeigen die Gültigkeit dieser Vereinfachung.

4 Validierung mittels inhaltlichen Vergleichs der beiden Textklassen

Für den inhaltlichen Vergleich von Dokumenten aus den beiden Textklassen Publikation und Patente wird auf die 'Medical Subject Headings' (MeSH) (National Library of Medicine (US) (NLM), 2022) zurückgegriffen, welche einmal jährlich von der 'National Library of Medicine' (NLM) in englischer Sprache herausgegeben werden. Dieser hierarchisch aufgebaute Thesaurus ermöglicht es medizinische Begriffe und ihre Synonyme auf einen einheitlichen Bezeichner zurückzuführen und somit den Inhalt eines medizinischen Textes zu beschreiben. Neben der englischen Version des MeSH werden in dieser Arbeit auch die offizielle deutsche (Deutsches Institut für Medizinische Dokumentation und Information (DIMDI), 2019) und die französische (Institut national de la santé et de la recherche médicale (INSERM), 2019) Übersetzung verwendet, weil bei den europäischen Patenten der EPO alle drei Sprachen äquivalent benutzt werden.

So werden aus den Volltexten der Patentbeschreibungen die MeSH-Terme mittels einer einfachen Wörterbuchsuche extrahiert und auf die englischen Hauptbegriffe zurückgeführt. Auf der Publikationsseite können die englischen MeSH-Hauptterme direkt den Metadaten des Datensatzes entnommen werden. So entsteht für jedes Dokument eine Liste von MeSH-Termen, welche den Inhalt beschreibt.

Durch die Verwendung des BERT-Base-Modells (Devlin et al., 2018) können diese Text-Kondensate in einem hochdimensionalen ($\text{dim} = 300$) Vektorraum abgebildet werden und die – per Namensvergleich – gefundenen rohen Patent-Publikations-Paare leicht mathematisch verglichen werden. Diese hochdimensionalen Vektoren (sog. „embeddings“) werden mithilfe der Kosinus-Ähnlichkeit verglichen: Der Wert liegt hierbei immer zwischen 0 und 1, wobei 1 übereinstimmende Dokumente sind und ein Wert von 0 keinerlei inhaltliche Übereinstimmung anzeigt. Untersuchungen mit größeren Datensätzen (Lippert und Förstner in Erarbeitung) haben gezeigt, dass Verknüpfungen mit drei oder mehr gleichen Namen eine hohe Kosinus-Ähnlichkeit ($> \text{ca. } 0,7$) aufweisen. Bei drei oder mehr gleichen Autoren- bzw. Erfindernamen wird also von einer Übereinstimmung ausgegangen und das Publikations-Patent-Paar als gültig angesehen. Der gefundene Schwellenwert wird auf die vorliegenden Daten übertragen.

Somit ist durch den beschriebenen inhaltlichen Vergleich mittels MeSH-Termen und "word embeddings" ein Werkzeug gefunden, um die gefundenen Roh-Paare zu validieren.

5 Weiterer Filter: „medizinische“ Patentklassen

Die internationalen Patentklassen (IPC) umfassen alle Teilbereiche des menschlichen Lebens, welche patentiert werden können. Sehr viele davon sind für Patente, die aus medizinischer Forschung hervorgegangen sind, undenkbar. Das Vorgehen des Filterns mittels Patentklassen ist üblich und erfolgt durch manuelles Festlegen erlaubter Patentklassen (z. B. bei Dornbusch & Neuhäusler, 2015). Im Unterschied dazu konnten Lippert und Förstner (in Erarbeitung) zeigen, dass sich die Teilmenge der gesicherten medizinischen Patente statistisch von der Gesamtmenge der Patente in ihren IPC-Klassen unterscheidet. Gesicherte medizinische Patente ergeben sich, wenn z. B. eine Universität und/oder Universitätsklinik als Patentbesitzer eingetragen ist. Es wurden dadurch spezielle medizinische Patentklassen gefunden, welche als weitere Filteroption für die rohen Publikations-Patent-Paare verwendet werden.

6 Ergebnis und Diskussion

In Abb. 1 sind die Patenteinreichungen der einzelnen Fachgruppen der drei Pilotfakultäten in die TOP 10 der Patentklassen als absolute Zahlen dargestellt. Auf der linken Seite sind jeweils die einzelnen Fachgruppen zu sehen, auf der rechten Seite die Patentklassen, wobei sich hier auf Sektion, Klasse und Unterklasse beschränkt wurde. Eine Definition dieser Klassen findet sich in Tab. 1.

Wegen der oben beschriebenen Mehrdeutigkeiten sind keine quantitative Aussage und auch kein Vergleich der Pilotfakultäten möglich. Die Mehrdeutigkeit auf Publikationsseite durch mehrfache Zuordnung zu einzelnen Einrichtungen wäre durch eine genauere Datengrundlage aufzulösen. Für einen Vergleich der Pilotfakultäten untereinander benötigt es jedoch Daten zu den Größen der einzelnen Einrichtungen, um eine Normierung durchzuführen. Auf der Patentseite liegt die Mehrdeutigkeit in der Natur der Daten und ist somit nicht auflösbar: Ein Patent ist gleichberechtigt in mehreren Patengruppen vertreten.

Dennoch kann ein qualitativer Vergleich der Pilotfakultäten erfolgen, bei dem die unterschiedlichen Forschungsausrichtungen deutlich werden. Im Folgenden einige Beispiele:

Tab. 1 Beschreibung der in Abb. 1 dargestellten Patent Klassen (IPC). nach: <https://ipcpub.wipo.int>

IPC	Beschreibung
A61K	Preparations for medical, dental or toilet purposes
A61P	Specific therapeutic activity of chemical compounds or medical preparations
A61Q	Specific use of cosmetics or similar toilet preparations
C07C	Acyclic or Carbocyclic compounds
C07D	Heterocyclic compounds
C07K	Peptides
C12N	Microorganisms or enzymes; compositions thereof; propagating, preserving, or maintaining microorganisms; mutation or genetic engineering; culture media
C12P	Fermentation or enzyme-using processes to synthesise a desired chemical compound or composition or to separate optical isomers from a racemic mixture
C12Q	Measuring or testing processes involving enzymes, nucleic acids or microorganisms
G01N	Investigating or analysing materials by determining their chemical or physical properties

Einreichungen in IPC A61P, C07K und G01N: während bei Pilotfakultät 1 die *Innere Medizin* und die *theoretischen Fächer* den Hauptanteil liefern, übernehmen bei Pilotfakultät 3 die *weiteren klinischen Fächer* den Part der *theoretischen Fächer*. Bei Pilotfakultät 2 sind die Unterschiede der Fachgruppen nicht so prägnant.

Einreichungen aus der Zahnmedizin: Nur in Pilotfakultät 1 und 2 wurden Patente aus dem Bereich der Zahnmedizin eingereicht. In beiden Fällen mit sehr geringer Anzahl.

7 Ausblick

Die im Rahmen dieses Teilprojektes entwickelte Datenverarbeitung (Lippert und Förstner in Erarbeitung) wurde auf die Daten der drei Pilotfakultäten des Projekts QuaMedFo angewandt. Die hierbei eingesetzte automatisierte statistische Einteilung in „erlaubte Patentklassen“ und der inhaltliche Vergleich beider Textklassen stellen sinnvolle Werkzeuge in der Bildung von validen Publikations-Patent-Paaren dar.

Auch wenn ein quantitativer Vergleich der drei Pilotfakultäten des Forschungstransfers in die Wirtschaft aufgrund der Datengrundlage nicht möglich war, stellen die Ergebnisse einen Mehrwert für Recherchen im medizinischen Bereich dar. Die validierten Verknüpfungen von Patenten und Publikationen dieser Pilotrichtungen und perspektivisch für alle vorhandenen Datensätze sollen in Zukunft bei der Recherche über LIVIVO, das Suchportal für Lebenswissenschaften (Müller et al., 2017), einem breiten Publikum zur Verfügung gestellt werden. Damit wird jeder Nutzende u. a. befähigt Vergleiche auf verschiedenen Aggregations Ebenen durchzuführen: angefangen bei Einzelpersonen, über einzelne Projekte und Fachgruppen, bis hin zu Instituten. Mit zusätzlichen Informationen, z. B. Personalschlüsseln, können auch universitätsübergreifend Vergleiche erstellt werden. Neben der Bewertung der allgemeinen wirtschaftlichen Verwertbarkeit auf den beschriebenen Aggregationsebenen, werden die zur Verfügung gestellten Ergebnisse durch die Information der Patentklassen Einblicke in die fachliche Ausrichtung aus Patent-technischer Sicht ermöglichen.

Danksagung Die Autoren danken V. Aman (DZHW) für die Bereitstellung und Aufarbeitung (u. a. mit Identifiern) dieser Daten. Großer Dank auch an C. Herrmann-Lingen (Universitätsmedizin Göttingen) für die Ausarbeitung der Fachgruppen und die Zuordnung jeder Einrichtung in diese.

Literatur

- Dahlborg, C., Lewensohn, D., & Sundberg, C. J. (2013). Investigating inventive productivity at sweden's largest medical university. *International Journal of Technology Transfer and Commercialisation*, 12(1–3), 102–120. <https://doi.org/10.1504/IJTTC.2013.064154>
- Deutsches Institut für Medizinische Dokumentation und Information (DIMDI). (2019). German mesh, 2019. <https://www.dimdi.de/dynamic/de/klassifikationen/weitere-klassifikationen-und-standards/mesh/>. Zugegriffen: 1. Sept. 2020.
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. <https://arxiv.org/abs/1810.04805>
- Dornbusch, F. & Neuhäusler, P. (2015). Academic patents in germany. Studien zum deutschen Innovationssystem 6–2015, Berlin. <http://hdl.handle.net/10419/156616>
- Gurulingappa, H., Müller, B., Klinger, R., Mevissen, H.-T., Hofmann-Apitius, M., Friedrich, C. M., & Fluck, J. (2010). Prior art search in chemistry patents based on semantic concepts and co-citation analysis. Text REtrieval Conference (TREC) 2010. <https://publica.fraunhofer.de/handle/publica/368883>
- Institut national de la santé et de la recherche médicale (INSERM). (2019). Le mesh bilingue anglais français. <http://mesh.inserm.fr/FrenchMesh/>. Zugegriffen: 20. Apr. 2022.

- Lippert, K. & Förstner, K. U. (in Erarbeitung). Knowledge transfer from science to economy illustrated via patent – publication pairs: Reducing ambiguities with word embeddings and references. *Im Einreichungsprozess bei: Scientometrics*.
- Stasa Lissoni, F., Llerena, P., Mckelvey, M., & Sanditov, B. (2008). Academic patenting in europe: New evidence from the keins database. *Research Evaluation*, 17(2), 87–102. <https://doi.org/10.3152/095820208X287171>
- Milojević, S. (2013). Accuracy of simple, initials-based methods for author name disambiguation *Journal of Informetrics*, 7(4), 767–773, 2013. <https://doi.org/10.1016/j.joi.2013.06.006>
- Müller, B., Poley, C., Pössel, J., et al. (2017). LIVIVO – The Vertical Search Engine for Life Sciences. *Datenbank Spektrum*, 17, 29–34. <https://doi.org/10.1007/s13222-016-0245-2>
- National Library of Medicine (US) (NLM). (2022). Medical Subject Headings MeSH. www.nlm.nih.gov/mesh/meshhome.html. Zugegriffen: 20. Apr. 2022.
- van Dongen, P., Winnink, J., & Tijssen, R. (2014). Academic inventions and patents in the netherlands: A case study on business sector exploitation. *World Patent Information*, 38, 27–32. <https://doi.org/10.1016/j.wpi.2014.03.002>
- Wikipedia. Strasbourg agreement — Wikipedia, the free encyclopedia. (2022). https://en.wikipedia.org/w/index.php?title=Strasbourg_Agreement_Concerning_the_International_Patent_Classification&oldid=1082987276. Zugegriffen: 20. Apr. 2022.

Lippert, Klaus Dr. Klaus Lippert studierte Geophysik an der LMU München und der Universität zu Köln. Anschließend promovierte er im Bereich der zerstörungsfreien offshore Grundwassersuche mittels Elektromagnetischer Verfahren. Nach mehrjähriger Erfahrung in der IT-Beratung und Datawarehouse-Entwicklung war er ab August 2019 bis Projektende bei ZB MED im QuaMedFo-Projekt.

Förstner, Konrad Konrad Förstner hat eine gemeinsame Professur für Data und Information Literacy an der TH Köln und bei ZB MED – Informationszentrum Lebenswissenschaften. Seine Forschung umfasst u.a. bioinformatische Analysen von Hochdurchsatzsequenzierdaten und die Analyse wissenschaftlicher Literatur mittels maschineller Lernverfahren und Strukturierung extrahierten Wissens sowie die Kombination beider Felder. Zudem ist seine Gruppe für die Entwicklung von LIVIVO – eine Suchmaschine für lebenswissenschaftliche Literatur – verantwortlich.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Input-Output-Relationen – Zu den Voraussetzungen einer Effizienzmessung medizinischer Fakultäten anhand von Drittmitteln

Judith Hartstein

1 Einleitung

Der Aufstieg von Forschungsevaluationen wird in der Literatur vor allem im Zusammenhang mit dem New Public Management diskutiert (vgl. Hornbostel, 2010), fußt aber auf der quantitativ vergleichenden Evaluation von Höheren Lehranstalten in den USA, wo sie Anfang des 20. Jahrhunderts noch stark disziplinär geprägt und auf angewandte Wissenschaften ausgerichtet war (Espeland & Sauder, 2016, S. 9 ff.; Wilbers et al., 2021). Die vergleichende Bewertung von Medical Schools in den USA, hat sich dabei im Zeitraum 1850–1930 von vagen Diskurskonstellationen hin zu tabellarischen Rankings entwickelt und auch zur *inhaltlichen* Standardisierung medizinischer Wissenschaft beigetragen, weil die gewählten Qualitätsmaßstäbe eine Grenzziehung zwischen wissenschaftlicher Medizin und sogenannter „Quacksalberei“ bewirkten (Wilbers et al., 2021). Die soziologische Forschung zu Standardisierung vertritt noch umfassender, dass Standards erstens niemals wertfrei sein können und zweitens immer soziale Konsequenzen haben (Bowker & Star, 2000; Timmermans & Epstein, 2010). Insbesondere der quantitative Vergleich erfordert nach Heintz (2010) eine Kombination von „Gleichheitsunterstellung und Differenzbeobachtung“ (S. 164), in

J. Hartstein (✉)

Robert K. Merton Zentrum für Wissenschaftsforschung, Humboldt-Universität zu Berlin, Berlin, Deutschland

E-Mail: hartstein@dzhw.eu

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Berlin, Deutschland

© Der/die Autor(en) 2024

S. Biesenbender und J. Hartstein (Hrsg.), *Qualitätsmessung als Prisma*, Higher Education Research and Science Studies, https://doi.org/10.1007/978-3-658-43683-4_7

109

Gestalt von Kriterien und Indikatoren, welche neue Vergleichsrealitäten schaffen – eine kommunikationssoziologische Perspektive, die unter der Phrase „Trust in Numbers“ (Porter, 1996) in der Soziologie der Bewertung etabliert ist. Im Zuge strategischer Kalkulationen werden diese Vergleichsrealitäten durch Forschungsevaluation dann auch handlungsleitend für Organisationen (Wiesenthal, 2007).¹ In diesem Sinne soll hier die Evaluation medizinischer Fakultäten als Ausprägung organisationaler Steuerung verstanden werden, wobei Evaluationsergebnisse jeweils strategische Kursanpassungen zur Folge haben (sollen), welche sich häufig in veränderter Mittelzuweisung ausdrücken, weshalb die Mittelempfänger im Wettbewerb stehen.

Medizinische Fakultäten in Deutschland müssen ihre strategischen Kalkulationen dabei gleich auf mehrere Wettbewerbe beziehen: die *inter-universitäre* Konkurrenz um Reputation und Mittel schlägt dabei auf die unteren Organisationsebenen durch, sodass medizinische Fakultäten, Institute oder auch einzelne Lehrstühle oder Wissenschaftler*innen mit ihren Äquivalenten anderer Hochschulen konkurrieren. *Intra-universitär bzw. inter-fakultär* konkurrieren Fakultäten derselben Hochschule um Mittel und Einfluss, ebenfalls mit Wirkung auf die darunter liegenden Organisationsebenen und erzeugen so *intra-fakultär* subdisziplinäre Konkurrenz zwischen medizinischen Instituten oder Lehrstühlen. Die Steuerungsinstanzen in den Fakultäten – üblicherweise die Dekanate – müssen alle diese Wettbewerbe adressieren und dabei zwei gegensätzliche Rollen einnehmen. Inter-universitär und inter-fakultär sind sie Wettbewerbsteilnehmer und bestrebt, dass die eigene Fakultät ein möglichst großes „Stück vom Kuchen“ abbekommt, wenn es um die Mittelzuteilung geht (z. B. über Hochschulverträge oder öffentliche Drittmittel). Intra-fakultär sind sie hingegen zentral für die Mittelverteilung auf Institute und Lehrstühle unter Berücksichtigung von Steuerungsanliegen und führen die Leistungsorientierte Mittelvergabe (LOM) durch – ein zentrales Instrument der leistungsorientierten Mittelallokation in medizinischen Fakultäten in Deutschland (Herrmann-Lingen in diesem Band). Durch die Einbeziehung von inter-universitär und publikationsbasiert gedachter „Exzellenz“ in die LOM sind diese beiden Wettbewerbe verbunden, etwa über die Verwendung von Impact-Faktoren, deren Bezugssystem ja die weltweite Publikationslandschaft ist.

Der vorliegende Beitrag greift nun die intra-fakultäre Perspektive heraus und versteht die Sicherstellung von Effizienz und Effektivität als zentrale Aufgabe des Controllings durch Dekanate (vgl. Dyckhoff & Ahn, 2002). Empirisch wird

¹Für eine ausführlichere Betrachtung der sozialen Konsequenzen quantitativer Forschungsevaluation wird auf Biesenbender (in diesem Band) verwiesen.

untersucht, ob an die an den Fakultäten gehaltenen Daten für die Betrachtung von Effizienz geeignet sind, also des mengenmäßigen Outputs in Bezug auf den Input – im Gegensatz zu Effektivität, welche die Zielgerichtetheit des Outputs beschreiben würde (vgl. Nullmeier, 2001). Dieser Betrachtungsweise unterliegt die Vorstellung, dass Input und Output durch eine Produktionsfunktion verbunden sind,² woran sich etablierte Methoden wie die Data-Envelope-Analyse (Messung von Effizienz einer Einheit relativ zur Vergleichsgruppe) anschließen ließen (beispielsweise bei Clermont, 2016; Gralka et al., 2018).

Dieses Verständnis von Effizienz betrachtet die medizinische Fakultät gewissermaßen als eine Black Box, bei der nur Zu- und Abflüsse beobachtet werden und deren interne Mechanismen verborgen bleiben.³ Dabei wird gemeinhin unterstellt, dass Input und Output einen gerichteten Zusammenhang aufweisen, d. h. bei einem Zuwachs an Input wird auch eine Steigerung des Outputs zu erwarten sein. Es ist in verschiedenen Disziplinen üblich, solche gerichteten Zusammenhänge zunächst als linear anzunehmen, soweit nichts anderes bekannt ist. Im hiermit vorgelegten Beitrag sollen deswegen die Voraussetzungen der Effizienzmessung als Modellannahmen für die lineare Regression verstanden und exemplarisch diskutiert werden.

Es ist dabei zunächst zu klären, welche Kennzahlen eigentlich als Input und welche als Output betrachtet werden sollen, um die Effizienz einer medizinischen Fakultät zu berechnen. Hier sind die Variablen *wissenschaftliche Beschäftigte* (Input) und *Publikationen* (Output) etabliert. Zusätzlich werden in neuerer Literatur auch *Drittmittel* als Output verstanden (Angelova et al., 2021), jedoch müsste hier eigentlich je nach Evaluations- oder Steuerungsanliegen eine unterschiedliche Zuordnung getroffen werden. So ist aus der Perspektive der Universitäten und Forschungseinrichtungen eine große Zahl an eingeworbenen Drittmitteln zwar ein positiver Indikator für Forschungsperformanz, gleichzeitig ist für ein Bundesministerium, welches oft als Drittmittelgeber auftritt, sicher auch relevant, ob mehr Drittmittel beispielsweise zu mehr exzellenter Forschung führen. Drittmittel an den Fakultäten ließen sich somit auch als Input auffassen, was auf die zentralen Forschungsfragen führt:

² Damit soll keineswegs unterstellt werden, dass Controller*innen medizinische Fakultäten als Wissensfabriken verstehen. Diese Unterstellung spielt gleichwohl im Diskurs um die LOM und die Forschungsevaluation generell eine gewisse Rolle und es entzündeten sich daran auch Debatten um Gerechtigkeit von Forschungsevaluation über Fächer oder Disziplinen hinweg.

³ Zu den vermuteten Mechanismen, die über die evaluationsbasierte LOM in mögliche Selbstverstärkungskreise münden, s. Herrmann-Lingen (in diesem Band).

Wäre für die Effizienzmessung eine Perspektive auf Drittmittel als Input möglich und sinnvoll? Und welche Drittmitteldarstellung wäre als Output-Parameter am geeignetsten?

Zur Beantwortung dieser Fragen wurden Daten zu Beschäftigten, Publikationen und Drittmitteln aus drei deutschen medizinischen Fakultäten empirisch verglichen. Die Datenerhebung ging dabei zunächst von der übergeordneten Perspektive der Forschungsevaluation des Wissenschaftsrates aus, der den „KDSF – Standard für Forschungsinformationen in Deutschland“ für Universitäten und Forschungseinrichtungen vorgeschlagen, spezifiziert und evaluiert hat (Wissenschaftsrat, 2013, 2016, 2020). Dort wird etwa empfohlen, innerhalb der Hochschulen über Beschäftigte Angaben zur Personenzahl und zu Vollzeitäquivalenten bereit zu halten, Drittmittelleinnahmen und Drittmittelzusagen zu erheben sowie Publikationszahlen zu verfolgen – und zwar jeweils mit spezifischen Ausdifferenzierungen etwa zu Organisationseinheit und Fach. Der KDSF-Standard konkurriert dabei gewissermaßen mit der wilden Standardisierung von Forschungsinformationen im Zuge der Verbreitung von Forschungsinformationssystemen, die den Hochschulen ihre Reporting-Aufgaben erleichtern sollen (Ebert et al., 2021), die teilweise für die Umsetzung der Leistungsorientierten Mittelvergabe (LOM) zum Einsatz kommen und von denen wir annehmen können, dass sie auf die administrativen Strukturen der Hochschulen zurückwirken.

Trotz der Anerkennung von Standardisierungsbemühungen muss die Beforschung medizinischer Forschung in besonderem Maße die (publikations-)kulturellen Unterschiede zwischen den verschiedenen Subdisziplinen berücksichtigen. Dies schließt an die kritische Diskussion zur unzureichenden Berücksichtigung von Fächerunterschieden bei der Forschungsevaluation im Allgemeinen an (vgl. Wilsdon et al., 2015), die auch der Wissenschaftsrat aufgreift, sich deshalb für die fachbezogene Datenerhebung ausspricht (Wissenschaftsrat, 2016, 2020) und dafür die Fächerklassifikation des Statistischen Bundesamtes hervorhebt (Wissenschaftsrat, 2016, S. 36), welche mit dem Hochschulstatistikgesetz (HStatG) für die Berichterstattung durch die Hochschulen festgeschrieben ist. Fächerklassifikationen bringen dabei jedoch in der Praxis häufig Probleme mit sich, da zum Beispiel interdisziplinäre Felder oft auf disziplinäre Hintergründe etwa von Personen oder Publikationen heruntergebrochen werden müssen (siehe z. B. Frietsch et al., 2021).⁴

⁴ Folgerichtig wurde auf Initiative des Wissenschaftsrates durch die Projektgruppe „Fächerklassifikation und Thesauri“ eine Forschungsfeldklassifikation erarbeitet, die Analysen für interdisziplinäre Forschungsfelder möglich machen soll (Stiller et al., 2021), jedoch zum Beginn der hier vorgestellten Analyse noch nicht vorlag.

In der hier berichteten Untersuchung wird deshalb eine Binnendifferenzierung vorgenommen, etwa zwischen theoretischen und klinischen Fächern und auch zwischen verschiedenen Fächern und Fächergruppen mit jeweils unterschiedlichen Forschungs- und Publikationstraditionen (s. QuaMedFo-Fächerklassifikation in diesem Band).

Der Beitrag ist im Folgenden so aufgebaut, dass zunächst kurz die Datenerhebung beschrieben wird, woran sich der empirische Ergebnisteil anschließt. Die Diskussion der Ergebnisse mündet dann in Empfehlungen zu den Voraussetzungen einer Effizienzmessung an medizinischen Fakultäten.

2 Daten und Methoden

Zur Beantwortung der aufgeworfenen Forschungsfragen wurde mit den drei Pilotfakultäten des Projekts QuaMedFo vereinbart, dass sie Daten über den Zeitraum 2005–2018 gemäß KDSF-Standard zur Verfügung stellen. Die Untersuchung versteht sich als Fallstudie, deren Datenerhebung nah am Gegenstand ist, die damit aber keine Repräsentativität beanspruchen kann. Durch die vergleichende Betrachtung mehrerer Fakultäten miteinander sollen jedoch trotzdem Erkenntnisse generiert werden, die außerhalb der betrachteten Fakultäten relevant sein können – eine Herangehensweise, die Morgan (2020, S. 206) als „thinking with cases“ bezeichnet hat.

Die Datenanfrage bezog sich auf die Bereiche a) Beschäftigte, b) Drittmittel und c) Publikationen. Gebeten wurde um Datenlieferungen nach dem KDSF-Standard, wobei es sich überwiegend um Aggregatdaten handelt, welche über einzelne Beschäftigte, Publikationen und Projekte Auskunft geben. Die Fakultäten haben umfänglich Daten geliefert, soweit es ihnen möglich war, teilweise auf Basisniveau und teilweise auf Aggregationsebenen unter- oder oberhalb der Organisationseinheiten. Daten wurden für die Analysen dann berücksichtigt, wenn sie für mindestens zwei der drei Fakultäten vergleichbar auf Fächergruppe und Jahr aggregiert oder aggregierbar waren.

2.1 Datenlieferung und -aufbereitung

Die *Beschäftigtendaten* wurden von den drei Fakultäten in unterschiedlicher Ausprägung geliefert: teilweise bezogen sich Angaben nur auf Drittmittelpersonal, teilweise nur auf Köpfe, teilweise nur auf Vollzeitäquivalente (VZÄ), wobei zusätzlich die Daten von Fakultät A nicht den gesamten Beobachtungszeitraum

Tab. 1 Betrachtete Kategorien im Analysedatensatz

	A	B	C
Drittmittelbeschäftigte (VZÄ)	2008–2020	2005–2020	–
Drittmittelausgaben	2009–2020	2007–2018	–
Drittmittelzusagen	–	2005–2018	2015–2018

umfassten. Aus Datenschutzgründen waren die Daten auf verschiedene Weise durch die Fakultäten anonymisiert und aggregiert worden. Aufgrund der Heterogenität der gelieferten Daten musste der Analysedatensatz stark eingeschränkt werden und umfasst nur die Vollzeitäquivalente von Drittmittelbeschäftigten aus den Fakultäten A und B, damit die Gleichheitsunterstellung (vgl. Heintz, 2010) gewährleistet bleiben konnte.

Die gelieferten *Drittmitteldaten* waren ebenfalls sehr divers und bezogen sich auf Einnahmen, Ausgaben sowie Zusagen. Der Analysedatensatz wurde auf diejenigen Kategorien beschränkt, welche mindestens den Detailgrad der QuaMedFo-Fächerklassifikation aufwiesen: Drittmittelausgaben und Drittmittelzusagen. Die schlechteste Datenlage gab es zu den Drittmittelleinnahmen, welche eigentlich nach KDSF-Standard erhoben werden müssten: diese wurden überhaupt nur von einer der drei Fakultäten vorgehalten und waren nur auf einem Aggregationsniveau oberhalb der QuaMedFo-Klassifikation (s. Tab. 1) lieferbar, sodass sie für Analysen ausscheiden mussten.

Die *Publikationsdaten* werden hier nur nachrichtlich erwähnt, da sie Teil derselben Datenerhebung waren wie die Drittmittel- und Beschäftigtendaten. Die Analyse der Publikationsdaten wird im Beitrag „Publikations- und Rezeptionsanalyse von drei Pilotfakultäten“ (Aman in diesem Band) berichtet.

2.2 Ergänzung der Daten um eine Fachklassifikation

Damit die drei Fakultäten miteinander verglichen werden konnten, musste eine gemeinsame Betrachtungsperspektive entwickelt werden. Von den im Projekt QuaMedFo beteiligten Praktiker*innen wurde die zum Anfragezeitpunkt 2018 vorliegende Fächerklassifikation des Statistischen Bundesamtes (DESTATIS)⁵

⁵ Der Wissenschaftsrat (2013, S. 43) hatte 2013 die Fächerklassifikationen nach DESTATIS noch als im Grunde unzureichend beschrieben. Bis zum Jahr 2016 wurden diese Klassifikationen (auch durch den Austausch zwischen Wissenschaftsrat und Statistischem Bundesamt) stark verändert, sodass sie in der Spezifikation des KDSF-Standards neu bewertet werden und

jedoch nicht als differenziert genug angesehen um die medizinische Forschung zielgerecht abzubilden. Darüber hinaus waren interdisziplinäre Forschung sowie lehrbezogene Sonderformate dort nicht geeignet abbildbar und teilweise waren an den Fakultäten vertretene Subdisziplinen gar nicht der Medizin zugeordnet (z. B. die medizinische Psychologie). Im Zuge der Datenaufbereitung, zunächst bei den Publikationen, wurde deshalb für das Projekt QuaMedFo eine eigene Fächerklassifikation erarbeitet, welche sich an die Klassifikation des medizinischen Fakultätentages (MFT) anlehnt und 11 inhaltliche Fächergruppen umfasst (s. QuaMedFo-Fächerklassifikation in diesem Band). Für die Analyse der Drittmittel- und Beschäftigtendaten war es notwendig, die zusätzlichen beiden Klassen „Interdisziplinäre und translationale Medizin“, welche vor allem Zentren umfasst, sowie „Lehre“, welche bei Beschäftigtenanalysen zum Tragen kommt, zu ergänzen, sodass hier im Unterschied zu anderen Analysen in QuaMedFo 13 Klassen verwendet wurden. Zur Umsetzung dieser Klassifikation in den Analysen wurde durch manuelle Zuordnung eine Konkordanzliste erstellt, welche die Organisationseinheiten der Fakultäten der Klassifikation zuordnet.

2.3 Methoden

Wie in der Einleitung geschildert gehen wir bei der Effizienzmessung grundsätzlich zunächst vom linearen Modell aus. Hierfür sollten Inputvariablen a) untereinander unabhängig sein, da sonst Modelle unnötig kompliziert werden für wenig Informationsgewinn (davon abgesehen destabilisiert starke Multikollinearität die Schätzung der Regressionskoeffizienten). Das bedeutet für unsere Fragestellung, ob Drittmittel als Inputparameter mit betrachtet werden sollten, dass dies nur sinnvoll ist, wenn sie von dem anderen etablierten Parameter – den Beschäftigten-VZÄ – stochastisch unabhängig sind. Zusätzlich sollten b) die Störterme der Variablen im Verhältnis zu den Werten möglichst klein sein. Es ist also die Modellannahme der Homoskedastizität für die als Output vorgeschlagenen Drittmittel zu prüfen.

Da mit den vorliegenden Daten nur eine kleine Fallstudie möglich ist, soll sich hier auf die exemplarische Prüfung dieser genannten beiden Modellannahmen a) *Unabhängigkeit der Regressoren* und b) *Homoskedastizität der Prädiktoren* beschränkt werden. Um die oben genannte Forschungsfrage zu beantworten:

als tauglich befunden werden konnte (Wissenschaftsrat 2016, S. 24). Die Klassifikationen werden seitdem laufend weiter überarbeitet.

Wäre für die Effizienzmessung eine Perspektive auf Drittmittel als Input möglich und sinnvoll? Und welche Drittmitteldarstellung wäre als Output-Parameter am geeignetsten?

Als zu ergänzender Inputparameter (Regressor) wurden von den in Qua-MedFo beteiligten Praktiker*innen die Drittmittelausgaben vorgeschlagen, da sie für die Forschenden eine Ressource darstellen. Die Drittmittelzusagen sollen hingegen als Output (Prädiktor) verstanden werden, da sie wissenschaftliche Erfolge darstellen.

Die Datenanalyse erfolgte mit der statistischen Programmiersprache R (R Core Team, 2022) sowie zusätzlichen Paketen, darunter *tidyverse* (Wickham et al., 2019), *ggplot2* (Wickham, 2016), und *lmtree* (Zeileis & Hothorn, 2002).

3 Ergebnisse der Studie

Die Analyse der Daten hat gezeigt, dass die Unterschiede beim Zusammenhang der verschiedenen erhobenen Variablen zwischen Fächern und Fakultäten sehr groß sind. Dies soll hier nur exemplarisch anhand des Zusammenhangs zwischen Drittmittelausgaben und Drittmittelpersonal illustriert werden. In der grafischen Darstellung (Abb. 1) unterscheiden sich sowohl die Steigungen der Regressionsgeraden zwischen den Fächern als auch die Skalen (Achsenbeschriftungen) zwischen den Fakultäten deutlich. Dies bestätigt sich auch rechnerisch im linearen Modell (Tab. 2), in welchem die Effekte von Fakultäten und Fächergruppen signifikant sind.

3.1 Drittmittel aus Input-Output-Perspektive: Unabhängigkeit

Input-Output-Analysen, welche Personal als Input und Publikationen als Output verstehen, lassen sich leicht als eindimensional verkürzt kritisieren, weshalb auch durch den KDSF-Standard weitere Erfassungsdimensionen vorgeschlagen werden.⁶ In diesem Zusammenhang sind besonders die Drittmittel interessant, da sie zum einen als Drittmittel-Ausgaben die Ressourcen erhöhen, also Input

⁶ Wie z. B. eine Differenzierung nach Kategorien von Drittmittelgeber*innen, oder eine Differenzierung der Vollzeitäquivalente etwa nach Personalkategorien, Geschlecht, Finanzierungsform und/oder Befristung. Im Übrigen werden auch sonst für die Universitätsbewertung häufig zusammengesetzte Indikatoren verwendet (vgl. Benito & Romera, 2011).

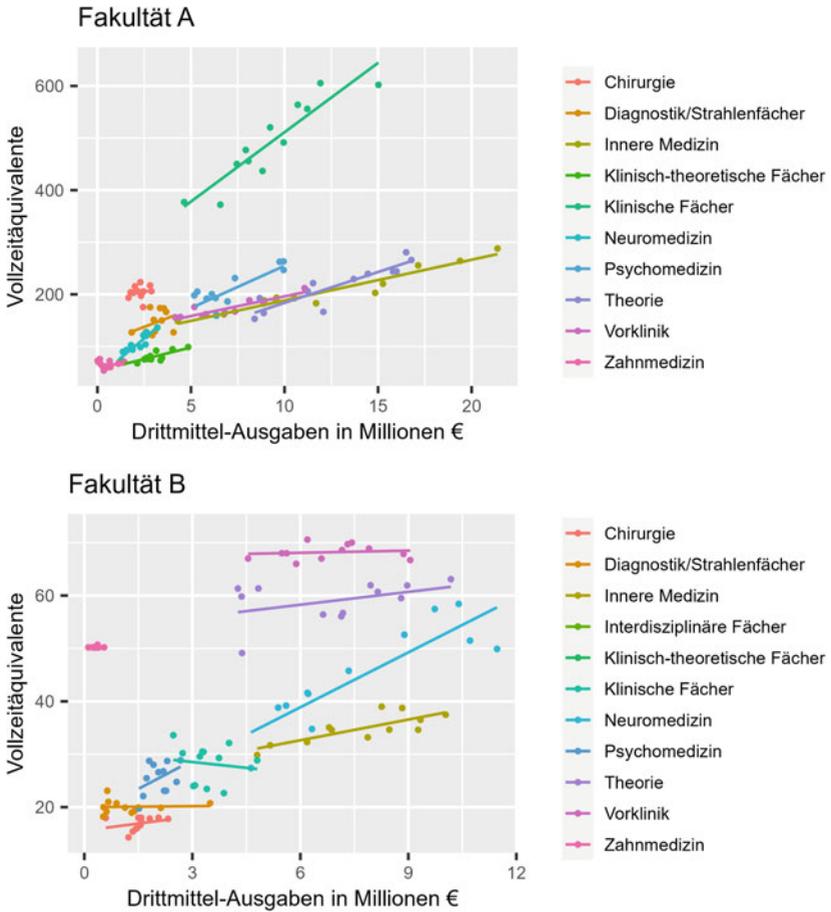


Abb. 1 Drittmittelausgaben und Vollzeitäquivalente: Die Darstellung zeigt a) die Unterschiede in den Forschungsprofilen der Fakultäten als unterschiedliche Mittelausstattung der Fächer; b) die Unterschiede in der Mittelausstattung zwischen den Fächern innerhalb der Fakultäten; c) die Unterschiede im Zusammenhang von Drittmittel-Ausgaben und Vollzeitäquivalenten (Steigungen der Regressionsgeraden)

Tab. 2 Lineare Regression von Drittmittelausgaben auf Drittmittelpersonal mit Kontrollvariablen: signifikante Unterschiede zwischen Fakultäten und Fächern werden hier deutlich

```

Call:
lm(formula = Ausgaben ~ Anzahl_VK + Jahr + Fak + Fächergruppen,
    data = IO_Daten)

Residuals:
    Min       1Q   Median       3Q      Max
-4262042 -1149136  -170796   1006747  7130875

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -514233860   75036482  -6.853  7.51e-11 ***
Anzahl_VK      20767         2028    10.239 < 2e-16 ***
Jahr          254738         37289     6.831  8.50e-11 ***
FakFak B     1843406        413037     4.463  1.30e-05 ***
FächergruppenDiagnostik/Strahlenfächer
  838306         550684     1.522   0.1294
FächergruppenInnere Medizin
  7931229        548215    14.467 < 2e-16 ***
FächergruppenKlinisch-theoretische Fächer
  2381692        719893     3.308   0.0011 **
FächergruppenKlinische Fächer
  1343219        626369     2.144   0.0331 *
FächergruppenNeuromedizin
  3771224        552869     6.821  9.02e-11 ***
FächergruppenPsychomedizin
  2567589        547946     4.686  4.94e-06 ***
FächergruppenTheorie
  7397510        550646    13.434 < 2e-16 ***
FächergruppenVorklinik
  4623591        548079     8.436  4.81e-15 ***
FächergruppenZahnmedizin
  -403860        558162    -0.724   0.4701
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1898000 on 215 degrees of freedom
(296 Beobachtungen als fehlend gelöscht)
Multiple R-squared:  0.8047, Adjusted R-squared:  0.7938
F-statistic: 73.83 on 12 and 215 DF,  p-value: < 2.2e-16

```

sind, und zum anderen als eingeworbene Drittmittel einen Indikator für Forschungsleistung, also Output, darstellen. Um die Leistungsmessung sinnvoll zu ergänzen, müssten Drittmittelausgaben jedoch stochastisch möglichst unabhängig vom Personal sein, sonst würden sie zwar Regressionsmodelle komplizierter machen, jedoch keinen inhaltlichen Beitrag leisten, sondern die Modellgüte (Stabilität der Regressionskoeffizienten) sogar beeinträchtigen (vgl. Backhaus et al., 2018, S. 79 ff.). Zur Prüfung auf Kollinearität wurde mit den gelieferten Daten aus Fakultät A und Fakultät B eine lineare Regression durchgeführt, wodurch sich zeigt (s. Tab. 3), dass von einer Unabhängigkeit von Drittmittelausgaben und drittmittelfinanziertem Personal keinesfalls gesprochen werden kann, sondern dass der Zusammenhang zwischen beiden Zahlen hochsignifikant ist, was auch so bleibt, wenn auf Jahr, Fach und Fakultät kontrolliert wird (s. Tab. 2), weshalb die Einführung von Drittmittelausgaben als zusätzliche Input-Variable eher abzulehnen ist.

Dieses Ergebnis mag trivial erscheinen, da die Erhöhung von Drittmittelpersonal durch verfügbare Drittmittel sofort plausibel ist – hierzu gibt es jedoch zwei Anmerkungen zu machen. Zum ersten wäre natürlich alternativ denkbar, dass

Tab. 3 Regression Drittmittelausgaben auf Drittmittelpersonal ohne Kontrollvariablen: Der Zusammenhang zwischen Drittmittelausgaben und Drittmittelpersonal ist deutlich signifikant

```

Call:
lm(formula = Ausgaben ~ Anzahl_VK, data = IO_Daten)

Residuals:
    Min       1Q   Median       3Q      Max
-5414317 -2673064 -1080259  2411700 13034168

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2844292    337085   8.438 3.88e-15 ***
Anzahl_VK     19105       2044   9.345 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3557000 on 226 degrees of freedom
(296 Beobachtungen als fehlend gelöscht)
Multiple R-squared:  0.2787, Adjusted R-squared:  0.2755
F-statistic: 87.33 on 1 and 226 DF, p-value: < 2.2e-16

```

die Drittmittelausgaben fast ausschließlich für Sachkosten (z. B. Laborausstattung) oder Investitionen (z. B. Forschungsbauten) aufgewendet würden, sodass hier prinzipiell auch nicht-personelle Ressourcen vorliegen könnten, die dann als Input berücksichtigt werden müssten – dies scheint jedoch angesichts der empirischen Untersuchung nicht so zu sein. Zum zweiten wird die Verwendung von Drittmittelausgaben als Input-Parameter ja tatsächlich aus der Fachcommunity vorgeschlagen (Herrmann-Lingen et al., 2014; Herrmann-Lingen in diesem Band), kommt angesichts der empirischen Untersuchung jedoch für ein lineares Modell eher nicht infrage – was in der Gesamtschau bedeutet, dass die Verwendung nichtlinearer Modelle ggf. stärker in Erwägung gezogen werden muss, wenn die Drittmittelausgaben in die Evaluation einbezogen werden sollen.

3.2 Drittmittel aus Input-Output-Perspektive: Varianzen

Außerdem untersucht wurden die Drittmittelzusagen, welche von zwei Fakultäten geliefert wurden. Im Bild ist erkennbar, dass die Varianz der Drittmittelzusagen im Fall der Fakultät B wenig homogen ist (s. Abb. 2), es kann also Heteroskedastizität vermutet werden, weshalb der Goldfeld-Quandt-Test angeraten ist um diese zu prüfen (vgl. Backhaus et al., 2008, S. 86). Im Ergebnis (Tab. 4) muss die Nullhypothese der Homoskedastizität verworfen werden und unsere Vermutung der Heteroskedastizität trifft also zu. Etwas überraschend ist, dass dies trotz der hohen Aggregationsebene auf Fächergruppen noch so deutlich zu Tage tritt.

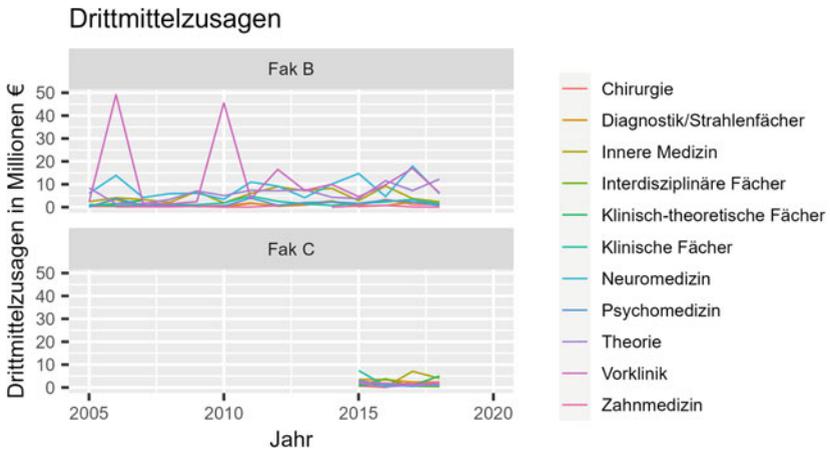


Abb.2 Drittmittelzusagen an den Fakultäten: In Fakultät B weisen die Drittmittelzusagen besonders im Bereich Vorklinik eine deutliche Varianz auf

Tab.4 Lineare Regression von Drittmittel-Zusagen auf Jahre mit darauffolgendem Goldfeld-Quandt-Test zur Prüfung der Nullhypothese von Homoskedastizität: die Nullhypothese muss zurückgewiesen werden – Heteroskedastizität liegt vor

```

Call:
lm(formula = Zusagen_kumuliert ~ Jahr, data = .)

Residuals:
    Min       1Q   Median       3Q      Max
-4171915 -3129790 -1976471  553550 45558442

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -105158644  266696437  -0.394   0.694
Jahr           54197     132582   0.409   0.683

Residual standard error: 6386000 on 143 degrees of freedom
Multiple R-squared:  0.001167,    Adjusted R-squared:  -0.005818
F-statistic: 0.1671 on 1 and 143 DF,  p-value: 0.6833

Goldfeld-Quandt test

data: model
GQ = 4.7461, df1 = 71, df2 = 70, p-value = 2.576e-10
alternative hypothesis: variance increases from segment 1 to 2

```

Im Wesentlichen bedeutet das, dass die Drittmittelzusagen im Verhältnis zu ihrem Mittelwert jährlich zu stark fluktuieren, um aussagekräftig zu sein. In der Tat wäre es auch in der Praxis kaum zu rechtfertigen, wenn seltene Ereignisse wie etwa die Drittmittelzusage für die Einrichtung einer neuen Klinik oder

eines neuen Zentrums dann als Arbeitsergebnis eines bestimmten Jahres verbucht würde. Anstelle dessen wäre es eher als Ertrag einer langjährigen erfolgreichen Forschungstätigkeit zu interpretieren.

Grundsätzlich könnte die analytische Verteilung von zugesagten Fördersummen als Einkommen auf mehrere Jahre diese Varianzverteilung homogenisieren. Allerdings würde das ohnehin bestehende Problem des Zeitverzugs zwischen Leistung (wissenschaftliche Beschäftigung mit dem Gegenstand) und Output (Zusage der beantragten Drittmittel) hierdurch noch verschärft.

Alternativ werden in den Wirtschaftswissenschaften oft rollierende Darstellungen und Planungen verwendet, welche die Mittelwerte über mehrere Jahre überlappend darstellen. Durch diese Darstellung lassen sich Trends unter Umständen besser erkennen – auch hier besteht jedoch wieder das Problem des Zeitverzugs zwischen Leistungserbringung und dem Erhalt aussagekräftiger Evaluationskennzahlen.

3.3 Einschränkungen

Die Betrachtung der drei medizinischen Fakultäten von außen ermöglicht zwar den Vergleich, ist aber notwendig auf die Herausgabe von Daten angewiesen, wodurch Opazität, Aufwand und Verfügbarkeit zu zentralen Herausforderungen werden. Erstens wird die direkte Herausgabe von Daten ohne Anonymisierung oder Aggregation durch Beschäftigtendatenschutz und Betriebsgeheimnisse gehemmt. Zweitens erfordert die aggregierte Übermittlung beim Datenlieferanten hohen Vorverarbeitungsaufwand und intensive Kommunikation zwischen den Datenempfängern und den Datengebern hinsichtlich Verfügbarkeit und Anforderungen. Hier konnte das Projekt auf die große Bereitschaft der Dekanate bauen, welche teilweise auch mehrmals die Daten extrahiert, anonymisiert und aggregiert haben und dabei auch die aus QuaMedFo entwickelte Klassifikation eingesetzt haben. Drittens ist die Verfügbarkeit von Daten, welche normalerweise von den Zuständigen für das Controlling nicht verwendet werden und dadurch nicht zu beschaffen waren, schwer abzuschätzen – empirisch war das bei den Drittmitteleinnahmen der Fall.

Ergänzend kann die Aussagekraft der Daten aus verschiedenen Gründen eingeschränkt sein. Beispielsweise hat eine Fakultät berichtet, dass durch das Integrationsmodell die Forschenden und die in der Patientenversorgung Tätigen administrativ nicht unterschieden werden (können). Bei Drittmitteln kommt zusätzlich eine Besonderheit in der medizinischen Forschung zum Tragen: Bei klinischen Studien, vor allem bei Auftragsforschung, gibt es oft im Vorhinein

kein festes Budget, weshalb hier viele fehlende Werte in den Tabellen existieren, obwohl die Projekte erfasst sind. Außerdem ist die Datenhaltung in den Fakultäten fragmentiert – teilweise werden verschiedene Systeme je nach Mittelherkunft verwendet (z. B. begutachtete Drittmittel, Auftragsforschung etc.), welche dann schwer zu harmonisieren sind.

Zusätzlich bleibt auch die Anwendung der neuen Fächergruppenklassifikation nicht ohne Probleme. Durch Zusammenlegung und Umstrukturierung innerhalb der Fakultäten kann sich die Zuordnung von Lehrstühlen zu Fächergruppen im Zeitverlauf verändern. Dadurch würden Fächer aus Perspektive der Analyse im Verhältnis zueinander größer oder kleiner, obwohl sich inhaltlich möglicherweise nicht viel geändert hat. Bei historisch vorhandenen, aber zum Evaluationszeitpunkt nicht mehr existierenden Instituten ist die retrospektive Zuordnung zudem noch schwieriger – ein Problem, was jedoch bei jeder Klassifikation, welche von außen kommt, auftreten wird.

4 Empfehlungen und Anwendungspotenziale

Steuerung durch Finanzierung greift auf ein wettbewerbliches Gedankenmodell zurück und soll die effiziente Mittelallokation im Hinblick auf optimalen Output gewährleisten. Auf Chancen und Grenzen dieses modellhaften Verständnisses wird in Biesenbender (in diesem Band) näher eingegangen. Dessen ungeachtet benötigt Steuerung geeignete Daten auf der Basis belastbarer und transparenter Konzepte und Definitionen, wie sie zum Beispiel mit dem KDSF-Standard für Forschungsinformationen spezifiziert werden. Im Projekt QuaMedFo wurde die Datenverfügbarkeit und Geeignetheit für drei Pilotfakultäten untersucht, indem für den Zeitraum 2005–2018 verschiedene Input–Output-Daten angefragt und in der Folge analysiert wurden, woraus sich einige Empfehlungen ableiten lassen.

4.1 Inhaltliche Anforderungen an die Datenhaltung in den Fakultäten

Drittmitteldaten kommen in drei Ausprägungen vor: Ausgaben, Zusagen und Einnahmen. Der empirischen Analyse nach muss von der Hinzunahme von Drittmittelausgaben als Input-Indikator jedenfalls für die Effizienzmessung abgeraten werden, da Drittmittelausgaben stochastisch abhängig vom Drittmittelpersonal

sind, wodurch sie keinen Erkenntnisgewinn im linearen Modell bieten.⁷ Ebenfalls mit Problemen behaftet scheinen Drittmittelzusagen als Output-Indikator, da hier eine hohe Fluktuation vorliegt, welche die linearen Modellannahmen verletzt. Und schließlich unterliegen jährlich betrachtete Drittmiteleinnahmen zwar dem Problem des Zeitverzugs zwischen Leistungserbringung und Leistungserträgen, lassen sich jedoch als geglättete Form der Drittmittelzusagen vermutlich rechnerisch am leichtesten verarbeiten. Insofern wäre den Fakultäten anzuraten, dass künftig Drittmiteleinnahmen strukturiert erfasst werden sollten – im Einklang mit den Empfehlungen des KDSF-Standards.

4.2 Berücksichtigung von Fächerunterschieden

Eine für die Steuerung verwendete Fächerklassifikation sollte eng an der Struktur der Fakultät liegen um Steuerungsanliegen adäquat abzubilden. Gleichzeitig ändert sich bei Strukturanpassungen der Fakultäten regelmäßig der Bezugsrahmen von Lehrstühlen und Kliniken, wodurch die zeitkontinuierliche Betrachtung nur schwer gewährleistet werden kann. Eine nachträgliche Klassifikation ist entsprechend hoch fehleranfällig, deshalb sollte die Klassifikation möglichst gleich bei der Datenerfassung erfolgen.

Zusätzlich wird am Paradigma der translationalen Medizin (vgl. Blümel et al., 2015) sowie mit der Gründung von organ- oder syndrombezogenen medizinischen Zentren deutlich, dass die fächerübergreifende Zusammenarbeit funktional im Sinne einer modernen Medizin ist, welche sich entsprechend auch nicht mehr nur fachbezogen evaluieren lässt. Vor diesem Hintergrund gewinnen themen- oder problembezogene Systematiken, die quer zu disziplinären Klassifikationen liegen, stetig an Relevanz.

4.3 Organisationale und technische Anforderungen an die Datenhaltung in den Fakultäten

Der Export aus den verschiedenen Speichersystemen und die Befassung der Mitbestimmungsorgane in den Pilotfakultäten waren zentrale Herausforderungen im Projekt. Auf die frühzeitige Einbindung der Mitbestimmungsorgane

⁷ Drittmittelausgaben hier zu berücksichtigen würde bedeuten, das Drittmittelpersonal mit doppelter Gewichtung in das Modell eingehen zu lassen, was schon auf den ersten Blick ungerechtfertigt erscheint.

bei Datenerfassung und -herausgabe sollte bei Folgeprojekten geachtet werden. Bei der Auswahl von Forschungsinformationssystemen sind Konfigurierbarkeit mit eigenen Klassifikationen, Flexibilität hinsichtlich struktureller Veränderungen sowie die Kompatibilität mit dem KDSF-Standard und interoperabler Datenexport zentral. Auf die Datenhaltung und Formatierung in Excel-Tabellen sollte eher verzichtet werden.

4.4 Hinweise für die Weiterentwicklung des KDSF-Standards

In der Untersuchung insbesondere der Datenverfügbarkeit hat sich gezeigt, dass alle betrachteten medizinischen Fakultäten die Drittmittelausgaben kleinteilig erfassen und für ihr Controlling verwenden, diese Daten also funktional sind. Die Drittmittelausgaben sind gleichwohl nicht im KDSF-Standard enthalten. Da die die Funktionalität von Kennzahlen für die Evaluierten aus soziologischer Perspektive nicht negiert werden sollte, wird hiermit der Einbezug von Drittmittelausgaben bei der Weiterentwicklung des KDSF-Standards vorgeschlagen.

Danksagung Vielen Dank an Prof. Dr. Christoph Herrmann-Lingen für die Entwicklung der Fächergruppenklassifikation und die Erstellung der Konkordanzliste bzw. Klassifikation der Institute nach Fächern und Fächergruppen. Vielen Dank an Dr. Sophie Biesenbender, Alexander Kratzenberg und Christopher Traylor für die Akquise der Daten. Vielen Dank an die Pilotfakultäten für ihre Kooperation und ihren Einsatz, besonders in Zeiten der Corona-Pandemie.

Literatur

- Angelova, D., Göser, M., Wimmer, S., & Sauer, J. (2021). How efficient are German life sciences? Econometric evidence from a latent class stochastic output distance model. *PLoS ONE*, *16*(3), e0247437. <https://doi.org/10.1371/journal.pone.0247437>
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2018). Regressionsanalyse. *Multivariate Analysemethoden* (S. 57–124). Springer Gabler. https://doi.org/10.1007/978-3-662-56655-8_2
- Benito, M., & Romera, R. (2011). Improving quality assessment of composite indicators in university rankings: A case study of French and German universities of excellence. *Scientometrics*, *89*(1), 153–176.

- Blümel, C., Gauch, S., Hendriks, B., Krüger, A. K., & Reinhart, M. (2015). In search of translational research. Report on the Development and Current Understanding of a New Terminology in Medical Research and Practice. *iFQ-BIH-Report*. Berlin institute of health.
- Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT press.
- Clermont, M. (2016). Effectiveness and efficiency of research in Germany over time: An analysis of German business schools between 2001 and 2009. *Scientometrics*, 108, 1347–1381 (2016). <https://doi.org/10.1007/s11192-016-2013-3>
- Dyckhoff, H., & Ahn, H. (2002). Kernaufgaben des Controlling — grundlegende Anmerkungen im Hinblick auf die Sicherstellung der Effektivität und Effizienz. In Weber, J., & Hirsch, B. (Hrsg.), *Controlling als akademische Disziplin*. Schriften des Center for Controlling & Management (CCM), (Bd. 7). Deutscher Universitätsverlag. https://doi.org/10.1007/978-3-322-81430-2_8
- Ebert, B., Tobias, R., & Biesenbender, S. (2021). „Welches System Für Welchen Zweck? Abschluss Und Ausblick.“ *Blog der DINI AGs FIS & EPUB*, 2021. <https://doi.org/10.57689/DINI-BLOG.20210419>
- Espeland, W. N., & Sauder, M. (2016). *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. Russell Sage Foundation. <https://doi.org/10.7758/9781610448567>
- Frietsch, R., Neuhäusler, P., Gruber, S., Zielinski, A., Wortmann, B., Wolny, M., ... & Kolb, M. (2021). Projekt: „Interdisziplinarität von Wissenschaftseinrichtungen – Strukturen und Effekte“ (INTERDIS). <https://doi.org/10.24406/publica-fhg-301254>
- Gralka, S., Wohlrabe, K., & Bornmann, L. (2018). How to Measure Research Efficiency in Higher Education? Research Grants vs. Publication Output. *CESifo Working Paper*, No. 7055. Center for Economic Studies and ifo Institute (CESifo), Munich. <https://www.cesifo.org/en/publications/2018/working-paper/how-measure-research-efficiency-higher-education-research-grants-vs>
- Heintz, B. (2010). Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs/Numerical difference. Toward a sociology of (quantitative) comparisons. *Zeitschrift für Soziologie*, 39(3), 162–181. <https://doi.org/10.1515/zfsoz-2010-0301>
- Herrmann-Lingen, C., Brunner, E., Hildenbrand, S., Loew, T. H., Raupach, T., Spies, C., Treede, R.-D., Vahl, C.-F., & Wenz, H.-J. (2014). Evaluation of medical research performance – position paper of the Association of the Scientific Medical Societies in Germany (AWMF). *GMS German Medical Science; 12:Doc11; ISSN 1612–3174*. <https://doi.org/10.3205/000196>
- Hornbostel, S. (2010). (Forschungs-)Evaluation. In: Simon, D., Knie, A., & Hornbostel, S. (Hrsg.), *Handbuch Wissenschaftspolitik* (S. 293–309). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91993-5_20
- HStatG (2016). Hochschulstatistikgesetz vom 2. November 1990 (BGBl. I S. 2414), das zuletzt durch Artikel 3 des Gesetzes vom Dezember 2016 (BGBl. I S. 2826) geändert worden ist.
- Morgan, M. S. (2020). ‘If p? Then What?’ Thinking within, with, and from cases. *History of the Human Sciences*, 33(3–4), 198–217. <https://doi.org/10.1177/0952695119899349>
- Nullmeier, F. (2001). Input, Output, Outcome, Effektivität und Effizienz. In Blanke, B., von Bandemer, S., Nullmeier, F., & Wewer, G. (Hrsg.), *Handbuch zur Verwaltungsreform*

- (S. 357–363). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-93216-7_40
- Porter, T. M. (1996). *Trust in numbers*. In *Trust in Numbers*. Princeton University Press.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Stiller, J., Trkulja, V., Biesenbender, S., & Petras, V. (2021). *Entwicklung einer Klassifikation für interdisziplinäre Forschungsfelder im Rahmen des Kerndatensatz Forschung Dokumentation des Projekts und der Projektergebnisse*. Berlin, https://www.kerndatensatz-forschung.de/docs_ff/2021-04-30_ffk-projekt_dokumentation.pdf
- Timmermans, S., & Epstein, S. (2010). A world of standards but not a standard world: Toward a sociology of standards and standardization. *Annual review of Sociology*, 36(1), 69–89. <https://doi.org/10.1146/annurev.soc.012809.102629>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., ... & Yutani, H. (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Wiesenthal, H. (2007). Forschungsevaluation als Organisationslernen. In: *Rationalität und Organisation I. Wirtschaft + Gesellschaft* (S. 309–322). Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-20000-8_14
- Wilbers, S., Ringel, L., & Werron, T. (2021). Homöopathen, „Quacksalber“ und wissenschaftliche Mediziner. Zu den Anfängen der Hochschulrankings in der medizinischen Ausbildung der USA, 1850–1930. In: Meier, F., & Peetz, T. (Hrsg.), *Organisation und Bewertung. Organisationssoziologie* (S. 393–424). Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-31549-8_16
- Wildon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Wissenschaftsrat. (2013). *Empfehlungen zu einem Kerndatensatz Forschung* (Drs. 2855–13). https://www.wissenschaftsrat.de/download/archiv/2855-13.pdf?__blob=publicationFile&v=1
- Wissenschaftsrat. (2016). *Empfehlungen zur Spezifikation des Kerndatensatz Forschung* (Drs. 5066–16). https://www.wissenschaftsrat.de/download/archiv/5066-16.pdf?__blob=publicationFile&v=1
- Wissenschaftsrat.(2020). *Stellungnahme zur Einführung des Kerndatensatz Forschung* (Drs. 8652–20). https://www.wissenschaftsrat.de/download/2020/8652-20.pdf?__blob=publicationFile&v=5
- Zeileis, A., & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News* 2(3), 7–10. <https://CRAN.R-project.org/doc/Rnews/>

Hartstein, Judith Judith Hartstein forscht seit 2016 am DZHW zu Bewertungspraktiken in der Wissenschaft und zu digitalen Infrastrukturen. Seit 2019 promoviert sie an der

Humboldt-Universität zu Berlin im Fach Soziologie. Zuvor hat sie einen M.A. in Wissenschaftsforschung an der Humboldt-Universität erworben sowie einen B.Sc. in Wirtschaftsmathematik an der Hochschule für Technik und Wirtschaft Berlin.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Qualitätsmaße zur Evaluierung medizinischer Forschung – Erkenntnisse und Empfehlungen aus dem Projekt QuaMedFo

Isabella Peters und Sophie Biesenbender

1 Einleitung

Das Verbundprojekt QuaMedFo widmete sich insbesondere der Erforschung der praktischen Verwendung und Eignung von verschiedenen, in Evaluationskontexten der deutschen medizinischen Fakultäten und Universitätskliniken verwendeten Indikatoren. Dabei wählte das Projekt einen forschungsinformation-basierten Ansatz, der verschiedene Daten bzw. Datenzugänge in den Blick nahm und ihre methodisch-theoretische wie auch praktische Nutzbarkeit für verschiedene Anwendungskontexte analysierte. Zentrale Forschungsfragen betrafen dabei die Herausforderungen bei der Evaluation von medizinischer Forschung auf institutioneller und personeller Ebene, die sich durch

I. Peters (✉)

ZBW – Leibniz-Informationszentrum Wirtschaft, Christian-Albrechts-Universität zu Kiel,
Kiel, Deutschland

E-Mail: ipe@informatik.uni-kiel.de

S. Biesenbender

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Berlin,
Deutschland

E-Mail: biesenbender@dzhw.eu

I. Peters

Christian-Albrechts-Universität zu Kiel, Kiel, Deutschland

© Der/die Autor(en) 2024

S. Biesenbender und J. Hartstein (Hrsg.), *Qualitätsmessung als Prisma*, Higher
Education Research and Science Studies,

https://doi.org/10.1007/978-3-658-43683-4_8

- die verbreitete Nutzung von publikations- und zitationsbasierten Indikatoren für bestimmte Output-Typen in der Leistungsorientierten Mittelvergabe (LOM),
- die dadurch entstandene Nichtberücksichtigung von Leistungs- und Forschungsdimensionen, die nicht in publikations- und zitationsbasierten Indikatoren abgebildet werden (können), und
- die enggeführte Definition von Impact, d. h. wissenschaftliche Resonanz reflektiert im Journal Impact Factor,

ergeben haben. Komplizierend kommt hinzu, dass die deutsche Evaluationskultur im Bereich der medizinischen Forschung durch diverse unterschiedliche Richtlinien und Empfehlungen (z. B. Deutsche Forschungsgemeinschaft, 2004; Herrmann-Lingen et al., 2014; Herrmann-Lingen in diesem Band) geprägt ist, was eine Fragmentierung der Evaluierungspraktiken zur Folge hat. Außerdem haben sich dadurch institutionensspezifische Umsetzungspraktiken etabliert, die nur wenig einrichtungsübergreifenden Standards folgen.

Um einen konstruktiven Beitrag zu einer erweiterten Betrachtung und Einbeziehung von Leistungsdimensionen zu leisten, wurden aus einer output-basierten Perspektive neben der vergleichenden Betrachtung vorhandener Datenzugänge und Indikatoren in QuaMedFo (Aman; Hartstein; Lemke et al.; Traylor und Aman sowie Lippert und Förstner in diesem Band) auch neue Indikatoren zur Abbildung der praktischen Resonanz und Reichweite von Forschung und Forschungsergebnissen entwickelt.

QuaMedFo arbeitete unter der Annahme, dass eine objektive und einfache Messung von Leistung nicht möglich ist, weil einerseits die Indikatoren an sehr spezifische Kontexte und Messkonzepte gebunden sind und andererseits die Bewertungsdimension nicht ohne die Formulierung strategischer Ziele (abhängig vom Verwendungskontext, institutionellen Profil, aber auch individuellen Strategien) denkbar ist. Daher war das Ziel des QuaMedFo-Projekts nicht ein allgemeingültiges Evaluierungsverfahren oder System zur Mittelvergabe zu entwickeln, sondern eine Reihe unterschiedlicher Indikatoren, die die unterschiedlichen Leistungsdimensionen berühren, zu analysieren und zu bewerten. Dazu wurde eng mit drei deutschen medizinischen Fakultäten zusammengearbeitet, die als sogenannte „Pilotfakultäten“ Zugriff auf u. a. ihre Publikationsdaten erlaubten und die Befunde des QuaMedFo-Projekts regelmäßig mit den Projektmitarbeitenden diskutierten.

Dieser Beitrag fasst die zentralen Ergebnisse von QuaMedFo zusammen, beschreibt die Herausforderungen für die Evaluation medizinischer Forschung

und formuliert Erkenntnisse und Empfehlungen für die zukünftige Evaluierungspraxis. Dabei spiegeln die folgenden Einordnungen und Interpretationen der Ergebnisse des Vorhabens QuaMedFo und seiner Teilprojekte sowie die darauf basierenden Empfehlungen die Einschätzungen der Autorinnen wider und umfassen keine systematische Bewertung von die Umsetzung betreffenden Fragen von Machbarkeit, Praktikabilität oder Akzeptanz im Allgemeinen.

2 Zusammenfassung der Projektergebnisse

Entlang der in Biesenbender (in diesem Band) diskutierten Bewertungs- und Leistungsdimensionen sollen die Projektergebnisse im Folgenden kurz zusammengefasst werden.

2.1 Wissenschaftliche Reichweite und wissenschaftliche Resonanz

Die wissenschaftliche Reichweite medizinischer Forschungsergebnisse lässt sich u. a. anhand des Journal Impact Factors der Fachzeitschriften (JIF) schätzen, in denen die Ergebnisse in Form von wissenschaftlichen Artikeln veröffentlicht wurden. Die Nützlichkeit, Relevanz, Qualität oder der Einfluss wissenschaftlicher Publikationen lässt sich außerdem mit der Häufigkeit beschreiben, mit der sie in anderen wissenschaftlichen Publikationen zitiert werden bzw. dort Resonanz erfahren (Wolbring, 2015; Deutsche Forschungsgemeinschaft | AG Publikationswesen, 2022).

Medizinische Fakultäten nutzen für die Leistungsorientierte Mittelvergabe (LOM) und den Vergleich der Fächer innerhalb einer Institution am häufigsten einen unkorrigierten JIF für einzelne Artikel (Medizinischer Fakultätentag, 2022; Herrmann-Lingen in diesem Band). Für Artikel, die in Zeitschriften ohne JIF erschienen sind, werden manchmal Ersatzpunkte vergeben (gemäß den Empfehlungen der Deutschen Forschungsgemeinschaft von 2004; Deutsche Forschungsgemeinschaft, 2004). Die in QuaMedFo untersuchten Daten der Pilotfakultäten zeigen, dass sich die JIFs der Zeitschriften, in denen der wissenschaftliche Output der Pilotfakultäten erschienen ist, positiv entwickeln. Zudem werden zunehmend mehr Zeitschriftenaufsätze publiziert, wobei der Anteil an deutschsprachigen Publikationen zurückgeht (Aman in diesem Band). Ob dies den fakultätsinternen Anreizsystemen der Fakultäten oder den natürlichen Schwankungen der JIFs geschuldet ist, lässt sich mittels bibliometrischer Verfahren nicht feststellen. Da

aber bspw. der Rückgang deutschsprachiger Publikationen geringer an solchen Institutionen ausfällt, die den JIF bei der LOM betrachten (Aman in diesem Band), liegt der Schluss nahe, dass die institutsinterne Steuerung des Publikationsverhaltens über die LOM eine Wirkung zeigt (Aman & van den Besselaar, 2023).

Zitationsraten und JIFs sind disziplin- und fachbereichsspezifisch und spiegeln die jeweilige Zitations- und Publikationskultur und -praxis unmittelbar wider (Aman in diesem Band). Die vorklinischen Fächer zeigen z. B. den größten Anteil an zitierten Publikationen insgesamt. Publikationen aus dem Bereich der Grundlagenfächer werden in Journalen mit höherem JIF veröffentlicht als Publikationen aus dem Bereich der angewandten Fächer (und dies unter Umständen nicht aus mangelnder Qualität, sondern weil es keine geeigneten wissenschaftlichen Zeitschriften mit höheren JIFs gibt).

LOM sowie intra- und inter-institutionelle Vergleiche allein auf Basis von unkorrigierten JIFs benachteiligen insbesondere Forschende aus der angewandten und klinischen Forschung (s. QuaMedFo-Fächerklassifikation in diesem Band). Eine Alternative stellen feld- oder journal-normalisierte Indikatoren dar (Aman in diesem Band).

2.2 Praktische Resonanz

Analysen von Zitationsdynamiken und der Rezeption von wissenschaftlichen (Journal-)Publikationen in medizinischen Leitlinien geben beispielhaft Aufschluss darüber, wie wissenschaftliche Ergebnisse Eingang in die medizinische Praxis finden. Sie bestätigen grundsätzlich den in britischen Studien beobachteten Zusammenhang von dem JIF des veröffentlichenden Journals einer wissenschaftlichen Publikation und der Wahrscheinlichkeit ihrer Zitierung in medizinischen Empfehlungen und Leitlinien (Lewison, 2003). Dieser Zusammenhang nimmt jedoch ab, sobald themenübergreifende Journale mit einem sehr hohen JIF aus der Betrachtung ausgeschlossen werden. Die Ergebnisse in QuaMedFo zeigen, dass ein großer Teil leitlinien- und anwendungsrelevanter Ergebnisse medizinischer Forschung in fachspezifischen Journalen mit relativ geringen JIFs veröffentlicht wird. Auf diesen Befunden basierende vertiefende fallstudienbasierte Analysen zu den Eigenschaften der Referenzen einzelner Leitlinien über Vergleiche von Artikeln einer bestimmten Zeitschrift eines bestimmten Jahrgangs zeigen, dass Artikel, die in Leitlinien referenziert werden, höhere mittlere Zitationszahlen aufweisen als vergleichbare nicht in Leitlinien referenzierte Artikel in dieser Zeitschrift (Traylor und Aman in diesem Band). Der Befund, dass jedoch ein großer

Anteil leitlinien- und anwendungsrelevanter Ergebnisse medizinischer Forschung in fachspezifischen Journalen mit relativ geringen JIFs veröffentlicht wird, bildet die Grundlage für die Entwicklung des Leitlinien-Impact-Faktors als ein zusätzliches Qualitätsmaß für Journals, welches eine andere Leistungsdimension als der JIF beschreibt, im Rahmen des Nachfolgeprojekts QuaMedFo-Transfer.¹

2.3 Gesellschaftliche Resonanz

Da Social-Media-Plattformen und andere Medienformate nicht nur Forschende adressieren, gelten sie als Möglichkeit, die Beschäftigung mit Forschungsergebnissen durch die breitere Öffentlichkeit abzubilden (als sogenannte Altmetrics). Für die drei an QuaMedFo beteiligten Pilotfakultäten hat sich gezeigt, dass nur sehr wenige Publikationen auf den verschiedenen Social-Media-Plattformen überhaupt erwähnt werden und, falls mehrfach erwähnt, dann insbesondere auf Twitter (Lemke et al. in diesem Band). Medizinische Forschungsergebnisse von deutschsprachigen Institutionen sind zudem nicht gleichmäßig über die verschiedenen Plattformen verteilt zu finden, sondern es herrschen eher spezialisierte Themenbereiche pro Plattform vor (Fraser et al., 2021). Interessant scheint die Analyse von sogenannten Policy-Dokumenten (z. B. Empfehlungen der Weltgesundheitsorganisation – WHO) zu sein, die insbesondere Ergebnisse klinischer Forschung häufig zitieren. Auf diese Weise lässt sich u. a. darstellen, wie medizinisches Wissen in die Gesellschaft wirkt bzw. welchen praktischen Einfluss dieses hat. Allerdings werden auch bei diesem Indikator Verzerrungen im Hinblick auf bestimmte Themen (z. B. Onkologie) oder im Hinblick auf die Anzahl multinationalaler Ko-Autor:innen sichtbar (Lemke et al. in diesem Band).

Ähnlich wie bei der Betrachtung von wissenschaftlicher Resonanz und Reichweite sind intra- und interinstitutionelle sowie interfachliche Vergleiche nur auf höheren Aggregationsebenen (z. B. große Publikationsmengen, lange Betrachtungszeiträume) sinnvoll möglich. Die Verwendung von Altmetrics ist nur als komplementär zu anderen quantitativen und qualitativen Formen der Wissenschaftsevaluation angeraten.

¹ https://www.dzhw.eu/forschung/projekt?pr_id=724.

2.4 Wirtschaftliche Verwertung von Forschung durch Patente

Ein Wissenstransfer von der Theorie in die Praxis lässt sich unter anderem bei der Patentierung von Forschungsergebnissen vermuten. Um diesen Transfer analytisch nachzuvollziehen, ist die Zuordnung von Patenten zu wissenschaftlichen Publikationen erforderlich, was bedeutet, dass ein Zusammenhang zwischen Daten aus verschiedenen Datenbanken und damit verbundenen Logiken hergestellt werden muss. An der ZB MED wurde ein Verfahren entwickelt, Patent-Publikation-Paare spezifisch für medizinische Forschung automatisiert zu finden (Lippert und Förstner in diesem Band). Hierbei kommen die Medical-Subject-Headings (Indexterme einer medizinischen Fachsystematik abgekürzt als MeSH) sowie mustererkennende Verfahren zur Anwendung. Nach der Zuordnung von Patenten zu Publikationen können diese (mit Einschränkungen) für Analysen zur wirtschaftlichen Verwertbarkeit der Forschung verwendet werden.

Obwohl daraufhin quantitative Aussagen über einen Wissenstransfer oder ein interinstitutioneller Vergleich nur eingeschränkt möglich sind, kann das Wissen über den Informationsfluss von Forschung zu Praxis hilfreich für die Situierung und fachlich-praktische Kontextualisierung von Instituten medizinischer Forschung sein.

3 Herausforderungen für die Evaluation medizinischer Forschung

Für die Evaluation medizinischer Forschung wurden im Rahmen von QuaMedFo, neben methodischen und theoretischen Anforderungen, insbesondere Herausforderungen im Hinblick auf Datenverfügbarkeit und Machbarkeit identifiziert.

Die größte Problematik ergibt sich aus der mangelnden Verfügbarkeit und der Unstrukturiertheit von Daten, die für eine verantwortungsvolle Anwendung szientometrischer Verfahren notwendig sind. In den analysierten Fakultäten gibt es keine einheitliche, vollständige Datenvorhaltung, die alle Bewertungsdimensionen umfänglich berücksichtigt (Biesenbender in diesem Band). Während Publikations- und Personaldaten vorhanden sind und zum großen Teil den KDSF-Standard (*Kerndatensatz Forschung*; KFiD – Kommission für Forschungsinformationen in Deutschland, 2022) abbilden, folgen die Informationen zu Drittmitteln einer anderen Logik: Anstelle von Drittmittelleinnahmen werden in der Regel Drittmittelausgaben zugrunde gelegt (Hartstein in diesem Band) und die komplexen

Selbstverstärkungskreise durch Drittmittel als Input- und Output-Parameter nicht abgebildet (Herrmann-Lingen in diesem Band).

Auch aufseiten der alternativen Plattformen, die Aufschluss über die gesellschaftliche Resonanz von medizinischer Forschung geben können, ist man mit einem Problem des Mangels konfrontiert (Fraser et al., 2021, Lemke et al. in diesem Band): Der größte Teil der Publikationen weist einen nicht messbaren altmetrischen Impact auf, d. h. medizinische Publikationen werden nicht auf Social-Media-Plattformen erwähnt. Eine Ausnahme stellt Twitter (seit 2023: X) dar, wo sich Nutzende mit zahlreichen medizinischen Publikationen auch mehrfach beschäftigen.

Die Durchführung umfassender und valider bibliometrischer Analysen oder Evaluationen der wissenschaftlichen als auch gesellschaftlichen Leistungsdimensionen erfordert in der Regel die Zusammenführung von institutionell (d. h. durch Einrichtungen oder Fakultäten) verarbeiteten Daten mit bibliographischen Informationen, die durch Drittanbieter auf unterschiedlichen Plattformen bereitgestellt werden. Dies birgt Raum für Fehler, erfordert ein Mindestmaß an Kenntnissen der Datenverarbeitung und sorgt für einen hohen Aufwand bei der Datenbeschaffung.

Bei der bibliometrischen Analyse ergeben sich auf methodischer Ebene vor allem Herausforderungen hinsichtlich der Nutzung von Indikatoren, die z. B. nicht vergleichend auf die verschiedenen Fächer angewendet werden können, wenn ein faires Ergebnis entstehen soll. Maßgeblich für faire Vergleiche ist die Kenntnis über u. a. das fachbereichsspezifische Publikations- und Zitationsverhalten. Das Projekt QuaMedFo konnte Aufschluss darüber geben, dass die bisherige JIF-basierte LOM-Praxis Nachteile für Fächer bringt, die im Web of Science unterrepräsentiert sind (z. B. angewandte klinische Forschung), in Monographien publizieren, eher interdisziplinär forschen oder in deutscher Sprache veröffentlichen (Aman in diesem Band).

Problematisch ist, dass für interinstitutionelle Vergleiche oder für die Normalisierung der Indikatoren auf Fächerebene bislang keine geeignete einheitliche, deutschlandweite oder internationale Fächerklassifikation vorliegt, sodass in QuaMedFo eine solche erarbeitet wurde, um eine gemeinsame Betrachtungsperspektive für die drei zu vergleichenden Pilotfakultäten zu schaffen (s. QuaMedFo-Fächerklassifikation in diesem Band). Für eine flächendeckende Zuordnung der Daten, sollte die Klassifikation gleich bei der Erfassung der Daten in den Institutionen genutzt werden. Beachtet werden muss allerdings, dass derartige Klassifikationen immer auch individuelle Steuerungsanliegen der Fakultäten und Einrichtungen widerspiegeln sollten (Hartstein in diesem Band) und eine (inter-)nationale Vergleichbarkeit nicht immer gewünscht oder notwendig ist.

Im Hinblick auf Altmetrics ergab sich ebenfalls ein interessantes Bild: Hier zeigte sich, dass vor allem anwendungsnahe Fächer auf Twitter oder auch in Policy-Dokumenten zahlreich und häufig Erwähnung finden (im Gegensatz zu grundlagenorientierten Fächern, die besser in den Indikatoren der wissenschaftlichen Resonanz und Reichweite abschneiden). Allerdings fokussiert sich das Interesse der alternativen Plattformen auf bestimmte Themen, sodass die weite Anwendbarkeit von Altmetrics hier auch eingeschränkt ist.

In Deutschland beschränkt sich die LOM in der medizinischen Forschung nicht nur auf die Bewertung von Publikationen, sondern sie bezieht auch die Drittmittelstärke der Leistungseinheiten einer medizinischen Fakultät ein. In Qua-MedFo konnte gezeigt werden, dass ein starker statistischer Zusammenhang zwischen den Drittmittelausgaben und dem Personalaufkommen besteht und dass Drittmittelausgaben daher als zusätzliche Input-Variable (da sie als Ausgaben die Ressourcen des Instituts erhöhen) für die Effizienzmessung medizinischer Forschung im linearen Modell ungeeignet, da redundant sind. Ähnlich ungeeignet sind Drittmittelzusagen als zusätzliche Output-Variable, welche das Ergebnis der Forschungsleistung darstellen könnten, leider aber durch starke Schwankungen ebenfalls die linearen Modellannahmen verletzen (Hartstein in diesem Band).

Auf theoretischer Ebene ergeben sich Herausforderungen durch den Einbezug von neuartigen Quellen, die z. B. die gesellschaftliche Resonanz medizinischer Forschung spiegeln sollen wie Policy-Dokumente (Lemke et al. in diesem Band). Im Gegensatz zu Zitationen in wissenschaftlichen Publikationen fehlen in der bibliometrischen Forschung allerdings noch systematische Erkenntnisse bspw. über die Anlässe und Motivationsstrukturen hinter der Referenzierung von Publikationen in Policy-Dokumenten. Es fehlt die theoretische Auseinandersetzung mit dem Wert einer Zitation einer wissenschaftlichen Publikation in einem Policy-Dokument und damit bleibt auch offen, wie das Vorkommen oder die Häufigkeit des Vorkommens interpretiert und sinnvoll genutzt werden kann.

4 Erkenntnisse und Empfehlungen des QuaMedFo-Projekts

Berücksichtigt man Fragen der Machbarkeit und bezieht man die Ausgangssituation ein, wie sie durch bestehende Systeme, Verfahren, Datenzugänge sowie methodisch-fachliche Kompetenzen und Ressourcen in Organisationen (in diesem Fall medizinischen Fakultäten) gegeben ist, legt die Zusammenarbeit mit den Pilotfakultäten den Schluss nahe, dass Anpassungen und Verbesserungen der Bewertung medizinischer Forschungsleistung nur schrittweise umsetzbar sind

(Aman sowie Hartstein in diesem Band). Aus diesem Grund beziehen sich die in QuaMedFo herausgearbeiteten Empfehlungen und entwickelten Vorschläge im Wesentlichen auf die Erweiterung etablierter Verfahren insbesondere der institutionellen LOM mit ihrem Fokus auf Drittmittel und Publikationen (bestimmten Typs) über die Ergänzung von Datenzugängen und Indikatoren für bestimmte Anwendungsfälle (Lemke et al.; Lippert und Förstner sowie Traylor und Aman in diesem Band).

Aus den Befunden der Teilprojekte von QuaMedFo über den momentanen Status der Evaluation medizinischer Forschung an deutschen Institutionen lassen sich daher folgende Erkenntnisse und Empfehlungen für die zukünftige Praxis ableiten:

- **Eine verantwortungsvolle und faire Forschungsbewertung erfordert eine multidimensionale Betrachtung von Forschungsleistung mit angepassten Indikatoren.** Die Analysen in QuaMedFo haben gezeigt, dass die zurzeit an deutschen medizinischen Forschungsinstitutionen durchgeführten Evaluierungspraktiken und die Nutzung von insbesondere publikationsbasierten und zitationsbetonten Indikatoren in der LOM (häufig auf Basis der Anzahl der Publikationen und eines unkorrigierten JIFs) zu einer unfair vergleichenden Bewertung der medizinischen Fachbereiche führt (Medizinischer Fakultätentag, 2022; Wolbring, 2015; Herrmann-Lingen et al., 2014). Speziell die anwendungsorientierte Forschung wird dadurch benachteiligt. Eine verantwortungsvolle Bewertungspraxis erfordert eine multidimensionale Betrachtung von Forschungsleistung mit an die medizinische Forschung angepassten Datenerhebungsinstrumenten, Datensätzen und Indikatoren (Herrmann-Lingen in diesem Band).
- **Die Ziele, Instrumente und Objekte der Evaluierung müssen zusammenpassen.** Evaluierungsaufgaben erfordern immer einen Abgleich der Evaluierungsziele (z. B. hohe wissenschaftliche Reichweite), der Evaluierungsinstrumente (z. B. JIF) und der Evaluierungsobjekte (z. B. wissenschaftliche Publikation). Nicht jedes Evaluierungsziel lässt sich mit jedem Instrument oder mit gerade vorhandenen Daten überprüfen. Optimalerweise beginnen Evaluierungsaufgaben mit der Definition eines Evaluierungsziels (Medizinischer Fakultätentag, 2022; Ringelhan et al., 2015). Danach folgt die Auswahl des oder der Instrumente und zuletzt die Erhebung der Daten. Die Qualität des Evaluationsergebnisses (z. B. ob die Steuerung gelungen ist) steht und fällt mit der Passung der drei Bestandteile (Ist das Instrument das richtige für das Ziel bzw. die Frage? Sind die Objekte passend für die Beantwortung der Frage?) sowie mit der Qualität des Instruments und der Daten über die Objekte.

- **Die Qualität der Daten ist für Evaluierungszwecke essentiell.** Datenqualität umfasst vor allem die Aspekte Vollständigkeit, Aktualität und Korrektheit. In dieser Perspektive hängen die Qualität von Indikatoren und ihre Nutzbarkeit wesentlich auch von der Qualität der Prozesse und Systeme der institutionellen Datenverarbeitung ab (Biesenbender, 2019). Dabei kommt Standards (wie z. B. Klassifikationen) eine besondere Bedeutung zu: sowohl für die Informationsverarbeitung (wie über den KDSF-Standard) als auch für die Entwicklung und Anwendung von Bewertungsinstrumenten im Sinne der Responsible-Metrics-Initiative (Wilson et al., 2015). Datenbanken und Standards müssen jedoch so flexibel sein, dass sie auf die disziplinären und evaluierungsbezogenen Anforderungen der Institute eingehen können.
- **Eine Ausweitung der Forschungsbewertung auf weitere Leistungsdimensionen und Datenquellen ermöglicht eine differenziertere Betrachtung von medizinischer Forschung.** Momentan konzentriert sich die Bewertung medizinischer Forschungsleistung auf nur sehr wenige Leistungsdimensionen, insbesondere Publikationen in Fachzeitschriften gelistet in PubMed, Scopus oder Web of Science und mit JIF – seltener auf Monographien und deutschsprachige Artikel in lokalen Zeitschriften ohne JIF (dafür mit JIF-Ersatzpunkten; Deutsche Forschungsgemeinschaft, 2004). Wenige Institutionen betrachten weitere Publikationstypen (z. B. News Items) oder andere Datenquellen, die wissenschaftliche, gesellschaftliche oder wirtschaftliche Reichweite, Resonanz oder Transfer (z. B. Preprint-Repositories, Altmetrics, Patente) abbilden können. Eine Ausweitung der zu berücksichtigenden Leistungsdimensionen (z. B. Organisation von Konferenzen, Veröffentlichung von Forschungssoftware; Hermann-Lingen sowie Biesenbender in diesem Band) sowie der Datenquellen (z. B. Policy-Dokumente, Twitter bzw. X seit 2023) zu weiteren Formen der Resonanz und Reichweite, ist im Hinblick auf den Zugewinn an Informationen zu medizinischer Forschungsarbeit und ihrer Wahrnehmung zu empfehlen. Aktuell wird ein Ansatz zur Abbildung der Leitlinienrelevanz medizinischer Forschung sowie zur Berücksichtigung von Leitlinienengagement medizinischer Forscher:innen in Erfassungs- und Bewertungssystemen erarbeitet und in Hinblick auf Fragen der Machbarkeit, Praktikabilität, Datenzugänglichkeit und Akzeptanz teilweise in bestehende Bewertungssysteme an Pilotfakultäten implementiert (Traylor und Aman in diesem Band). Über dieses Vorhaben lassen sich perspektivisch medizinische Leitlinien und darin enthaltene Referenzen als neue output-basierte Grundlage für Qualitätsmaße erschließen (Biesenbender in diesem Band). Weitere Outputs sind in diesem Zusammenhang denkbar und für die diesbezügliche Erschließung empfohlen. Hier liegt außerdem das Potenzial, die durch die

momentane LOM-Praxis entstehende Überbetonung von grundlagenorientierter Forschung zu dämpfen – auch wenn durch den Einbezug weiterer Quellen und Publikationstypen sehr wahrscheinlich andere Betonungen entstehen. An dieser Stelle müssen die Evaluierungsziele im Blick behalten werden.

- **Die Steuerungsfunktion der Forschungsbewertung sowie das Potenzial von Forschungsinformation strategisch nutzen.** Evaluierungen greifen immer auch steuernd in Prozesse ein (Hartstein in diesem Band). Die Institutionen sollten ein stärkeres Bewusstsein dafür entwickeln, dass sich natürliches Verhalten, z. B. im Hinblick auf Publikationen und Drittmiteleinwerbung, nicht mehr beobachten lässt, wenn Evaluierungsaufgaben oder Steuerungsziele bekannt sind. Dies bedeutet nicht, dass Institutionen ihre strategischen Planungen verheimlichen sollten – im Gegenteil: Im Einklang mit bekannten Forderungen und Initiativen (z. B. DORA,² Leiden Manifesto,³ COARA⁴) sollten Transparenz über die Evaluierungsaufgaben hergestellt und die LOM-Systeme entsprechend angepasst werden. Die Mittelverteilung ist dabei nicht die einzige Aufgabe, die über entsprechende Daten und Instrumente erledigt werden kann. Möglicherweise ist es für manche Institutionen sinnvoll, den Blick von der LOM auf den Nutzen der verschiedenen Informationen für die eher qualitativ ausgerichtete Betrachtung der Einrichtung, im Sinne eines Monitorings der größeren strategischen Ziele, zu lenken. Daher regt QuaMedFo an, das Anwendungsspektrum von Forschungsinformationen jenseits ihrer Einsatzmöglichkeiten im Rahmen von Evaluationen von Leistungseinheiten (LOM-Systeme) oder Personen (etwa in Berufungsverfahren) zu erschließen. Dies umfasst bspw. evidenzbasierte strategische Orientierungen oder eine kritische Einschätzung der Leistungsfähigkeit. Standardisierte, über einen längeren Zeitraum vorgehaltene und verarbeitete Informationen zu Forschung in Institutionen bzw. Organisationseinheiten ermöglichen vor diesem Hintergrund eine Auskunftsfähigkeit über die eigenen Leistungsprozesse und -ergebnisse für die Darstellung nach außen, aber insbesondere auch für eine Erhöhung der Steuerungsfähigkeit (Wissenschaftsrat, 2020).
- **Die Forschungsbewertung und Nutzung von Forschungsinformation ist immer von individuellen Zielen geprägt.** Die Nutzung von Forschungsinformation sowie Evaluierungsaufgaben und ihre Passung mit Instrumenten und Objekten sind zum großen Teil individuell und müssen sich nach den Gegebenheiten der Institution, ihren Zielen, ihrer Spezialisierung etc. richten. Folgt

² <https://sfedora.org/>

³ <http://www.leidenmanifesto.org/>

⁴ <https://coara.eu/>

man diesem Gedanken konsequent, kann es keine deutschlandweite oder gar internationale Blaupause geben – was zwangsweise zu Spannung im interinstitutionellen Wettbewerb führen würde. Hier ist die fachwissenschaftliche Community gefragt zu diskutieren, in welcher Hinsicht eine interinstitutionelle und ggf. interfachliche Vergleichbarkeit hergestellt werden soll und mit welchen Mitteln.

Literatur

- Aman, V., & van den Besselaar, P. (2023). Effects of performance-based funding systems on publication behavior – A case study of German medical faculties. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4388750>
- Biesenbender, S. (2019). The governance and standardisation of research information in different science systems: A comparative analysis of Germany and Italy. *Higher Education Quarterly*, 73(1), 116–127. <https://doi.org/10.1111/hequ.12193>
- Deutsche Forschungsgemeinschaft. (2004). *Empfehlungen zu einer Leistungsorientierten Mittelvergabe (LOM) an den Medizinischen Fakultäten. Stellungnahme der Senatskommission für Klinische Forschung der Deutschen Forschungsgemeinschaft* (S. 1–23). https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/stellungnahmen_papiere/2004/stellungnahme_klinische_forschung_04.pdf
- Deutsche Forschungsgemeinschaft | AG Publikationswesen. (2022). *Wissenschaftliches Publizieren als Grundlage und Gestaltungsfeld der Wissenschaftsbewertung*. <https://doi.org/10.5281/ZENODO.6538163>
- Fraser, N., Bräuer, P., & Peters, I. (2021). Altmetrics for evaluation of medical research in Germany. *Proceedings of 18th International Conference on Scientometrics & Informetrics*, 1471–1472. <http://hdl.handle.net/11108/504>
- Herrmann-Lingen, C., Brunner, E., Hildenbrand, S., Loew, T. H., Raupach, T., Spies, C., Treede, R.-D., Vahl, C.-F., & Wenz, H.-J. (2014). Evaluation of medical research performance – Position paper of the Association of the Scientific Medical Societies in Germany (AWMF). *GMS German Medical Science; 12:Doc11; ISSN 1612–3174*. <https://doi.org/10.3205/000196>
- KFiD – Kommission für Forschungsinformationen in Deutschland. (2022). *Dokumentation der Spezifikation des KDSF – Standard für Forschungsinformationen in Deutschland: Version 1.3*, Berlin. <https://doi.org/10.58010/kdsf:Spez:1.3:2022>
- Lewison, G. (2003). Beyond outputs: New measures of biomedical research impact. *Aslib Proceedings*, 55(1/2), 32–42. <https://doi.org/10.1108/00012530310462698>.
- Medizinischer Fakultätentag. (2022). *Impulspapier: Weiterentwicklung der Indikator-gestützten Mittelallokation für die Forschung in der Hochschulmedizin*. Medizinischer Fakultätentag. https://medizinische-fakultaeten.de/wp-content/uploads/2022/10/MFT-Impulspapier_IMA_LOM_2022.pdf
- Ringelhan, S., Wollersheim, J., & Welpel, I. M. (2015). Performance management and incentive systems in research organizations: Effects, limits and opportunities. In Welpel, I.,

- Wollersheim, J., Ringelhan, S., & Osterloh, M. (Hrsg.), *Incentives and Performance* (S. 87–103). Springer, Cham. https://doi.org/10.1007/978-3-319-09785-5_6
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Ker-ridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. <https://doi.org/10.13140/RG.2.1.4929.1363>
- Wissenschaftsrat. (2020). *Stellungnahme zur Einführung des Kerndatensatz Forschung* (Drs. 8652–20). https://www.wissenschaftsrat.de/download/2020/8652-20.pdf?__blob=publicationFile&v=5
- Wolbring, T. (2015). Anatomie des Journal Impact Faktors. *Soziale Welt*, 66(2), 121–140. <https://doi.org/10.5771/0038-6073-2015-2-121>

Peters, Isabella Isabella Peters ist Professorin für Web Science an ZBW – Leibniz-Informationszentrum Wirtschaft und der Christian-Albrechts-Universität zu Kiel. Sie erforscht den Einsatz von Social Media in der wissenschaftlichen Kommunikation, z. B. Altmetrics, und in Open Science.

Biesenbender, Sophie Sophie Biesenbender ist promoviert in Politik- und Verwaltungswissenschaften. Sie forscht und arbeitet seit Oktober 2012 am DZHW (bzw. der Vorgängerorganisation, dem Institut für Forschungsinformation und Qualitätssicherung – iFQ) in Berlin mit einem Fokus auf die Themen Forschungsinformationen und Indikatorik. Zwischen 2020 und Anfang 2022 war sie kommissarische Leiterin der Abteilung „Forschungssystem und Wissenschaftsdynamik“ des DZHW. Seit 2022 ist sie Leiterin der Geschäftsstelle der Kommission für Forschungsinformationen in Deutschland (KFID).

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



QuaMedFo-Fächerklassifikation

Im Verbundprojekt QuaMedFo haben vier Forschungseinrichtungen aus unterschiedlichen theoretischen, methodischen und empirischen Perspektiven die von drei Pilotfakultäten zur Verfügung gestellten Daten analysiert. Von Prof. Dr. Christoph Herrmann-Lingen wurde eine von allen Teilprojekten verwendete Fächerklassifikation erarbeitet, zu welcher die Einrichtungen an den Pilotfakultäten zugeordnet wurden. Diese stellt eine Modifikation der im Jahr 2010 im Medizinischen Fakultätentag (MFT) vorgeschlagenen Klassifizierung in Fächergruppen als Grundlage für die Erhebung von Forschungskennzahlen dar, welche Eingang in das Projekt GOMED – Governance Hochschulmedizin gefunden hat (Krempkow et al., 2013a; siehe insbesondere Krempkow et al., 2013b, S. 58). So lassen sich die verschiedenen in diesem Band dargestellten Teilergebnisse aufeinander beziehen. Die folgende Tabelle gibt einen Überblick, welche Fächer jeweils in den Fächergruppen enthalten sind.

Fächergruppe	Fächer
Vorklinische Fächer	<ul style="list-style-type: none">• Anatomie (Neuroanatomie, Embryologie, Zellbiologie, Neurobiologie etc.)• Biochemie (Biochemie, Entwicklungsbiochemie, Molekularbiologie etc.)• Physiologie (inkl. Neuro-, Sinnes-, Herz-Kreislauf-, Pathophysiologie etc.)

(Fortsetzung)

(Fortsetzung)

Fächergruppe	Fächer
Chirurgie	<ul style="list-style-type: none"> • Allgemein- und Visceralchirurgie • Transplantationschirurgie • Herz-, Thorax- und Gefäßchirurgie • Kinder(herz)chirurgie • Neurochirurgie • Unfallchirurgie • Orthopädie • Urologie • Sonstige
Innere Medizin	<ul style="list-style-type: none"> • Gastroenterologie, Hepatologie, Endokrinologie • Infektiologie • Hämatologie, Onkologie • Rheumatologie • Nephrologie • Kardiologie, Pneumologie • Angiologie • Sonstige (Sportmedizin) • Kinderkardiologie
Psychomedizinische Fächer	<ul style="list-style-type: none"> • Allg. Psychiatrie und Psychotherapie • Psychosomatische Medizin • Kinder- und Jugendpsychiatrie • Sonstige (Med. Psychologie, Med. Soziologie, Gerontopsychiatrie) • Psychopathologie/Neuropsychologie
Neuromedizinische Fächer	<ul style="list-style-type: none"> • Neurologie • Neurophysiologie • Neuroimmunologie • Neuropathologie • Neuroradiologie • Neurodegeneration • Neurorestauration • Neuropädiatrie

(Fortsetzung)

(Fortsetzung)

Fächergruppe	Fächer
Weitere klinische Fächer	<ul style="list-style-type: none"> • Anästhesiologie, Notfallmedizin, Intensivmedizin, Palliativmedizin • Augenheilkunde (inkl. Neuroophthalmologie) • Dermatologie, Venerologie, Allergologie • Gynäkologie, Geburtshilfe • HNO (Pädaudiologie, Phoniatrie) • Kinderheilkunde (Kinderkardiologie, -onkologie, -neurologie) • Neonatologie • Sonstige
Diagnostische und Strahlenfächer	<ul style="list-style-type: none"> • Diagnostisch-interventionelle Radiologie (inkl. Radiodiagnostik, Strahlenheilkunde/Radioonkologie, Nuklearmedizin) • Klin. Chemie • Pathologie • Gastroenteropathologie • Strahlentherapie/Radioonkologie • Nuklearmedizin
Weitere klinisch-theoretische Fächer und Serviceeinheiten	<ul style="list-style-type: none"> • Hygiene, Mikrobiologie • Tropenhygiene • Virologie • Humangenetik • Transfusionsmedizin • Sonstige
Theoretische Fächer	<ul style="list-style-type: none"> • Arbeits- und Sozialmedizin, Umweltmedizin • Allgemeinmedizin • Epidemiologie, Med. Biometrie und Med. Informatik (inkl. Biostatistik, Med. Statistik, Gen. Epidemiologie, Bioinformatik) • Magnetresonanztomografie in der Neurologie und Psychiatrie • Geschichte, Theorie, Ethik der Medizin • Immunologie • Zelluläre Immunologie • Serologie • Rechtsmedizin • Pharmakologie und Toxikologie (Klin. Pharmakologie, Psychopharmakologie, Molekulare Pharmakologie) • Sonstige (Molekulare Onkologie, Stammzellbiologie, Med. Physik etc.)

(Fortsetzung)

(Fortsetzung)

Fächergruppe	Fächer
Zahnmedizin	<ul style="list-style-type: none"> • Kieferorthopädie • Mund-, Kiefer-, Gesichtschirurgie • Prothetik • Zahnerhaltung, Präventive Zahnheilkunde, Parodontologie • Sonstige

Referenzen

Krempkow, R., Landrock, U., Neufeld, J., & Schulz, P. (2013a). Intendierte und nicht-intendierte Effekte dezentraler Anreizsysteme am Beispiel der fakultäts-internen leistungsorientierten Mittelvergabe in der Medizin. iFQ – Institut für Forschungsinformation und Qualitätssicherung. <https://www.forschungsinform.de/projekte/GOMED/GOMED-Abschlussbericht.pdf>

Krempkow, R., Landrock, U., Neufeld, J., & Schulz, P. (2013b). Anhang zum Abschlussbericht des Projektes GOMED – Governance Hochschulmedizin. iFQ – Institut für Forschungsinformation und Qualitätssicherung. <https://www.wissenschaftsmanagement-online.de/beitrag/intendierte-und-nicht-intendierte-effekte-dezentraler-anreizsysteme-am-beispiel-der-fakult>

Datensatz für Sekundäranalysen: Befragung von wissenschaftlichen und administrativen Beschäftigten an medizinischen Fakultäten im Projekt QuaMedFo

Die Befragung wissenschaftlicher und administrativer Beschäftigter an medizinischen Fakultäten in Deutschland war Teil des Moduls „Praktikabilität, Akzeptanz und Verhaltenswirksamkeit“ im Verbundprojekt „Qualitätsmaße zur Evaluierung medizinischer Forschung“ (QuaMedFo). Sie wurde in enger Zusammenarbeit des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (DZHW) und der Universitätsmedizin Göttingen (UMG) geplant und durchgeführt. Als Projektergebnis wurden die aufbereiteten Daten als Scientific Use File

(SUF) im Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) publiziert.

Die Befragung schließt inhaltlich an die Befragung im Projekt „Governance Hochschulmedizin“ (GOMED) aus dem Jahr 2011 am damaligen Institut für Forschungsinformation und Qualitätsentwicklung (iFQ – heute: Abteilung 2 des DZHW) an. Wichtige und noch heute relevante Fragen aus GOMED zur Leistungsorientierten Mittelvergabe (LOM) an medizinischen Fakultäten wurden aktualisiert und um Fragen zu möglichen neuen Indikatoren und etwaigen Verhaltensanpassungen ergänzt.

Es handelt sich um eine querschnittliche Erhebung, die als standardisierte Onlinebefragung im Zeitraum vom 15. April bis zum 9. August 2021 durchgeführt wurde. Die offene Einladung zur Umfrage (Respondenten-gesteuerte Auswahl) mit Unterstützung der QuaMedFo-Pilotfakultäten ergab eine Nettostichprobe von $n = 516$ Personen.

Der Datensatz steht im FDZ-DZHW für wissenschaftliche Zwecke kostenlos zur Verfügung. Nach Abschluss eines Datennutzungsvertrags, der die Einhaltung von Datenschutzbestimmungen regelt, kann der Datensatz heruntergeladen werden:

- Biesenbender, S., Hartstein, J., Herrmann-Lingen, C., Kratzenberg, A., Traylor, C., Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), & Universitätsmedizin Göttingen. (2022). *Befragung von wissenschaftlichen und administrativen Beschäftigten an medizinischen Fakultäten im Projekt QuaMedFo (Qualitätsmaße zur Evaluierung medizinischer Forschung)*. German Centre for Higher Education Research and Science Studies (DZHW). <https://doi.org/10.21249/DZHW:QMF2021:1.0.0>

Als Begleitdokumente stehen der Methodenbericht zur Umfrage sowie der Fragebogen ohne Zugangseinschränkungen zum Download zur Verfügung:

- Biesenbender, S., Hartstein, J., Herrmann-Lingen, C., Kratzenberg, A., & Traylor, C. (2022). *Methodenbericht zur Befragung von wissenschaftlichen und administrativen Beschäftigten an medizinischen Fakultäten im Projekt QuaMedFo (Qualitätsmaße zur Evaluierung medizinischer Forschung)*. <https://doi.org/10.21249/DZHW:QMF2021-DMR:1.0.0>

- Biesenbender, S., Hartstein, J., Herrmann-Lingen, C., Kratzenberg, A. & Traylor, C. (2022). *Fragebogen zur Befragung von wissenschaftlichen und administrativen Beschäftigten an medizinischen Fakultäten im Projekt Qua-MedFo (Qualitätsmaße zur Evaluierung medizinischer Forschung)*. Hannover: FDZ-DZHW. <https://doi.org/10.21249/DZHW:qmf2021-ins1-att1:1.0.0>

Die Datenerhebung wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 16PU17011A gefördert.

Literatureempfehlungen aus dem Projektkontext

- Aman, V. (2022). Do authorship regulations in performance-based funding systems affect authorship practices? A case study of German medical faculties. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Hrsg.), *Proceedings of the 26th International Conference on Science, Technology and Innovation Indicators (STI 2022)*. Granada. <https://doi.org/10.5281/zenodo.6912639>.
- Aman, V., & van den Besselaar, P. (2023). Effects of performance-based funding systems on publication behavior – A case study of German medical faculties. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4388750>.
- Brierley, L., Nanni, F., Polka, J. K., Dey, G., Pálffy, M., Fraser, N., & Coates, J. A. (2022). Tracking changes between preprint posting and journal publication during a pandemic. *PLOS Biology*, 20(2), e3001285. <https://doi.org/10.1371/journal.pbio.3001285>.
- Brierley, L., Nanni, F., Polka, J., Dey, G., Pálffy, M., Fraser, N., & Coates, J. A. (2021). Preprints in motion: Tracking changes between preprint posting and journal publication during a pandemic. *BioRxiv*. <https://doi.org/10.1101/2021.02.20.432090>.
- Fraser, N., Bräuer, P., & Peters, I. (2021). Altmetrics for evaluation of medical research in Germany. In *Proceedings of 18th International Conference on Scientometrics and Informetrics (ISSI2021)*. Zenodo. <https://doi.org/10.5281/zenodo.4743900>.
- Fraser, N., Brierley, L., Dey, G., Polka, J. K., Palfy, M., Nanni, F., & Coates, J. A. (2021). *preprinting-a-pandemic/pandemic_preprints (V3.0)* [Software]. GitHub: https://github.com/preprinting-a-pandemic/pandemic_preprints.
- Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálffy, M., Nanni, F., & Coates, J. A. (2021c). Preprinting the COVID-19 pandemic. *BioRxiv*. <https://doi.org/10.1101/2020.05.22.111294>.
- Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálffy, M., Nanni, F., & Coates, J. A. (2021d). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*, 19(4), e3000959. <https://doi.org/10.1371/journal.pbio.3000959>.
- Herrmann-Lingen, C., Brunner, E., Hildenbrand, S., Loew, T. H., Raupach, T., Spies, C., Treede, R.-D., Vahl, C.-F., & Wenz, H.-J. (2014). Evaluation of medical research performance – Position paper of the Association of the Scientific Medical Societies in Germany (AWMF). *GMS German Medical Science; 12:Doc11; ISSN 1612–3174*. <https://doi.org/10.3205/000196>.

- Lemke, S., Witthake, A. & Peters, I. (2022). Altmetrics for German medical research: What leads to research articles achieving policy impact? In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Hrsg.), *Proceedings of the 26th International Conference on Science, Technology and Innovation Indicators (STI 2022)*. Granada. <https://doi.org/10.5281/zenodo.6912582>.
- Traylor, C., & Herrmann-Lingen, C. (2023). Does the journal impact factor reflect the impact of German medical guideline contributions? *Scientometrics*, 128(3), 1951–1962. <https://doi.org/10.1007/s11192-023-04647-z>.

Stichwortverzeichnis

A

Altmetrics, [XV](#), [XVII](#), [27](#), [34](#), [36](#), [65](#), [70](#),
[79](#), [133](#), [136](#), [138](#)
Altmetrics-Abdeckungsrate, [79](#)

B

Bewertungsdimension, [130](#), [134](#)
Bewertungssystem, [6](#), [15](#), [37](#), [66](#), [138](#)
Bibliometrie, [XV](#), [37](#), [43](#), [44](#), [61](#), [81](#), [135](#)

C

Coalition for Advancing Research
Assessment (CoARA), [22](#), [139](#)

D

Datenbank, [4](#), [22](#), [37](#), [44](#), [68](#), [88](#), [98](#), [134](#),
[138](#)
Datenqualität, [138](#)
Datenquelle, [67](#), [70](#), [81](#), [138](#)
Drittmittelindikator, [10](#), [111](#), [116](#)

E

Effizienzmessung, [35](#), [115](#), [122](#), [136](#)
Evaluation
auf Personenebene, [13](#), [21](#), [139](#)
institutionelle, [23](#)
interinstitutionelle, [132](#)

intrainstitutionelle, [2](#), [110](#), [132](#)
Evaluationssystem, [4](#), [5](#), [37](#)

F

Fächerklassifikation, [53](#), [112](#), [114](#), [123](#),
[132](#), [143](#)
Feldnormalisierung, [53](#)
Forschungsevaluation, [XV](#), [XVI](#), [XVIII](#), [8](#),
[19](#), [37](#), [80](#), [93](#), [109](#), [112](#), [134](#)
Forschungsinformation, [22](#), [112](#), [122](#), [129](#),
[134](#), [139](#), [147](#)
Forschungsinformationssystem, [22](#), [112](#),
[124](#)
Forschungsleistung, [XIII](#), [XIV](#), [8](#), [15](#), [29](#),
[97](#), [118](#), [136](#)
Forschungoutput, [10](#), [28](#), [31](#), [35](#), [48](#), [67](#),
[130](#), [131](#)
Forschungssteuerung, [22](#), [43](#), [122](#), [139](#)

G

Guideline Impact Factor (GLIF), [35](#)

I

Impact
gesellschaftlicher, [15](#), [32](#), [86](#), [133](#)
wissenschaftlicher, [4](#), [25](#), [33](#), [60](#), [130](#),
[131](#)

Indikatorgestützte Mittelallokation (IMA),
1, 24
Informationssystem, 22
Input-Output-Relation, 11, 111
Integrationsmodell, 10, 121
Interdisziplinarität, 13, 25, 112
Internationale Patentklasse (IPC), 99

J

Journal Impact Factor (JIF), 33, 56, 131

K

KDSF-Standard, 27, 112, 113, 122, 134
Kerndatensatz Forschung, 134
Klassifikation, 38, 44, 77, 135
Klassifikationsmodell, 121
Ko-Okkurenz-Netzwerk, 71, 73

L

Leiden Manifesto for Research Metrics, 80,
139
Leistungsdimension, 36, 130, 131, 138
Leistungsorientierte Mittelvergabe (LOM),
1, 14, 24, 44, 110, 112, 130, 131,
147
Leitlinienindikatorik, 35, 93, 133
Leitlinienreferenz, 86

M

Matthäus-Effekt, 33
MeSH-Ontologie, 99
Metadaten, 32, 101, 102
Metrik

leistungsorientierte, XVII
verantwortungsvolle, 22, 134, 138

O

Output, wissenschaftlicher, 10, 28, 31, 35,
48, 67, 130, 131

P

Patentindikator, 97
Peer Review, 12
Plattform, altmetrische, 66, 72, 133
Policy-Dokument, 66, 72, 75, 133
Praxistransfer, XIV, XVIII, 10, 29, 85
Predatory Publishing, 7
Publikationsanalyse, 44, 45, 50
Publikationsdatenbank, 21, 29, 44, 89
Publikationsindikator, 50
Publikations-Patent-Paar, 35, 98, 102

Q

Qualitätsbegriff, VII
Qualitätskriterium, VIII, 13
Qualitätsmessung, XIII, XIV, XVI
Qualitätssicherung, VIII, 3, 25
QuaMedFo-Fächerklassifikation, 113, 115,
132, 135, 143

R

Replikation, 6, 27
Repositorium, 81, 138
Responsible Metrics, 22, 134, 138
Rezeptionsanalyse, XVII, 34, 44, 50

S

San Francisco Declaration on Research
Assessment (DORA), 4, 139
Societal Impact, 15, 32, 86, 133
Steuerungsfunktion, 2, 23, 139
Steuerung, wissenschaftliche, 22, 43, 122,
139
Szientometrie, 21, 134

T

Transfer aus der Forschung, 9, 29, 31, 85,
97, 134, 138

V

Verwertung von Forschung, 97, 134

W

Web Science, [XV](#)

Wettbewerb, [2](#), [21](#), [43](#), [110](#), [140](#)

Wissenstransfer, [VIII](#), [2](#), [85](#), [97](#), [134](#)

Z

Zitationsmaß, [50](#), [89](#)