

Foto: utah51/stock.adobe.com

Medizinische Forschung

Zusammenarbeit in der Datenwissenschaft stärken

In der medizinischen Forschung spielt die technisch gestützte Datenwissenschaft eine nicht mehr wegzudenkende Rolle. In diesem Rahmen wirken verschiedene Fachbereiche mit ihren jeweiligen Kompetenzen zusammen – was Schnittstellenkoordination erforderlich macht.

Text | Antonia Zapf, Michael Altenbuchinger, Johanna Apfel-Starke, Tim Beissbarth, Erika Graf, Carolin Jenkner, André Karch, Matthias Katzensteiner, Jochem König, Claudia Ose, Erik Tute

Hochschule Hannover: Katzensteiner

Medizinische Hochschule Hannover: Tute

Medizinische Hochschule Hannover und Hochschule Hannover: Apfel-Starke

Universitätsklinikum Essen und Fliedner Fachhochschule: Ose

Universitätsklinikum Freiburg: Graf, Jenkner

Universitätsmedizin Göttingen: Altenbuchinger, Beissbarth

Universitätsklinikum Hamburg-Eppendorf: Zapf

Universitätsmedizin Mainz: König

Universitätsklinikum Münster: Karch

Medizinische Forschung basiert auf Theorien, die durch Daten gestützt werden. Die verfügbare Datenmenge ist in den letzten Jahrzehnten exponentiell gewachsen und mit dem technischen Fortschritt entstanden neue Möglichkeiten, die allerdings spezifische Expertise erfordern. In der medizinischen Datenwissenschaft sind verschiedene Fachrichtungen beteiligt – etwa Bioinformatik, Biostatistik, Epidemiologie, Datenmanagement und Medizininformatik. Zusätzlich hat sich die Berufsbezeichnung Medizinische Datenwissenschaftlerin/Medizinischer Datenwissenschaftler (im Englischen Medical Data Scientist) etabliert, ohne dass klar definiert wäre, welche Kompetenzen damit verbunden sind. Medical Data Science wird häufig als Oberbegriff für die genannten Bereiche verwendet, fokussiert aber stark auf Methoden des Maschinellen Lernens und der Künstlichen Intelligenz (KI). Offen bleibt, ob diese Generalisierung der Komplexität medizinischer Forschung gerecht wird. Zudem wissen medizinisch Forschende oft nicht, worin sich die Disziplinen unterscheiden und wer für ihre Bedarfe die richtigen Ansprechpartner sind – zumal sich sowohl Überschneidungen als auch Alleinstellungsmerkmale zeigen. So wird beispielsweise Biostatistik in experimentellen klinischen Studien, aber auch in

der Epidemiologie und Bioinformatik eingesetzt. Für Forschende ist daher oft unklar, welche Disziplin welche Aufgaben abdeckt [1]. Ziel dieses Artikels ist es, die Beiträge der einzelnen Fachrichtungen vorzustellen, Kernkompetenzen und Überschneidungsbereiche zu identifizieren und daraus Empfehlungen für eine bessere interdisziplinäre Zusammenarbeit abzuleiten. Hierbei orientiert sich die Vorstellung der Fachdisziplinen an der Durchführung und Analyse einer randomisierten kontrollierten klinischen Studie – die Tätigkeiten der einzelnen Disziplinen werden also bewusst nicht vollständig wiedergegeben. Medizinische Studien gliedern sich in Vorarbeiten, Studienplanung, Aufsetzen, Durchführung, Auswertung und Publikation. Im Folgenden werden die Aufgaben der Disziplinen entlang dieser Phasen beschrieben und im Onlineappendix konkret für eine Studie zum Vergleich zweier Bildgebungstechniken im Gebiet der Mammografie vorgestellt.

Tätigkeiten in den Studienphasen

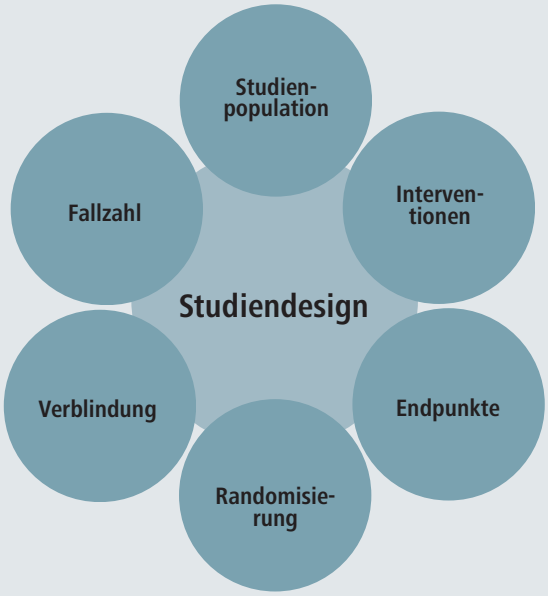
Vor einer Studie sind Laboruntersuchungen und frühe klinische Studien nötig, um zum Beispiel Dosis, Biomarker und Endpunkte festzulegen. Beobachtungsstudien helfen, Zielpopulationen und relevante Effekte zu identifizieren. Epidemiologie, Biostatistik und Bioinformatik tragen hier zur Methodik, Datenanalyse und Hypothesenbildung bei.

Ziel der Planungsphase ist es, eine relevante und überprüfbare Forschungsfrage zu definieren. Epidemiologie und Biostatistik helfen, die Fragestellung zu operationalisieren und ein dazu passendes optimales Studiendesign zu entwickeln *(siehe Grafik)* [2]. Geplante Zwischenanalysen und adaptive Designs wie auch die zu Fragestellung und Studiendesign passenden Analysen müssen hier bereits spezifiziert werden. Die Medizininformatik entwickelt und implementiert das Forschungsdatenmodell und plant zudem Datenfluss und -integration – insbesondere bei mehreren Datenquellen wie Routine- und Studiendaten. Das Datenmanagement erstellt darauf basierend Datenmanagementplan und Datenschutzkonzept mit klaren Rollen [3]. Für extern finanzierte Studien sind Förderanträge und Genehmigungen durch Zulassungs- und Aufsichtsbehörden erforderlich, an denen Epidemiologie, Biostatistik, Medizininformatik und Datenmanagement mitwirken.

Nach der Planung erfolgt die Vorbereitung der Datenerhebung. Medizininformatik und Datenmanagement entwickeln elektronische Case Report Forms (eCRFs), setzen die Datenbank auf und validieren diese, die Biostatistik prüft die eCRFs. Beide Disziplinen arbeiten am risikobasierten Qualitätsmanagement, um eine hohe Datenqualität sicherzustellen.

Grafik

Verschiedene Aspekte des Designs einer klinischen Studie



Literatur
https://daebl.de/7wj9

Epidemiologie und Biostatistik sind maßgeblich an der Erstellung des Studienprotokolls beteiligt, insbesondere bei Studiendesign, Methoden gegen Verzerrung, Fallzahlplanung und Analyseverfahren. Sie erstellen den statistischen Analyseplan und Randomisierungslisten. Gleichzeitig werden Gremien wie Steering Committee (SC) und Data Safety Monitoring Board (DSMB) besetzt, in denen auch Biostatistik vertreten sein sollte.

In der Durchführungsphase steht die Datenerhebung im Vordergrund, wobei zwischen Primärdaten und Sekundärdaten unterschieden wird. Primärdaten werden direkt für die Studie von study nurses, Studienärztinnen und -ärzten oder durch die Studienteilnehmenden selbst erhoben. Sekundärdaten, wie Kassendaten oder Daten aus Klinikinformationssystemen, wurden dagegen ursprünglich für einen anderen Zweck erhoben. Die Verknüpfung beider Datentypen erfordert Expertise aus Epidemiologie und Medizininformatik. Gemäß Datenmanagementplan werden Datenquellen verknüpft, bereinigt, qualitätsgeprüft und pseudonymisiert als Forschungsdatenbank bereitgestellt. Das DSMB überwacht Sicherheitsdaten und gibt Empfehlungen zur Fortführung der Studie. In adaptiven Designs führen unabhängige Biostatistiker Zwischenanalysen durch und berichten die Ergebnisse dem DSMB. Von der Biostatistik werden die finalen Analysen vorbereitet und unter Verwendung von

verblindeten oder Pseudodaten Analyseprogramme erstellt. Nach Abschluss der Datenerhebung prüft das Datenmanagement die Datenbank final. Nach Klärung aller Rückfragen wird sie „eingefroren“ (auf nicht mehr änderbar gesetzt) und an die Biostatistik übergeben. Diese führt die geplanten Analysen durch, erstellt den Auswertungsbericht und unterstützt bei der Interpretation und Publikation der Ergebnisse. Dabei müssen die Studienergebnisse unter Berücksichtigung möglicher Verzerrungsquellen interpretiert und durch entsprechende Sensitivitätsanalysen abgesichert werden.

Bei Beobachtungsstudien ist eine Datenanalyse durch die Epidemiologie unter Berücksichtigung von möglichem Confounding und kausaler Strukturen, etwa mithilfe von Mediationsanalysen oder komplexen Modellen, besonders wichtig. In ergänzenden Analysen können beispielsweise von Epidemiologie und Biostatistik emulated target trials auf Basis von Routinedaten durchgeführt werden, um Subgruppen mit einem größeren Nutzen zu identifizieren. Von der Bioinformatik können hochdimensionale Bild- oder Omicsdaten mit dem Ziel der personalisierten Medizin analysiert werden. Die Medizininformatik kann darauf aufbauend KI-basierte klinische Entscheidungsunterstützungssysteme entwickeln.

Onlineappendix
Anhand einer Beispielstudie mit fiktionalen Anteilen werden die Beiträge der einzelnen Fachrichtungen veranschaulicht.
https://daebl.de/t7b8

Fachdisziplinen im Überblick

- **Epidemiologie** erforscht Verbreitung, Ursachen und Folgen von Gesundheit und Krankheit in Bevölkerungen durch Design, Durchführung und Analyse von Beobachtungsstudien unter Nutzung moderner Methoden der kausalen Inferenz.
- **Biostatistik** übersetzt biomedizinische Forschungsfragen in mathematisch-statistische Datenmodelle, stellt die adäquate Methodik dafür bereit und trägt durch sorgfältige Planung, Datenerhebung und statistische Auswertung zum Erkenntnisgewinn aus medizinischen Daten bei.
- **Datenmanagement** ist ein begleitender Prozess, der den gesamten Lebenszyklus von (Forschungs-)Daten begleitet, steuert und sich dabei unterschiedlicher Werkzeuge und Methoden (zum Beispiel informatisch) bedient.
- **Medizininformatik** beschäftigt sich mit der systematischen Erschließung, Verwaltung, Aufbewahrung, Verarbeitung und Bereitstellung medizinischer Daten. Sie setzt Theorien und Methoden der Informatik und anderer Wissenschaften ein.
- **Bioinformatik** dient der Erforschung, Entwicklung und Anwendung computergestützter Methoden zur Beantwortung molekularbiologischer und biomedizinischer Fragestellungen.

Diskussion und Fazit

Eine der wichtigsten Schlussfolgerungen ist, dass Kooperationspartner für valide und effiziente Forschung wissen sollten, welche Fachrichtung welche Kernkompetenzen besitzt und welchen Beitrag sie entsprechend erwarten können. Dafür könnte unter anderem eine Koordination an den Forschungseinrichtungen ein zentrales Element darstellen. Zudem könnten Zertifikate helfen, Expertise transparent und nachvollziehbar auszuweisen.

Auch zeigen die dargestellten Tätigkeiten, dass die Aufgaben in der medizinischen Datenwissenschaft vielfältig und komplex sind. Eine einzelne Person kann nicht die gesamte Expertise aller Disziplinen abdecken, ohne Qualitätsverluste zu riskieren. Generative KI kann die Arbeit unterstützen, erfordert jedoch präzise Anleitung und kritische Prüfung – was ebenfalls Fachwissen voraussetzt. Medical Data Scientists mit breitem Grundverständnis können als wertvolle Schnittstellen zwischen den Disziplinen fungieren: Sie übernehmen grundlegende Aufgaben selbst, vermitteln zwischen Fachrichtungen und koordinieren komplexe Projekte gemeinsam mit Spezialisten. Daher ist es wichtig, bei komplexeren Projekten Generalisten und Spezialisten der entsprechend benötigten Fachdisziplinen zu involvieren, während bei weniger komplexen Fragestellungen auch Medical Data Scientists alleine der Aufgabe gerecht werden können.

Zwischen den Disziplinen besteht bislang zu wenig Abstimmung. Besserer methodischer Austausch sowie gegenseitiges Verständnis der jeweiligen Kompetenzen könnten die Zusammenarbeit in Studien effizienter machen. Dafür sollten Tätigkeiten und Expertise der jeweils anderen Fachdisziplinen im Studium vermittelt werden. Austauschplattformen wie interdisziplinäre Tagungen oder Fachgesellschaften wie die Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) können zur Vernetzung beitragen. Auch eine institutionelle Koordination, etwa an Universitätskliniken, wäre hilfreich.

Die medizinische Datenwissenschaft vereint unterschiedliche, sich ergänzende Disziplinen. Generalisten können als Bindeglieder fungieren, ersetzen jedoch nicht die notwendige Tiefe spezialisierter Expertise. Ein strukturierter Austausch, institutionelle Kooperation und gezielte Ausbildung sind entscheidend, um das Potenzial der Medical Data Science optimal zu nutzen. ■

Zusatzmaterial Heft 20/2026, zu:

Medizinische Forschung

Zusammenarbeit in der Datenwissenschaft stärken

In der medizinischen Forschung spielt die technisch gestützte Datenwissenschaft eine nicht mehr wegzudenkende Rolle. In diesem Rahmen wirken verschiedene Fachbereiche mit ihren jeweiligen Kompetenzen zusammen – was Schnittstellenkoordination erforderlich macht.

Text | Antonia Zapf, Michael Altenbuchinger, Johanna Apfel-Starke, Tim Beissbarth, Erika Graf, Carolin Jenkner, André Karch, Matthias Katzensteiner, Jochem König, Claudia Ose, Erik Tute

Literatur

1. Zapf A, Huebner M, Rauch G, Kieser M: What makes a biostatistician? STAT MED 2019; 38 (4): 695–701; Zapf A, Rauch G, Kieser M: Why do you need a biostatistician? BMC Med Res Methodol 5. Februar 2020; 20 (1): 23. DOI: 10.1186/s12874–020–0916–4.
2. Pocock SJ, Clayton TC, Stone GW: Design of Major Randomized Trials: Part 3 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol 22. Dezember 2015; 66 (24): 2757–66.
3. Katzensteiner M, Witte ML, Müller A, Schewe N, Bott OJ, Apfel-Starke J, Tümler K (AG-Datenmanagement der GMDS): Data-Steward(-ship) in der Digitalisierung des Gesundheitswesens. mdi – Forum der Medizin_Dokumentation und Medizin_Informatik des BVMI/DVMD 2022; 22(4): 123–5.