

Peters, Isabella; Biesenbender, Sophie

Book Chapter — Published Version

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

Suggested Citation: Peters, Isabella; Biesenbender, Sophie (2024) : Qualitätsmaße zur Evaluierung medizinischer Forschung – Erkenntnisse und Empfehlungen aus dem Projekt QuaMedFo, In: Biesenbender, Sophie Hartstein, Judith (Ed.): Qualitätsmessung als Prisma, ISBN 978-3-658-43683-4, Springer VS, Wiesbaden, pp. 129-141, https://doi.org/10.1007/978-3-658-43683-4_8

This Version is available at:
<http://hdl.handle.net/11108/632>

Kontakt/Contact

ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics
Düsternbrooker Weg 120
24105 Kiel (Germany)
E-Mail: info@zbw.eu
<https://www.zbw.eu/de/ueber-uns/profil-der-zbw/veroeffentlichungen-zbw>

Standard-Nutzungsbedingungen:

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

Terms of use:

This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.



<http://creativecommons.org/licenses/by/4.0/>

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 forschungsinformationbasierten 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

- 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 institutionsspezifische 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 der 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 multinationaler 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 Drittmiteleinahmen 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 QuaMedFo 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 Evaluierungsergebnisses (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 (Wilsdon 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 Drittmittelinwerbung, 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://sfdora.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 inter-institutionellen 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). Altimetrics 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., & Welp, I. M. (2015). Performance management and incentive systems in research organizations: Effects, limits and opportunities. In Welp, 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., Kerbridge, 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. Altimetrics, 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.

