

Lemke, Steffen; Witthake, Anne; Peters, Isabella

Book Chapter — Published Version

Altmetrics zur Evaluation medizinischer Forschung in Deutschland

Suggested Citation: Lemke, Steffen; Witthake, Anne; Peters, Isabella (2024) : Altmetrics zur Evaluation medizinischer Forschung in Deutschland, In: Biesenbender, Sophie Hartstein, Judith (Ed.): Qualitätsmessung als Prisma, ISBN 978-3-658-43683-4, Springer VS, Wiesbaden, pp. 65-84, https://doi.org/10.1007/978-3-658-43683-4_4

This Version is available at:

<http://hdl.handle.net/11108/631>

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/deed.de>

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/>

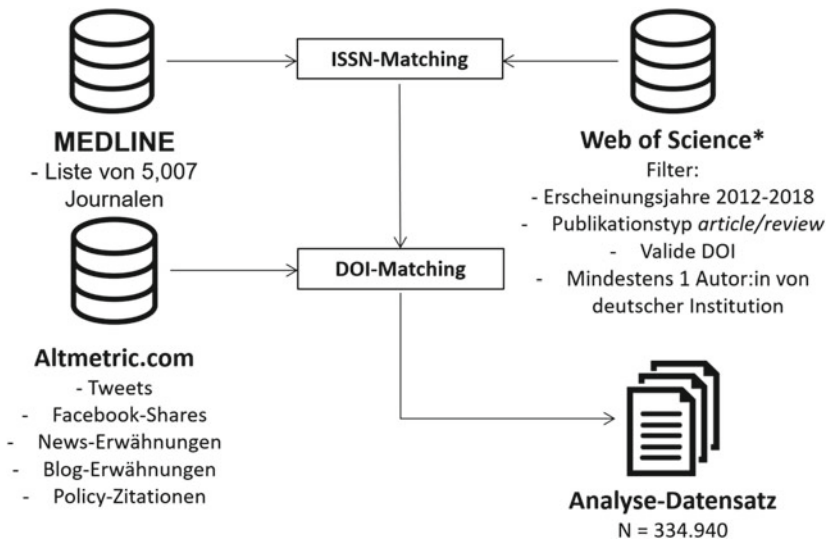


Abb. 1 Schematische Darstellung der Datensatzerstellung

2 Datensatz

Zur Erstellung eines Datensatzes unter deutscher Mitarbeit entstandener medizinischer Forschung beginnen wir mit der Liste in MEDLINE² indexierter 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 Altmatic.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 Altmatic.com nur Einträge für Artikel enthält, die mindestens eine Nennung auf mindestens einer der beobachteten Plattformen erhalten haben, wurden für alle Artikel ohne Altmatic.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 Altmatic.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.

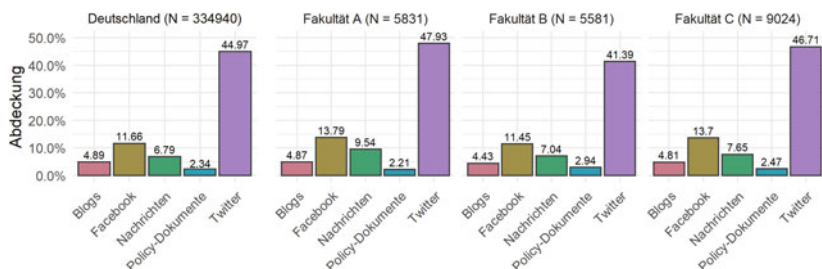


Abb. 2 Abdeckung der Publikationen des Hauptdatensatzes sowie der drei Pilotfakultäten in den fünf betrachteten altmetrischen Datenquellen.

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

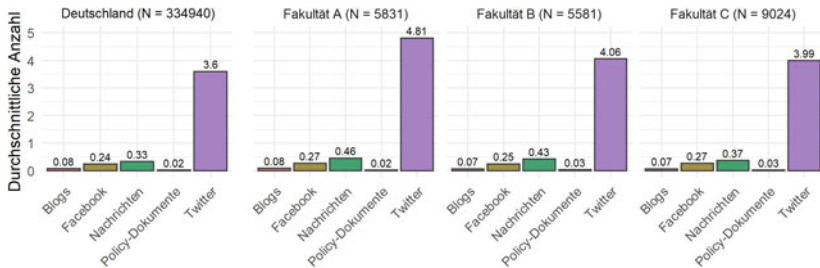


Abb. 3 Ausprägung von Nennungen der Publikationen des Hauptdatensatzes sowie der drei Pilotfakultäten in den fünf betrachteten altmetrischen Datenquellen.

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

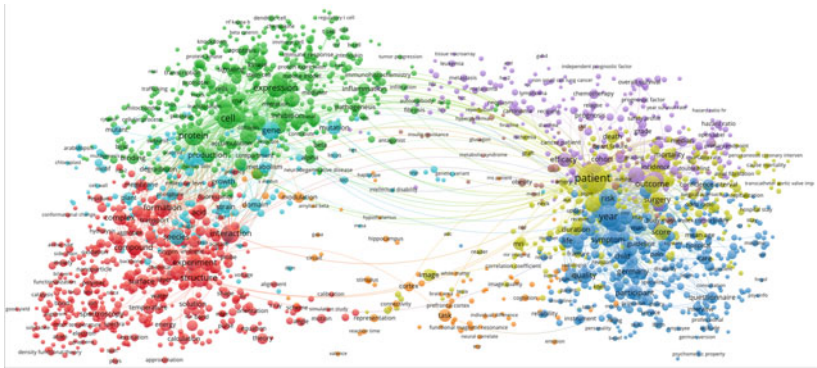


Abb. 4 Ko-Okkurrenz-Netzwerk der Terme aus 100.000 zufälligen Abstracts und Titeln des Datensatzes

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

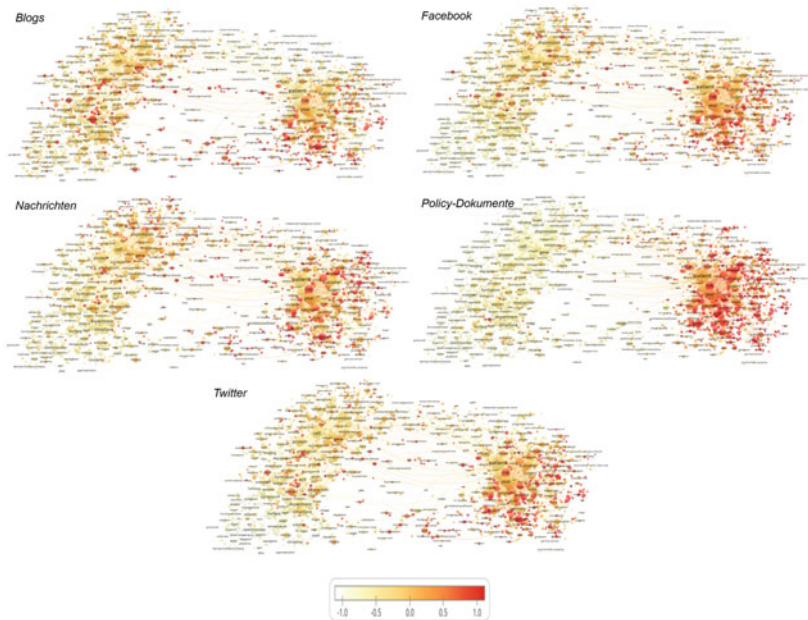


Abb. 5 Term-Ko-Okkurrenz-Netzwerk, eingefärbt entsprechend der Ausprägung altmetrischer Indikatoren bei jeweiligen Artikeln (o. l. Nennungen in Blogs, o. r. Nennungen auf Facebook, m. l. Nennungen in Nachrichten, m. r. Nennungen in Policy-Dokumenten, u. Nennungen auf Twitter). Die Farbvariationen drücken 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.

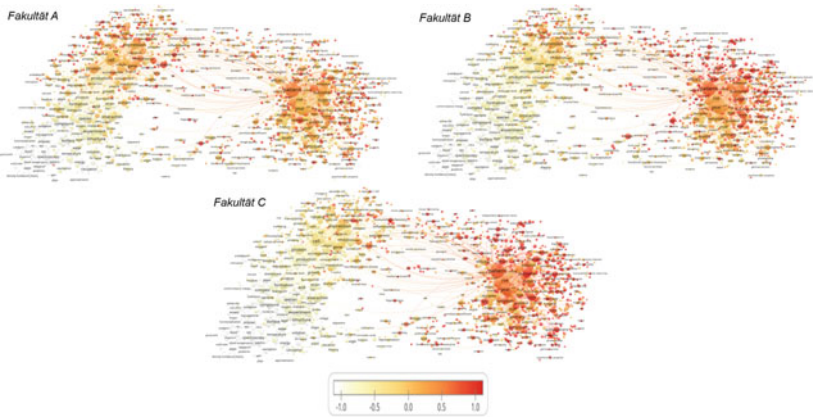


Abb. 6 Term-Ko-Okkurrenz-Netzwerk, eingefärbt entsprechend Ausprägung von Publikationen der drei Pilotfakultäten mit Vorkommen der jeweiligen Terme (o. l. Fakultät A, o. r. Fakultät B, u. Fakultät C). Die Farbvariationen drücken die relativen Häufigkeiten aus, mit denen jeweilige Terme in Abstracts und/oder 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

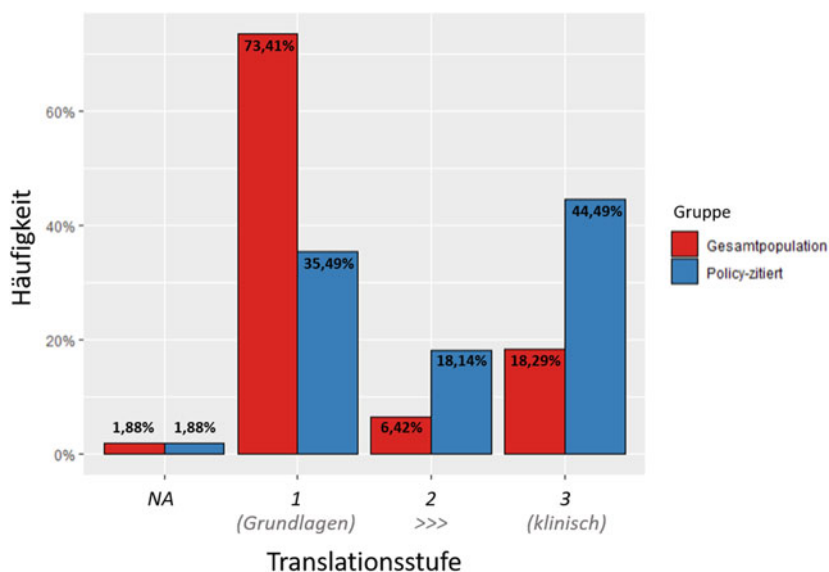


Abb. 7 Anteile per Text-Klassifikation bestimmter Translationsstufen an Artikeln der Gesamtpopulation (rot) und der Artikelgruppe mit je mindestens einer Policy-Dokument-Zitation (blau)

(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

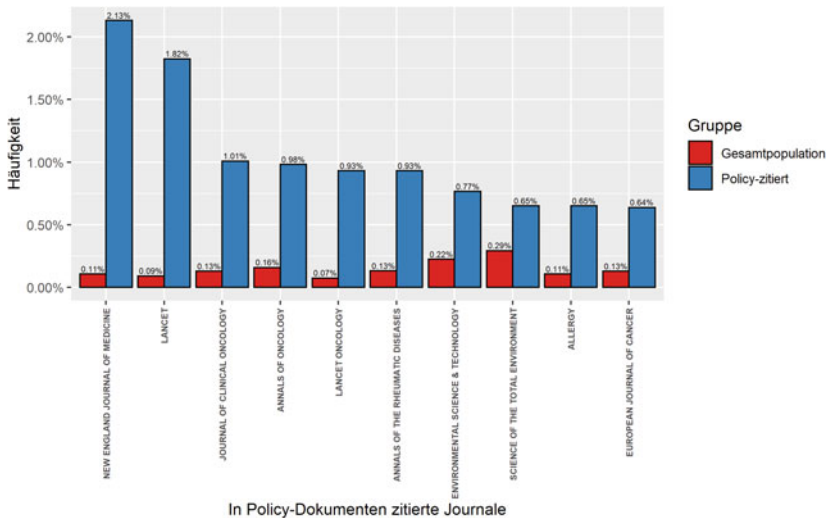


Abb. 8 Die zehn häufigsten Journale unter Artikeln mit mindestens einer Policy-Dokument-Zitation sowie ihre relativen Anteile in beiden untersuchten Gruppen

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 Nutzbar-machung altmetrischer Daten deutlich: in den meisten altmetrischen Quellen befinden sich die prozentualen Abdeckungs-raten 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

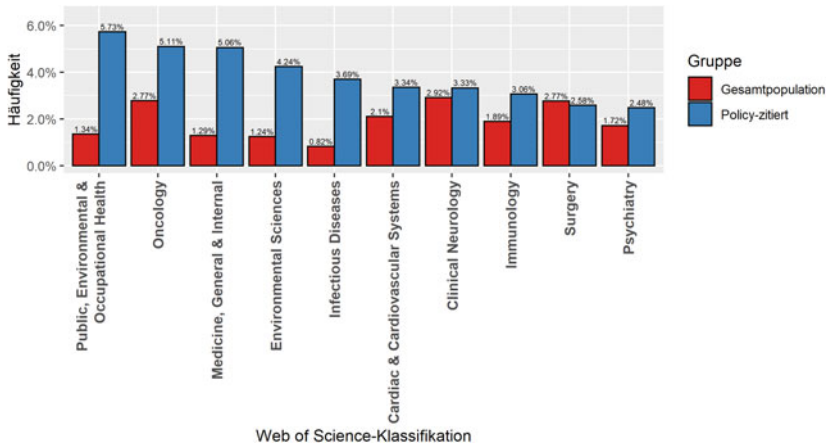


Abb. 9 Die zehn häufigsten Web of Science-Klassifikationen unter Artikeln mit mindestens einer Policy-Dokument-Zitation sowie ihre relativen Anteile in beiden untersuchten Gruppen

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. *Informaatitutkimus*, 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., 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>.
- 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 Datenanalystin 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.

