

# Daten, Informationen, Wissen

## Methodische und strukturelle Herausforderungen der Coronapandemie\*

### Einleitung

Ich halte diesen Vortrag mit zwei Hüten auf dem Kopf, einmal als (Inhalts-)Wissenschaftler an der Christian-Albrechts-Universität zu Kiel (CAU), zum anderen als Vertreter der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF) in Berlin. Viele der im Folgenden zu besprechenden Dinge betreffen neben meiner Arbeit an der CAU auch mein Engagement bei der TMF, und nach meiner Überzeugung wird gerade die TMF als eine die Wissenschaft unterstützende Einrichtung bei der Bewältigung der Probleme und Herausforderungen, um die es heute gehen soll, eine entscheidende Rolle spielen. Doch mehr dazu im Verlaufe meines Vortrags.

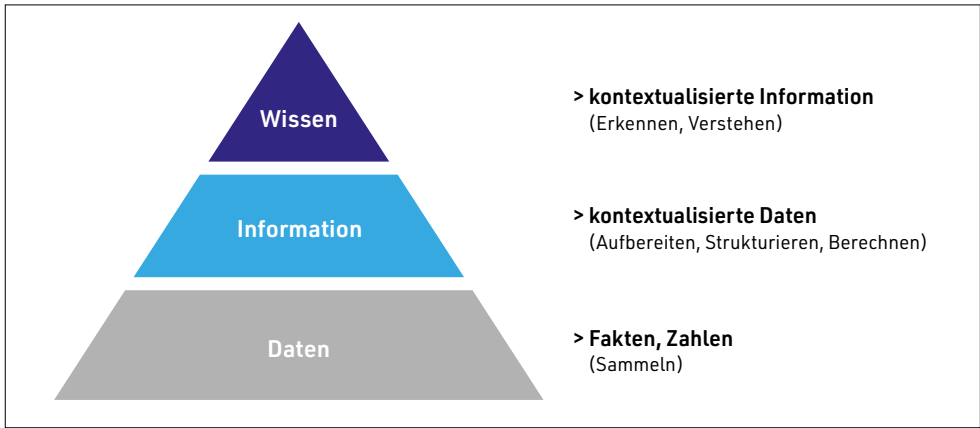
### Wissenschaftstheoretische Vorbemerkung

Ich beginne mit einem kurzen wissenschaftstheoretischen, Ihnen zum Teil vielleicht schon bekannten ›Primer‹, der den Titel meines Vortrags nachträglich motivieren soll. Daten, Informationen, Wissen bilden die drei Ebenen der sogenannten ›Wissenspyramide‹ (Abb. 1), eines gängigen Modells der Wissensgenerierung.

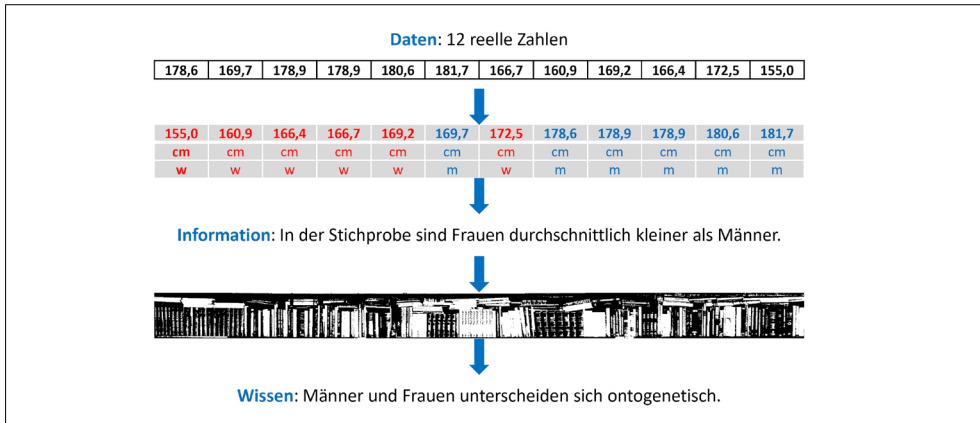
Die Wissenspyramide geht auf Agnar Aamodt und Mads Nygård, zwei norwegische Wissenschaftler, zurück. Zur eigentlichen Wissenspyramide gehören noch zwei weitere Schichten, die in Abbildung 1 nicht berücksichtigt sind, nämlich unterhalb der Daten die

---

\* Das Video zum Vortrag, der am 10.6.2021 im Rahmen der digitalen Ringvorlesung *Die Coronavirus-Pandemie und ihre Folgen II* der Christian-Albrechts-Universität zu Kiel gehalten wurde, ist einsehbar unter: <https://doi.org/10.5446/57167>.



**Abb. 1:** Die Wissenspyramide. Eigene Darstellung nach Agnar Aamodt, Mads Nygård: Different roles and mutual dependencies of data, information, and knowledge – An AI perspective on their integration, in: Data & Knowledge Engineering 16,3 (1995), S. 191–222. [https://doi.org/10.1016/0169-023X\(95\)00017-M](https://doi.org/10.1016/0169-023X(95)00017-M).



**Abb. 2:** Beispiel zur Wissenspyramide.

Zeichen und oberhalb des Wissens die Handlung. Ich konzentriere mich hier jedoch auf die drei mittleren Komponenten. Daten sind demnach zunächst einmal Fakten und Zahlen, die für sich genommen wenig Bedeutung haben. Erst durch ihre Kontextualisierung, d.h. ihre Strukturierung und Aufbereitung sowie die Berechnungen geeigneter Maßzahlen, werden sie zu Information. Diese wiederum geht durch weitere Kontextualisierung wie Erkennen oder Verstehen in Wissen über. Wichtig ist an diesem Bild, dass Daten die Basis von Information und Wissen bilden, und dass deren Güte somit entscheidend von der Güte und der Qualität der Daten abhängt.

Ich möchte den Schritt von Daten zu Information und Wissen kurz an einem Beispiel verdeutlichen: In der obersten Zeile der Abbildung 2 sieht man zwölf reelle Zahlen – Daten, die für sich genommen zunächst wenig aussagen. Wenn diese Zahlen aber kontextualisiert werden, indem man z.B. ihre Skalierung (hier Zentimeter) angibt und feststellt, dass es sich

um die Körpergrößen von Personen unterschiedlichen Geschlechts handelt, so ergibt sich folgende Information: Zumindest in der Stichprobe sind Frauen durchschnittlich kleiner als Männer. Wenn nun noch biologisches Vorwissen hinzutritt, dann kann daraus z.B. das Wissen entstehen, dass sich Männer und Frauen ontogenetisch unterscheiden.

Auch in unserem Beispiel wird der Schritt von den Daten zur Information und zum Wissen sehr von der Qualität der Daten bestimmt, neben der methodischen Herausforderung, geeignete Wege der Kontextualisierung zu finden. Mein Vortrag geht nun der Frage nach, was diese Abhängigkeit der Qualität von Information und Wissen von der Qualität von Daten für die Jahrhundertkatastrophe der Coronavirus-Pandemie bedeutet, in die wir 2020 hineingeraten sind.

## Abhängigkeit von Wissens- und Datenqualität

Zur Erinnerung: Das Ganze begann Ende 2019, als das Büro der Weltgesundheitsorganisation (WHO) in China über Fälle einer schweren Lungenerkrankung mit unbekannter Ätiologie informiert wurde. Die ersten Betroffenen waren angeblich Händler und Verkäufer auf dem Seafood Market, dem Fischmarkt von Wuhan. In der Meldung der WHO vom 5. Januar 2020 steht noch, dass es keine Evidenz für eine signifikante Übertragung der Krankheit von Mensch zu Mensch gäbe.<sup>1</sup> In der Zwischenzeit sind wir diesbezüglich leider eines Besseren belehrt worden. Interessanterweise wurde schon zu Beginn vermutet, dass das Wuhan Virology Institute, das Zentralinstitut für Virologie der Volksrepublik China, etwas mit dem Ausbruch zu tun haben könnte – ein Verdacht, der anschließend jedoch lange Zeit wieder in den Hintergrund trat.

Im Februar 2020 gab es im *New England Journal of Medicine* die erste wichtige wissenschaftliche Publikation zum Virus SARS-CoV-2, der die geheimnisvolle Krankheit COVID-19 verursacht.<sup>2</sup> Sie stammte von einer chinesischen Arbeitsgruppe um Wissenschaftler aus dem Institut in Wuhan. Danach ging es publizistisch stark voran: Schon im März 2020 war die Zahl der Veröffentlichungen zum Thema COVID-19 und Coronavirus auf etwas über 1000 gestiegen; danach, etwa im Mai 2020, stabilisierte sich das Aufkommen bei ungefähr 2000 Publikationen pro Woche. Es gibt in der Zwischenzeit bei PubMed unter der Bezeichnung ›LitCovid‹ einen eigenen Bereich zur Coronapandemie, den ich heute (am 10. Juni 2021)

---

1 »On 31 December 2019, the WHO China Country Office was informed of cases of pneumonia of unknown etiology (unknown cause) detected in Wuhan City, Hubei Province of China. [...] According to the authorities, some patients were operating dealers or vendors in the Huanan Seafood market. Based on the preliminary information from the Chinese investigation team, no evidence of significant human-to-human transmission and no health care worker infections have been reported.« World Health Organization: Covid-19 China. Meldung vom 5. Januar 2020. <https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229> (letzter Zugriff: 29.3.2022).

2 Na Zhu et al.: A Novel Coronavirus from Patients with Pneumonia in China, 2019, in: *New England Journal of Medicine* 382 (2020), S. 727-733. <https://dx.doi.org/10.1056/NEJMoa2001017>.

besucht habe.<sup>3</sup> In der Zwischenzeit sind dort mehr als 137.000 relevante Publikationen zu COVID-19 gelistet. Wenn man zählt, wie häufig bestimmte Länder in den Abstracts dieser Publikationen genannt werden, liegt ein klarer Fokus auf den USA und China. Das ist nicht verwunderlich, weil China wahrscheinlich das Ursprungsland der Pandemie ist und die USA numerisch am stärksten betroffen waren und sind. Das Aufkommen an Publikationen zu COVID-19 in Zentralasien und Afrika ist hingegen nicht sonderlich ausgeprägt.

Zur Vermutung, dass das Coronavirus etwas mit dem Virology Institute in Wuhan zu tun haben könnte, habe ich vor einer Woche von einem Kollegen den Hinweis auf einen lesenswerten Science-Blog erhalten.<sup>4</sup> Demnach waren 2012 in Yunnan – mehrere tausend Kilometer von Wuhan entfernt – in einer Mine sechs Arbeiter an Symptomen ähnlich denen von COVID-19 erkrankt, nachdem sie dort Fledermauskot entsorgt hatten. In einer 2014 in China erschienenen Masterarbeit wurde berichtet, dass Proben dieser Arbeiter schon 2013 an das Institut in Wuhan geschickt und dort positiv auf Coronaviren getestet worden waren. Im Jahre 2016 erschien dann eine chinesische Doktorarbeit, die dokumentierte, dass Fledermauskot-Proben aus der Mine 2014 im Institut ankamen und ebenfalls positiv auf Coronaviren getestet wurden.<sup>5</sup> Ein anonymes Facebook-Mitglied hat diese Doktorarbeit ins Englische übersetzen lassen und anschließend ins Internet gestellt.<sup>6</sup>

Aus der Doktorarbeit geht hervor, dass der 2013 in der Mine in Yunnan entdeckte Fledermausvirus BtCoV/4991 identisch ist mit dem, der 2020 vom Wuhan Institut angeblich erstmals in der oben genannten Publikation im *New England Journal* erwähnt wurde, wahrscheinlich aber mit dem schon 2018 dort untersuchten Virus RaTG13 übereinstimmt.<sup>7</sup> Dabei handelte es sich um den engsten genetischen Verwandten von SARS-CoV-2. All diese Informationen deuten darauf hin, dass es einen Zusammenhang zwischen dem Ausbruch der Coronapandemie und dem Virology Institute in Wuhan gegeben hat. Mit hundertprozentiger Sicherheit klären wird man das wohl nie, weil verlässliche und notwendige Daten dazu nicht mehr verfügbar oder wahrscheinlich nie verfügbar gewesen sind.<sup>8</sup>

---

3 LitCovid: <https://www.ncbi.nlm.nih.gov/research/coronavirus/> (letzter Zugriff: 7.1.2022). Zum Zeitpunkt des letzten Zugriffs, 10.6.2022, waren dort 259063 Veröffentlichungen verzeichnet.

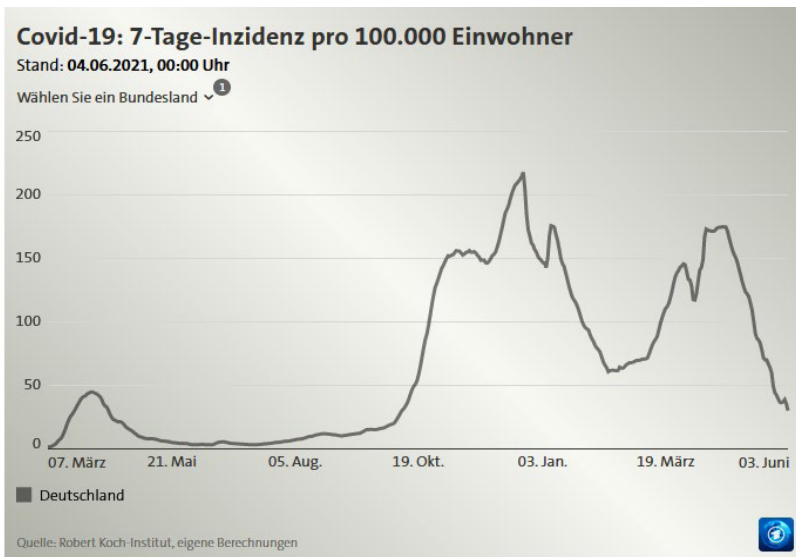
4 Jonathan Latham, Allison Wilson: A Chinese PhD Thesis Sheds Important New Light On The Origin of the COVID-19 Coronavirus, 11.5.2021. <https://www.independentsciencenews.org/commentaries/a-chinese-phd-thesis-sheds-important-new-light-on-the-origin-of-the-covid-19-coronavirus/> (letzter Zugriff: 29.3.2022).

5 Huang Canping: Novel Virus Discovery in Bat and the Exploration of Receptor of Bat Coronavirus HKU9. 2016. <https://doi.org/10.5281/zenodo.4008337>.

6 Anonyme Teilübersetzung ins Englische von Canping, Novel Virus Discovery in Bat (wie Anm. 5). <https://www.documentcloud.org/documents/20694207-canping-huang-phd-novel-virus-discovery-in-bat-isn-translation> (letzter Zugriff: 10.1.2022).

7 Siehe hierzu auch die Zusammenfassung des Vortrags *Fledermäuse als Virenschleudern: Was sagt die Forschung?* von Christian Voigt im vorliegenden Band: <http://dx.doi.org/10.38072/978-3-928794-82-4/p3>.

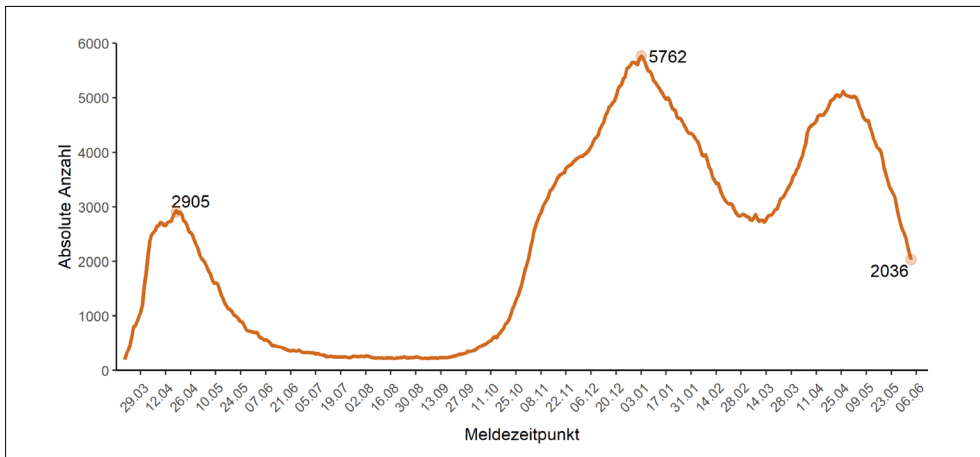
8 Siehe zu diesem Thema auch die Beiträge *Wohin führt uns die Coronavirus-Pandemie?* von Helmut Fickenscher (<https://dx.doi.org/10.38072/978-3-928794-82-4/p2>) und *Die Immunantwort bei SARS-CoV-2 und ihre Folgen* von Reinhold E. Schmidt (<https://dx.doi.org/10.38072/978-3-928794-82-4/p4>) in diesem Band.



**Abb. 3:** Screenshot der Tagesschau-Webseite vom 4.6.2021. Alle Rechte vorbehalten. Quelle: Tagesschau: [Die Corona-Lage im Überblick](#).

Spätestens seit März 2020 hat die Coronapandemie Deutschland fest im Griff. Die Abbildungen 3 und 4 zeigen zwei Kurven, die inzwischen schon fast ikonographischen Wert bekommen haben. In Abbildung 3 sind die sogenannten »Inzidenzwerte« dargestellt, die das Robert Koch-Institut (RKI) täglich veröffentlicht. Missverständlich bezeichnet die Tagesschau-Webseite, der die Abbildung 3 entstammt, diese Zahlen als »7-Tage-Inzidenz pro 100.000 Einwohner«, was so nicht stimmt: Der Inzidenzwert entspricht stattdessen der Anzahl der in sieben Tagen dem RKI gemeldeten Neuinfektionen pro 100.000 Einwohner, die sich aus mehreren Gründen von der Inzidenz unterscheidet. Trotzdem sieht man der Kurve an, dass es in Deutschland drei große Infektionswellen gegeben hat: eine vergleichsweise glimpfliche im Frühjahr 2020, dann eine sehr heftige Welle im Herbst 2020 und dann die dritte Welle, die glücklicherweise im Moment abzuklingen scheint. Selbst wenn man diesen Inzidenzwerten nicht traut, zeigt die in [Abbildung 4](#) dargestellte Anzahl der intensivmedizinisch behandelten COVID-19-Patienten, dass es diese Wellen zweifellos gegeben hat.

Warum betrifft uns die Coronapandemie so sehr? Sie ist nicht nur ein medizinisches oder epidemiologisches Problem, sie hat auch massive soziale und volkswirtschaftliche Auswirkungen. Gerade darin besteht eine wesentliche Motivation für meinen Vortrag: Angesichts der dramatischen Folgen, die die Pandemie für Wirtschaft, Bildung und Sozialleben hat, sind Politiker nicht zu beneiden, die in der aktuellen Situation Entscheidungen treffen müssen – bis hin zum Shutdown des gesamten öffentlichen Lebens. Die Politik muss dabei ständig Risiken und Nutzen ihrer Maßnahmen bewerten und kann dies nur auf der Grundlage validen Wissens und valider Information tun, d.h., wie wir gesehen haben, auf der Grundlage valider Daten. Und genau da liegt das Problem!



**Abb. 4:** Anzahl der gemeldeten COVID-19-Fallzahlen des jeweiligen Beobachtungstages (Stand 3.6.2021, 12:15 Uhr). Alle Rechte vorbehalten. Quelle: RKI: Archiv der täglichen Situationsberichte zu COVID-19, [Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 \(COVID-19\)](#) v. 3.6.2021, S. 9, Abb. 5.

## Coronapandemie als ›Missing-Data-Krise‹

Aus meiner Sicht ist die Coronapandemie in weiten Teilen eine ›Missing-Data-Krise‹. Vor zwei Monaten habe ich gemeinsam mit Amke Caliebe aus Kiel sowie mit Gerhard Antes und Friedhelm Leverkus einen Artikel im *Deutschen Ärzteblatt* veröffentlicht, in dem Sie die wesentlichen Argumente für diese Sichtweise noch einmal nachlesen können.<sup>9</sup> Im Kern geht es darum, dass viele wirklich gravierende (gesundheits-)politische Entscheidungen mit weitreichenden gesellschaftlichen Auswirkungen auf der Grundlage unzulänglicher und oftmals gänzlich fehlender Daten getroffen werden mussten und immer noch getroffen werden müssen.

Dass die Wirksamkeit der im Zuge der Pandemie ergriffenen Maßnahmen häufig nicht überprüft werden konnte, weil die dafür erforderlichen Daten und informationstechnischen Voraussetzungen fehlten, ist ein auf vielen Ebenen wiederkehrendes Phänomen:<sup>10</sup>

- Aussagekräftige Daten fehlen zum Beispiel in der Epidemiologie, wo sie die Grundlage für ein sinnvolles Pandemie-Management bilden sollten. Ich habe mehrfach betont, dass wir die Pandemie-Entwicklung sehr viel besser verstanden hätten, wenn uns von Anfang an bekannt gewesen wäre, wer zu welchem Zeitpunkt aus welchem Grund und mit welchem Ergebnis auf SARS-CoV-2 getestet wurde. Diese Daten gab

<sup>9</sup> Amke Caliebe, Gerd Antes, Friedhelm Leverkus, Michael Krawczak: Dateninfrastruktur: Ein Baustein gegen COVID-19, in: *Deutsches Ärzteblatt* 10,118 (2021), S. A501-504. <https://www.aerzteblatt.de/pdf.asp?id=218237> (letzter Zugriff: 29.3.2022).

<sup>10</sup> Siehe z. B. Wissenschaftsrat: Impulse aus der COVID-19-Krise für die Weiterentwicklung des Wissenschaftssystems in Deutschland. Positionspapier. Drucksache 8834-21. Januar 2021. <https://www.wissenschaftsrat.de/download/2021/8834-21.pdf> (letzter Zugriff: 29.3.2022).

und gibt es schlichtweg nicht und es ist nach wie vor so, dass dem RKI im Wesentlichen nur summarische Testergebnisse gemeldet werden.

- Zusätzliche Daten hätten potenziell auch aus einer Selbsterhebung der Bürgerinnen und Bürger kommen können. Diesbezüglich waren viele Hoffnungen auf die Corona-App gerichtet, die diese aber letztendlich nicht erfüllen konnte.
- Kein Geringerer als Karl Lauterbach hat vor Kurzem noch einmal daran erinnert, wie wertvoll es gewesen wäre, relativ früh während der Pandemie Kohortenstudien zu starten. Idealerweise wäre im Zuge dessen eine Zufallsstichprobe aus der deutschen Bevölkerung mit vielleicht 50.000 oder 100.000 Teilnehmern dazu verpflichtet worden, sich regelmäßig testen zu lassen. Dadurch hätten die Inzidenzwerte, die aus den Meldungen an das RKI entstehen, einfach präzisiert bzw. plausibilisiert werden können. Und es hätte dem aktuell kontrovers diskutierten Umstand Rechnung getragen, dass wir in Deutschland keine Impfreister haben, dass also nirgendwo zentral die Information hinterlegt wird, wer wann mit welchem Impfstoff geimpft wurde.

Auch in der klinischen Forschung (wo ich zugegebenermaßen kein Experte bin) sieht man, dass es schon relativ früh in der Pandemie eine ganze Reihe von Wirksamkeitsstudien gab, diese aber sehr unkoordiniert abliefen. Insbesondere gab es keine systematischen Ansätze, die Studien zu registrieren und zu koordinieren, oder die Daten, die in diesen Studien entstanden, auf sinnvolle Art und Weise zusammenzubringen.

Ein weiterer Grund für die aktuelle Datenkrise ist die Schwierigkeit, aus der Krankenversorgung Daten zu sichern und für die Forschung im Allgemeinen sowie für die Pandemie-Bewertung im Besonderen nutzbar zu machen. Das gilt für die Krankenhaus-IT ebenso wie für die Praxis-IT. Zur Praxis-IT kann ich wenig sagen, aber die Krankenhaus-IT ist insbesondere in den Universitätskrankenhäusern in Deutschland sehr heterogen und wenig darauf ausgelegt, daraus systematisch strukturierte Daten zu gewinnen. Daher haben wir im Moment immer noch große Probleme, Versorgungsdaten von COVID-19-Patienten für die Forschung zu verwenden. Alternativ ließen sich solche Daten auch von den Versicherern bekommen, was aber unter anderem wegen Fragen des Datenschutzes und des intellektuellen Eigentums problematisch, wenn nicht gar unmöglich sein dürfte.

## **Epidemiologie, klinische Forschung und Krankenversorgung**

Ich möchte nun die Bereiche Epidemiologie, klinische Forschung und Krankenversorgung etwas näher beleuchten, um dann am Ende noch einmal darauf einzugehen, wie im Bereich der Krankenversorgung dem Problem der mangelnden Verfügbarkeit forschungstauglicher Daten begegnet werden kann.

In der Epidemiologie gab es lange Zeit eine Auseinandersetzung darüber, wem die Politik mehr vertrauen sollte, der Eminenz oder der Evidenz. In seinem Porträt über die

Physikerin Viola Priesemann für die Wochenzeitung *Die Zeit* stellte Malte Henk die Frage, ob sie die Lösung (für die Pandemie-Problematik) hätte.<sup>11</sup> Eine Zeit lang hat man sehr darauf gehofft, dass die theoretischen Modellierer mit ihren Differenzialgleichungssystemen in der Lage wären, das Pandemie-Geschehen verlässlich zu beurteilen und vorherzusagen. In Teilen mag diese Hoffnung berechtigt gewesen sein, und ich zitiere dazu aus unserem bereits erwähnten Artikel im *Deutschen Ärzteblatt*, dass die Kompetenz der Theoretiker sehr hilfreich für die wissenschaftsgeleitete Politik ist.<sup>12</sup> Auf der anderen Seite kann aber der Rat von Experten nur dann sinnvoll sein, wenn er sich auf aussagekräftige Daten stützt. Kein noch so ausgefeiltes statistisches Modell kann eine unzureichende Datenlage kompensieren. Im Moment stellt der rapide Rückgang der Inzidenzwerte die Theoretiker vor große Fragen, und keiner von ihnen kann erklären, weshalb dieser Rückgang so rapide erfolgt. Alle Modelle hatten vorausgesagt, dass es eher ein langsames ›Phasing-Out‹ geben sollte.

Der amerikanische Mathematiker Xiao-Li Meng hat bei der Jahrestagung der Deutschen Gesellschaft für Epidemiologie 2020 einen sehr interessanten Vortrag gehalten, in dem er das Problem der Unsicherheit von Inzidenz-Schätzungen eingehend thematisierte.<sup>13</sup> Es kommt demzufolge zu systematischen Verfälschungen solcher Schätzungen, wenn sie aus Stichproben stammen, bei denen die Wahrscheinlichkeit einer Testung davon abhängt, ob die betreffende Person bereits infiziert ist oder nicht. Im Idealfall würde man eine Zufallsstichprobe ziehen, bei der es diesen Zusammenhang nicht gibt. Sobald aber eine positive Korrelation zwischen Testung und Infektion besteht, können selbst große Stichproben ihre Aussagekraft verlieren. Insbesondere wenn gezielt Personen getestet werden, die mit hoher Wahrscheinlichkeit infiziert sind, treibt dies die Schätzung der Inzidenz nach oben. Da sich Teststrategien aber oftmals gerade daran orientieren, möglichst effizient zu sein, wurden und werden Corona-Tests vordringlich im Umfeld von erkrankten oder bereits als infiziert erkannten Personen durchgeführt, was wiederum die Infektionswahrscheinlichkeit der Getesteten nach oben treibt und trieb.

Wo soll die Information über solche sogenannte ›Störgrößen‹ herkommen, die das Testgeschehen und damit die Schätzung der Inzidenzen beeinflussen? Eine Möglichkeit wäre die Selbstauskunft der Betroffenen, wann und aus welchem Grund sie getestet wurden und mit welchem Ergebnis. Dies wiederum hätte eine Aufgabe der Corona-App sein können, für die es allerdings nur einen relativ geringen Nutzungsanreiz gab. Die Freiwilligkeit der mit der App möglichen Infektionsmeldungen hat dazu geführt, dass deren Effizienz sehr gering war. Die Warn-Rate, d.h. der Anteil der tatsächlich Infizierten, die über die App eine entsprechende Meldung herausgegeben haben, lag schätzungsweise nur bei etwa 11 Prozent. Weil die App mit anonymen Daten arbeitet, hat sie zudem

---

11 Malte Henk: Viola Priesemann: [Hat sie die Lösung?](#), in: *Die Zeit* 53 (17.12.2020), S. 15-17. Siehe auch den Beitrag *Ausbreitung von COVID-19 und Strategien der Eindämmung* von Viola Priesemann im vorliegenden Band: <https://dx.doi.org/10.38072/978-3-928794-82-4/p7>.

12 Caliebe et al., Dateninfrastruktur (wie Anm. 9), S. A501.

13 Hierzu ausführlich Xiao-Li Meng: Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election, in: *The Annals of Applied Statistics* 12,2 (2018), S. 685-726. <https://doi.org/10.1214/18-AOAS1161SF>.



keine direkte Entlastung der Gesundheitsämter mit sich gebracht. Für die Epidemiologie standen essentielle Informationen, die sich mit der App hätten gewinnen lassen, nicht zur Verfügung, weil z.B. anonym kein Tracking von Infizierten (wer hat sich wann wo bewegt) möglich war. Dies veranlasste den Kontroversen nicht scheuenden Oberbürgermeister der Stadt Tübingen, Boris Palmer, zu der Aussage, die Corona-App schütze zwar unsere Daten, uns aber nicht vor dem Virus.<sup>14</sup>

Welche Bedeutung es hat, wenn die exakten Umstände von Infektionen mit dem Coronavirus außer Acht gelassen werden, illustriert sehr eindrücklich das Beispiel der Stadt Kiel. Als die Verantwortlichen sich dort erstmals entschlossen, die gemeldeten Infektionen auf das Stadtgebiet zu ›mappen‹,<sup>15</sup> wurde die regionale Unterschiedlichkeit des Infektionsgeschehens deutlich.<sup>16</sup> Die Stadt Kiel hat daraufhin gezielt Maßnahmen an den Brennpunkten der Infektion ergriffen, um die Verbreitung des Coronavirus einzudämmen, z.B. durch den Einsatz von Kräften, die mit den Personen vor Ort intensiv über das Problem gesprochen haben.

Meine Betonung der Wichtigkeit aussagekräftiger Daten soll nicht die Bedeutung epidemiologischer Modelle schmälern. Vielmehr treffen hier zwei Philosophien aufeinander, theoretische Modelle einerseits, empirische Daten andererseits, die aber beide das gleiche Ziel verfolgen. Diese Dualität wurde eindrücklich in einem Blog des *Boston Review* dargestellt, in dem zwei ausgewiesene Epidemiologen aus beiden Lagern zu Wort kamen:<sup>17</sup> Marc Lipsitch von der Harvard Universität in Boston und John Ioannidis aus Stanford. Lipsitch, der eher im Lager der theoretischen Modelle zu verorten ist, erteilte dabei starkem Skeptizismus gegenüber modellbasierten Vorhersagen eine Absage. Wissenschaftler könnten nicht darauf beharren, dass ohne empirische Evidenz keine Aussagen getroffen werden könnten. Sie sollten stattdessen dafür sorgen, dass alle Informationen, die zur Lösung eines Problems beitragen können, auch tatsächlich Berücksichtigung finden.<sup>18</sup>

Ioannidis, der als Protagonist einer eher datengetriebenen Herangehensweise gilt, hat wiederum eingeräumt, dass alles halbwegs Vernünftige und dem gesunden Menschenver-

---

14 Boris Palmer: Corona-Kontaktverfolgung: Weniger Bedenken, mehr Taiwan. Wie wir die Technologie nutzen müssen, um Corona-Infektionen wirklich nachverfolgen zu können, in: Die Zeit 1 (2021). <https://www.zeit.de/2021/01/corona-kontaktverfolgung-taiwan-suedkorea-app-datentechnologie> (letzter Zugriff: 29.3.2022).

15 Steffen Müller: Das sind die Infektionszahlen in den Kieler Stadtteilen, in: Kieler Nachrichten (3.4.2021). <https://www.kn-online.de/Kiel/Coronavirus-in-Kiel-Das-sind-die-Infektionszahlen-in-den-Stadtteilen> (letzter Zugriff: 29.3.2022).

16 Jonas Bickel: Maskenpflicht am Vinetaplatz in Kiel wird oft ignoriert, in: Kieler Nachrichten (6.4.2021). <https://www.kn-online.de/Kiel/Corona-in-Gaarden-Maskenpflicht-am-Vinetaplatz-in-Kiel-wird-oft-ignoriert> (letzter Zugriff: 29.3.2022).

17 Marc Lipsitch: Good Science is Good Science. For the sake of both science and action in the COVID-19 pandemic, we need collaboration among specialists, not sects. Mai 2020. <https://bostonreview.net/articles/marc-lipsitch-good-science-good-science/> (letzter Zugriff: 29.3.2022); John P. A. Ioannidis: The Totality of the Evidence. As policymakers debate the right response to COVID-19, they must take seriously the harms of pandemic policies, not just their benefits. Mai 2020. <https://bostonreview.net/articles/john-p-ioannidis-totality-evidence/> (letzter Zugriff: 29.3.2022).

18 Lipsitch, Good Science is Good Science (wie Anm. 17).

stand Folgende getan werden muss, solange es keine hochqualitativen Daten gibt. Sobald aber aussagekräftige Daten zur Verfügung stünden, müssten sich alle Handlungen und Maßnahmen schnellstens auf diese stützen.<sup>19</sup>

Beide Positionen wurden in besagtem Blog von Jonathan Fuller, einem Philosophen aus Pittsburgh, wunderbar in Einklang gebracht.<sup>20</sup> Ihm zufolge sollte die Epidemiologie beides nutzen: die Theorie mit ihren Modellen ebenso wie konkrete Daten. In der Art und Weise, wie Wissen und Informationen gesammelt werden, kann die Epidemiologie liberal sein, bei den Schlussfolgerungen aus diesen Quellen muss sie aber konservativ zu Werke gehen: Mit einer Hand Daten sammeln, mit der anderen Hand nach diesen Daten handeln. Diese Einsicht, Schaden und Nutzen von Handlungsempfehlungen mit der gleichen Seriosität abzuwägen, wie sie bei der Generierung der zugrundeliegenden Evidenz angewendet werden muss, sollte meines Erachtens auch der Politik als Richtschnur dienen.

Zur klinischen Forschung zwei kurze Anmerkungen: Betrachtet man die weltweit aktiven klinischen Studien zu COVID-19 während der zweiten starken Welle,<sup>21</sup> so fällt eine Konzentration von Studien an hospitalisierten Patienten auf, die im Wesentlichen in Kollektiven von etwa 50 bis 100 Probanden durchgeführt wurden. Es gab also offensichtlich viele kleine Studien, die wenig koordiniert voneinander abliefen und überwiegend in Krankenhäusern stattfanden. Über 80 Prozent der Studien waren nur zweiarmig, es wurde also genau ein Wirkstoff gegen ein Placebo getestet; bei 15 Prozent aller Studien handelte es sich dabei übrigens um Chloroquin. Hier hätte die klinische Forschung zu COVID-19 erheblich davon profitieren können, wenn es eine Möglichkeit zur systematischen Registrierung und inhaltlichen Koordination der Studien gegeben hätte, und wenn es möglich gewesen wäre, diese Daten hinterher zusammenzuführen und zu teilen.

Das *Review* von Alison Dillman et al. aus dem Juni 2020 berichtete von einer empirischen Untersuchung der Neigung zum ›Data-Sharing‹ und der dabei verfolgten ›Policy‹ in der klinischen Corona-Forschung.<sup>22</sup> Von den 35 im Sommer 2020 zu COVID-19 publizierten Studien (mit relativ kleinen Fallzahlen) waren aber nur etwa ein Drittel mit einem Studienregister verlinkt. Zwar hatten sich mehr als zwei Drittel der Studienleiter zur Teilung der Daten bereiterklärt, in fast allen Fällen allerdings nur »auf Anfrage«. Hier fehlte und fehlt nach meinem Kenntnisstand immer noch die Möglichkeit, das klinische Forschungsgeschehen systematisch zu koordinieren und die erzeugten Daten für eine Nachnutzung zusammenzutragen.

---

19 Ioannidis, *The Totality of the Evidence* (wie Anm. 17).

20 Jonathan Fuller: *Models vs. Evidence. COVID-19 has revealed a contest between two competing philosophies of scientific knowledge. To manage the crisis, we must draw on both.* Mai 2020. <https://bostonreview.net/articles/jonathan-fuller-models-v-evidence/> (letzter Zugriff: 29.3.2022).

21 Alison Dillman et al.: *The Landscape of Emerging Randomized Clinical Trial Evidence for COVID-19 Disease Stages: A Systematic Review of Global Trial Registries*, in: *Infect Drug Resist* 13 (2020), S. 4577-4587. <https://doi.org/10.2147/IDR.S288399>.

22 Alison Dillman et al.: *Reporting and design of randomized controlled trials for COVID-19: A systematic review*, in: *Contemporary Clinical Trials* 101 (2020). DOI: <https://doi.org/10.1016/j.cct.2020.106239>.

Damit komme ich zum letzten Teilbereich, den momentanen Möglichkeiten, Daten aus der Versorgung von COVID-19-Patienten in Kliniken und Praxen in geeigneter Art und Weise für die Forschung nutzbar zu machen. Dieses Thema ist nicht nur für COVID-19 relevant. Bezüglich der Erschließung von Versorgungsdaten gab und gibt es ein Leben jenseits von Corona, wie im Cartoon *Keeping the Eye on the Ball* von Charlie Padgett glänzend dargestellt.<sup>23</sup> Auf der rechten Seite des Cartoons sind all jene Gesundheitskrisen zu sehen, die uns bereits vor der Pandemie beschäftigt haben und die uns mit Sicherheit auch nach der Pandemie weiter beschäftigen werden. Insofern habe ich große Hoffnung, dass die Coronapandemie als Katalysator wirkt, um die Probleme bei der Erschließung medizinischer Versorgungsdaten für Forschungszwecke zu lösen. Im Moment ist diesbezüglich tatsächlich Licht am Horizont zu sehen.

Dafür möchte ich noch kurz auf die Arbeit der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. eingehen, die in diesem Bereich schon seit über 20 Jahren eine entscheidende Rolle spielt.<sup>24</sup> Die TMF ist die Dachorganisation der medizinischen Verbundforschung in Deutschland für organisatorische, rechtlich-ethische und technologische Belange. Sie wurde 1999 vom Bundesministerium für Bildung und Forschung (BMBF) im Zusammenhang mit der Etablierung der Kompetenznetze in der medizinischen Verbundforschung gegründet. Das BMBF hatte damals erkannt, dass es in den Bereichen IT, Datenschutz, Recht und Ethik Probleme gibt, die unabhängig vom jeweils beforschten Krankheitsbild immer wieder auftauchen, und dass es sinnvoll wäre, eine Plattform zu schaffen, wo diese Probleme im interdisziplinären Austausch und standortübergreifend gemeinsam bearbeitet werden.

## Diskussion

**Frage: Vielen Dank, lieber Herr Krawczak. Das sind die Worte eines Mahners, der sich immer für Datenqualität eingesetzt hat. Das Thema wirkt, wie so oft, etwas trocken, aber es ist viel Musik darin. Ein Zuschauer oder eine Zuschauerin fragt zum Thema Modelle: Welches Risiko schätzen Sie höher ein, das Risiko einer fehlerhaften Schlussfolgerung und damit einer fehlerhaft ergriffenen Maßnahme auf der Grundlage einer rein empirisch ermittelten Datenkorrelation ohne Absicherung des statistischen Ergebnisses durch ein dahinter liegendes Wissensmodell oder das Risiko, trotz einer eventuell nur unsicheren statistischen Evidenz zur Gefahrenabwehr wegen nur dünner Korrelation politisch nicht gehandelt zu haben? Unter welchem Unsicherheitsmaß**

---

23 Abbildung veröffentlicht in Eleftheria Zeggini et al.: Biomedical Research Goes Viral: Dangers and Opportunities, in: *Cell* 181,6 (2020), S. 1189-1193, Fig. 1. <https://doi.org/10.1016/j.cell.2020.05.014>.

24 Webseite der Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.: <https://www.tmf-ev.de/> (letzter Zugriff: 10.1.2022).

**sollte man in der Epidemiologie schon aufgrund eines rein empirisch ermittelten Erklärungsmodells politisch handeln, wenn es noch kein wissenschaftliches Erklärungsmodell gibt?**

**Antwort:** Die Frage geht in die Richtung des Blogs des *Boston Review*, von dem ich anfangs berichtet habe: Alle Beteiligten hatten sich dort dafür ausgesprochen, dass man zu jedem Zeitpunkt das Bestmögliche tun sollte. Solange noch keine verlässlichen und hinreichenden empirischen Daten vorliegen, ist es unumgänglich, sich auf die Vorhersagen von Modellen zu stützen, so sie denn wissenschaftlich fundiert sind. Das ist allemal besser, als sich nur auf das Bauchgefühl zu verlassen. Sobald aber hinreichende Daten zur Verfügung stehen, müssen theoretische Vorhersagen untermauert bzw. abgelöst werden durch Maßnahmen, die sich auf empirische Erkenntnisse gründen. Mein Vortrag sollte darauf aufmerksam machen, dass solche Daten zur Verfügung stehen müssen und der Mangel daran ein Kernproblem der Pandemie gewesen ist. Wir haben es versäumt, den Aufbau von Infrastrukturen zur Datenerhebung und Datennutzung in geeigneter Art und Weise voranzutreiben. Ich kann mich sehr gut daran erinnern, dass die Medizininformatik-Initiative (MII)<sup>25</sup> vor der Pandemie ziemlich auf der Stelle getreten ist und nicht die Wahrnehmung erhalten hat, die sie verdiente. Das hat sich durch die Pandemie geändert, allerdings immer noch fokussiert auf das Problem COVID-19. Aus meiner Sicht wird es einen ›Spin-off‹ geben müssen, der der Frage nach der Dateninfrastruktur ein größeres Gewicht verleiht. Um in aller Kürze auf die Frage zu antworten: Handeln muss man immer. Politik muss Entscheidungen treffen und ich glaube, sie hat das in Deutschland ganz gut gemacht. Allerdings hätte es aus meiner Sicht der Wahrnehmung und der Akzeptanz der Entscheidungen in der Bevölkerung gutgetan, wenn der dahintersteckende wissenschaftliche Input transparenter gewesen wäre. Vielleicht hätte man von Seiten des Bundesministeriums für Gesundheit ein monatliches wissenschaftliches Symposium für den offenen wissenschaftlichen Dialog zu COVID-19 veranstalten sollen. Das hätte den Eindruck verhindert, dass viele wissenschaftliche Empfehlungen zumindest in der Anfangszeit ohne vorherigen Diskurs abgegeben wurden und ohne Transparenz darüber, welche Wissenschaftsdisziplin welchen Input geleistet hat. Insbesondere Epidemiologen sind in der ersten Zeit der Pandemie kaum in Erscheinung getreten. Es waren im Wesentlichen Virologen und Physiker, die wahrnehmbar die Politik beraten haben.

**Frage: Sie haben mit einem leicht ironischen Unterton von deutscher Kleinstaaterei bei Initiativen wie dem Netzwerk Universitätsmedizin (NUM)<sup>26</sup> gesprochen. Sie haben außerdem – und mein Eindruck ist das an dieser Stelle auch – gesagt, dass wir aufgrund unserer spezifisch deutschen Historie eine Datenschutzwahrnehmung haben, die sehr intensiv und extensiv ist - die so intensiv in der Öffentlichkeit debattiert wird, dass da**

---

25 Webseite der MII: <https://www.medizininformatik-initiative.de> (letzter Zugriff: 29.3.2022).

26 Webseite des NUM: <https://www.netzwerk-universitaetsmedizin.de/> (letzter Zugriff: 29.3.2022).

**teilweise der Gesundheitsschutz geringer bewertet wird als der Datenschutz. Sehen Sie das auch so, oder würden Sie das nicht so gegeneinander ausspielen, wie ich es tue?**

**Antwort:** Ich finde es sehr wichtig, dass man hier nichts gegeneinander ausspielt. Datenschutz ist ein legitimes Interesse aller Bürger, und wir alle wollen, dass insbesondere sensible medizinische Daten geeignet geschützt werden. Mir fehlt jedoch oft ein risiko-adaptierter Umgang mit dem Thema Datenschutz: Welche Risiken entstehen denn tatsächlich dadurch, dass z.B. im Universitätsklinikum Schleswig-Holstein (UKSH) hinter dessen Firewall entsprechende Datenverarbeitung stattfindet und Patientendaten, die sowieso in den Systemen des UKSH vorliegen, in neuer Form für die Forschung zur Verfügung gestellt werden? Wo liegen die Risiken, wenn Experten aus ganz Deutschland in der MII zusammenarbeiten, um geeignete Techniken der Pseudonymisierung zu entwickeln und Treuhändlerlösungen auf den Weg zu bringen, damit Datenschutzrisiken minimiert werden? Wir leben in einer Zeit, in der viele Menschen bei Facebook, Google und Amazon Unmengen an persönlichen Daten bedenkenlos offenbaren, gleichzeitig aber teilweise emotionale Diskussionen im Zusammenhang mit der medizinischen Forschung geführt werden, obwohl es keinen bisher dokumentierten Fall des Missbrauchs medizinischer Forschungsdaten zum Nachteil einer einzelnen Person gegeben hat. Das muss sich meiner Meinung nach ändern, weil man dem Datenschutz ansonsten keinen großen Gefallen tut. In der Wahrnehmung führt der zähe Umgang mit dem Datenschutz fast reflexartig zu der Aussage, Datenschutz blockiere alles. Der Datenschutz muss aber nichts blockieren, wenn er sich an den tatsächlichen Risiken ausrichtet.

**Frage: Probleme ›schnell mal erhobener‹ Daten oder systematisch falsch erhobener Daten, Qualitätsfragen und eine unklare Datenprovenienz sind das eine, Denken ist das andere. Stecken wir nicht neben der Datenkrise in einer Modellkrise? Zum Beispiel sind Parameterräume zu orthogonalisieren oder zumindest zu normalisieren, vielleicht klug zu vergrößern oder eine Principal Component Analysis (PCA) durchführen. Aber sind Parameterräume oder Indikatoren überhaupt der richtige Ansatz, um sich auf die Hauptparameter zu konzentrieren und gleichzeitig orthogonalisierte Räume betrachten zu können?**

**Antwort:** Die Frage zielt vermutlich auf Verfahren, die eingesetzt werden, um hochdimensionale Daten in ihrer Dimensionalität zu reduzieren und so zu erlauben, sich auf die wesentliche Information in den Daten zu konzentrieren. Wir widmen uns diesem Thema im Moment in Kiel gerade bei der Zwischenauswertung der Daten aus dem COVIDOM-Projekt<sup>27</sup> im NUM. Da werden Dimensionsreduktionsverfahren eine große Rolle spielen. Allerdings werden wir auch dort mit der Datenqualität leben müssen, die uns die Versorgungssysteme zur Verfügung stellen. Viele dieser Daten wurden primär für Versorgungszwecke erhoben und von Seiten der Forschung gibt es wenig Möglichkeiten, an der Qualität etwas zu ändern. Datenerzeugung in der Versorgung orientiert

---

<sup>27</sup> Webseite der Studie: <https://www.covidom.de/> (letzter Zugriff: 29.3.2022).

sich primär an zwei Dingen: Der Patient muss wieder gesund werden und die Finanzierung der Versorgung muss sichergestellt sein. Die Forschung kann also nicht immer frei bestimmen, welche Daten sie in welcher Form in der Medizin nutzt – deswegen heißt das Ganze ja auch ›Sekundärnutzung‹. Anders sieht es aus, wenn man in prospektiven Studien geplant Studiendaten erhebt. Da ist man als Wissenschaftler relativ frei in der Gestaltung der Qualitätssicherung. Im Rahmen einer Pandemie lassen sich aber die allermeisten wichtigen wissenschaftlichen Fragestellungen nicht mit prospektiven Studien beantworten. Wir sind stattdessen auf Daten angewiesen, die wir von anderen erhalten, bis hin zu den von Patienten selbst mit Hilfe von Apps erhobenen Daten oder den Daten aus der Verwaltung. Damit gilt es umzugehen und unser Methodenwissen einzusetzen, um auch aus diesen Daten die bestmögliche Evidenz zu generieren.

**Frage: Sie sind Mitglied bei der COVIDOM-Gruppe, mögen Sie kurz etwas dazu sagen, worum es bei diesen Langzeituntersuchungen mehrerer Universitätskliniken an COVID-19-Patienten geht?**

**Antwort:** COVIDOM war ursprünglich ein schleswig-holsteinisches Projekt und verfolgte die Idee, in einem bestimmten geografischen Bereich möglichst alle auf SARS-CoV-2 positiv getesteten Personen in ein Langzeit-Follow-up-Programm einzuschließen. Ausgangspunkt waren die relativ frühen Anzeichen, dass selbst Personen, die in der Akutphase der Infektion keine großen gesundheitlichen Probleme bekamen, möglicherweise Langzeitschäden zu befürchten haben. Dabei waren auch andere Organsysteme als der Atemstrakt zunehmend in den Fokus gerückt. Deswegen haben wir in Kiel früh das COVIDOM-Projekt angeschoben, um in Kooperation mit den Gesundheitsämtern infizierte Personen anzuschreiben und sie zu bitten, sich in unser Langzeitprogramm einzubringen. Als dann das NUM Fahrt aufgenommen hatte und das NAPKON (Nationales Kohorten Netzwerk)<sup>28</sup> entstand, da war klar, dass wir uns von Kiel aus daran beteiligen müssen. Damit ist das COVIDOM-Projekt praktisch zum populationsbasierten Teil des NAPKON geworden. Maßgeblich beteiligt sind daran außer Kiel noch Würzburg und Berlin. Der gemeinsame Aufbau einer Kohorte von Infizierten unabhängig vom ursprünglichen Erkrankungszustand läuft nun unter dem Label ›NAPKON-POP‹ im Rahmen des NUM. Die Kollegen hier in Kiel arbeiten aber genauso weiter, als wenn es sich nach wie vor um die schleswig-holsteinische Studie handeln würde. Wir können mit Stolz sagen, dass wir bislang mehr als 80 Prozent der Daten zum bundesweiten Projekt beigetragen haben.

**Frage: Und dies trotz relativ niedriger Fallzahlen und einer geringen Bevölkerungsdichte.**

**Antwort:** Ja, aber es ist uns mit Unterstützung des Gesundheitsministeriums in Schleswig-Holstein gelungen, über die Gesundheitsämter eine sehr hohe Rekrutierungszahl zu

---

<sup>28</sup> Webseite von NAPKON: <https://napkon.de/> (letzter Zugriff: 1.4.2022).

erreichen. Auch im Zusammenhang mit dem Kieler Biobank-Projekt popgen<sup>29</sup> konnten wir erleben, dass die Bereitschaft der Bürger zur Teilnahme an solchen Studien in unserem Bundesland sehr hoch ist. Wir haben bei den ersten popgen-Studien Rekrutierungszahlen von 60 bis 70 Prozent der Angeschriebenen erreicht. Das ist sehr viel. Es scheint ein besonderes Kennzeichen von Schleswig-Holstein zu sein, dass die Bereitschaft der Bevölkerung zum Mitmachen bei solchen Projekten besonders stark ist.

**Frage: Eine weitere Frage: Es wurde eingangs erläutert, dass die Qualität beziehungsweise Validität unseres Wissens maßgeblich von der Qualität der Daten an der Basis der sogenannten ›Wissenspyramide‹ abhängt. Das haben Sie anhand einer Grafik dargestellt. Wie beurteilen Sie die Gleichmäßigkeit der länderspezifisch gelieferten Daten etwa aus Europa, China, USA, Asien etc.? Und wie können bekannte Unterschiede in der Datenqualität im Prozess der Wissensgenerierung gegebenenfalls ausgeglichen werden?**

**Antwort:** Eine relativ einfache Möglichkeit, solchen Unterschieden Rechnung zu tragen, ist das in der Statistik als ›Stratifizierung‹ bezeichnete Verfahren. Dabei wird die statistische Auswertung der Daten nach bestimmten Kriterien wie z.B. der Quelle der Daten stratifiziert. Einfach gesprochen wird die Analyse mit deutschen, italienischen und brasilianischen Daten getrennt durchgeführt und dann geprüft, ob man im Wesentlichen zu den gleichen Ergebnissen kommt. Eine weitere Möglichkeit ist die Einigung der Datenerzeuger auf Standards, die dafür sorgen, dass Daten in einer gemeinsamen Sprache beschrieben werden und die gleiche Qualität haben. Ich spreche zunächst vom Ausgleich über die Institutionen in Deutschland hinweg, die solche Daten beitragen: Wie werden die Daten erhoben, wie werden sie validiert, werden sie kontrolliert, wie werden sie eingegeben? Das ist übrigens ein wesentlich wirksameres und nachhaltigeres Werkzeug, um Datenqualität und -homogenität zu garantieren, als die erwähnten statistischen Verfahren. Wo ein Ausgleich organisatorisch aber nicht möglich ist, z.B. bei der Zusammenführung von Daten aus Kontinenten mit unterschiedliche Standards, wird an einer statistischen Korrektur kein Weg vorbeiführen.

**Frage: Es ist schon von erheblichem Interesse, ob z.B. der schwedische Weg durch die Pandemie der günstigere oder weniger günstige gewesen ist. Insofern sind Datenzusammenfügungen und -vergleiche immer mit der Frage verknüpft, ob wir es hier mit demselben Sachverhalt oder mit Äpfeln und Birnen zu tun haben.**

**Antwort:** Wenn wir von Schweden reden, würde ich davon ausgehen, dass dort nach den gleichen Standards und mit den gleichen Qualitätskriterien und Qualitätssicherungsmaßnahmen gearbeitet wurde wie in Deutschland. Insbesondere wenn die Datenerhebung *lege artis* erfolgte und wichtige Begleitdaten, die sogenannten ›Metadaten‹, in

---

<sup>29</sup> Webseite von popgen: <https://www.epidemiologie.uni-kiel.de/biobanking/biobank-popgen> (letzter Zugriff: 29.3.2022).

geeigneter Art und Weise miterfasst wurden, kann man die Vergleichbarkeit gut nachträglich beurteilen. Außerdem gibt es Indikatoren für die Qualität von Daten, die man dann nutzen kann, um möglicherweise erkennbare Unterschiede auszugleichen.

**Frage: Ein Kommentar und eine Frage zum Datenschutz: Die von Ihnen angesprochene Datenspende ist nach gegenwärtigem Kenntnisstand in der jetzigen Fassung der Datenschutz-Grundverordnung (DSGVO) nicht vorgesehen, weil es sich hier um höchst persönliche Daten handelt. Den möglichen großen Vertrauensverlust sollte man nicht unterschätzen. Es hat schon MRT-Bilder im Internet gegeben. Wäre nicht die vollständige Anonymisierung hier ein Ausweg? Vollständig anonyme Daten sind nicht von der DSGVO erfasst. Oder kommt es auf die Rückverfolgbarkeit zum Patienten an?**

**Antwort:** Vielleicht die letzte Frage zuerst: Die Rückverfolgbarkeit bereitet uns in Sachen Datenschutz das größte Kopfzerbrechen bei der Forschung. Wir können nicht mehr auf eine vollständige Anonymisierung setzen, weil wir für viele wissenschaftliche Fragestellungen Langzeitverläufe brauchen. Das hat übrigens nichts mit COVID-19 zu tun. Die gesamte Forschung zur Prävention von Krankheiten ist darauf angewiesen, dass man relativ früh (bestenfalls noch vor einer Erkrankung) erhobene Daten hinterher mit dem akuten Erkrankungszustand in Verbindung bringt. Das geht nicht mit vollständig anonymisierten Daten, sondern nur, wenn bereits bestehende Datenbestände personenbezogen ergänzt und fortgeschrieben werden können. Zu der ersten Frage: Ich hatte bereits erwähnt, dass der Kieler Jurist Sebastian Graf von Kielmansegg einen sehr umfangreichen Teil zum Datenspende-Gutachten für das Bundesministerium für Gesundheit beigetragen hat. Er machte darin die Aussage, dass die Datenspende zwar nicht explizit in der DSGVO vorgesehen, wohl aber mit ihr vereinbar ist, weil sie im Kern eben forschungsfreundlich ist. Die Frage ist, wie sie in nationales Recht umgesetzt wird. Man merkt der DSGVO und den Erwägungsgründen zu den einschlägigen Artikeln an, dass hier wirklich der Wunsch bestand, Forschung und insbesondere die medizinische Forschung zu privilegieren gegenüber anderen Formen der Datennutzung, z. B. für soziale Netzwerke, Werbung und so weiter. Das bringt meines Erachtens ein gewisses Grundvertrauen des europäischen Gesetzgebers gegenüber medizinischer Forschung zum Ausdruck. Man muss sich auch in Erinnerung rufen, dass die DSGVO das öffentliche Interesse (nach Abwägung gegen die Risiken der Betroffenen) als gleichwertige Rechtsgrundlage für die Verarbeitung personenbezogener Daten vorsieht – neben der informierten Einwilligung. In Deutschland gewinnt man jedoch immer wieder den Eindruck, es gäbe eine Hierarchisierung dieser beiden Rechtsgrundlagen in dem Sinn, dass die Einwilligung immer vorgeht. Erst wenn die nicht möglich ist, dürfe man sich etwas anderes überlegen. Das ist aber vom Gesetzgeber so nicht intendiert gewesen, und beide mögliche Rechtsgrundlagen, öffentliches Interesse und persönliches Einverständnis, stehen auf einer Stufe. Das hat Graf von Kielmansegg auch vor Kurzem in einem Gutachten für das NUM nochmals zum Ausdruck gebracht vor dem Hintergrund der Frage, ob wichtige Daten aus der frühen Phase der Coronapandemie auch ohne Einwilligung in das Netzwerk integriert und dort wissenschaftlich genutzt werden können.



**Frage: Das heißt also, dass es ein öffentliches Interesse gibt, das auf der gleichen Datenschutzebene wie der individuelle, persönliche Wille steht?**

**Antwort:** Genau! Wenn das öffentliche Interesse an der Nutzung von Daten die Risiken, die den Betroffenen daraus entstehen, stark überwiegt, dann ist das eine hinreichende Rechtsgrundlage für die Verarbeitung der Daten, auch wenn es keine Einwilligung gibt. Ganz wichtig ist dabei jedoch, dass das eine nicht gegen das andere ausgespielt werden kann: Wenn es eine Einwilligung gibt und der Patient seine Ablehnung klar zu verstehen gegeben hat, oder wenn sogar ein expliziter Widerspruch vorliegt, dann kann das öffentliche Interesse das nicht aushebeln. Wenn es aber keine Äußerung des Patientenwillens gibt, dann kann das öffentliche Interesse nach Abwägung der Risiken eine valide Rechtsgrundlage für die Datenverarbeitung sein. Das hat große Konsequenzen für das NUM, weil deswegen in der Tat Daten aus der Frühphase der Pandemie auch ohne Einwilligung wissenschaftlich genutzt werden können. Solche Daten sind sehr wichtig für Fragestellungen, wie ich sie vorhin geschildert habe, z.B. ob es Frühindikatoren für das spätere Gesundheitsrisiko von SARS-CoV-2-Infizierten gibt.

**Frage: Einmal angenommen, es könnten alle Daten ermittelt werden wie gewünscht und aufgrund einer Modellierung kommt man zu dem Schluss einer bestimmten Handlungsempfehlung. Ethisch ließe sich diese Maßnahme aber nicht verantworten. Sollte dann die Empfehlung aus dem objektiven Datenmodell dennoch Vorrang haben? Mich erinnert das Ganze ein wenig an die Diskussion um KI (Künstliche Intelligenz), etwa im Straßenverkehr oder auf sonstigen Gebieten. Sollten wir uns im Zweifel einer Herrschaft der Daten unterwerfen?**

**Antwort:** Ich kann mich gut erinnern, dass hochrangige Politiker (einschließlich Angela Merkel) mehrfach betont haben, unser Land würde nicht von Virologen regiert. Virologen geben Ratschläge aus Sicht der Wissenschaft; die daraus abgeleiteten Entscheidungen treffen die Entscheidungsträger, in diesem Fall die Politik. Daran sollte man nicht rütteln. Wissenschaftler haben kein Interesse an einem Automatismus, mit dem wissenschaftlicher Rat unmittelbar in einer bestimmten Entscheidung der Politik endet. Im Gegenteil: Wissenschaftler haben immer wieder darauf hingewiesen, dass Wissenschaft mit Dissens, mit Uneinigkeit verbunden ist. Sie ist immer Diskussion, sie ist immer Widerstreit. Wir als Wissenschaftler haben gelernt, das auszuhalten. Die Politik aber muss irgendwann entscheiden. Deswegen können Wissenschaftler nur Optionen und unterschiedliche Sichtweisen zu einem bestimmten Problem verständlich machen und transportieren. Das hat meines Erachtens nichts mit Wegducken zu tun. Ich glaube, die Rollenverteilung, so wie ich sie geschildert habe, tut der ganzen Sache gut. Deswegen gibt es auch keine Herrschaft der Daten oder Herrschaft der Wissenschaft.

**Frage: Das ist schon fast ein Schlusswort. Aber noch eine Frage: Würde man angesichts Ihrer Ausführungen auf struktureller Ebene für weniger kompetitive und mehr kooperative Forschung plädieren? Dem steht der gegenwärtige Universitätsbetrieb mit**

**seiner Eigenlogik gewissermaßen entgegen, beispielsweise durch seine Fragmentierung über die Drittmittel.**

**Antwort:** Man muss aus meiner Sicht deutlich zwischen wissenschaftlichem Wettstreit und Infrastrukturaufbau bzw. -pflege unterscheiden. Letztere würden immer davon profitieren, wenn sie kooperativ und nicht kompetitiv ablaufen. Infrastrukturaufbau in Konkurrenz zueinander zu betreiben, ist unsinnig. Das ist quasi die ›DNA‹ der TMF, die gerade dem wiederholten Entstehen unterschiedlicher Lösungen in unterschiedlichen Kontexten entgegenwirken soll. Das Rad muss meistens nicht neu erfunden werden! Davon deutlich trennen muss man die wissenschaftliche Kompetition, den Streit der Ideen und Methoden in der wissenschaftlichen Arbeit. Das ist unser Alltag. Im besten Fall führen diese Auseinandersetzungen irgendwann zu inhaltlicher Konvergenz, wir einigen uns mithin in einer bestimmten Sache und führen den Streit an anderer Stelle fort. Kompetition ist letztendlich der Treibstoff für Wissenschaft. Ohne Kompetition würde die Wissenschaft nicht dort stehen, wo sie heute steht. Aber noch einmal sei gesagt: Wo es um Infrastruktur geht, ist Einigung allemal besser, als im Widerstreit immer wieder Altes neu aufzubauen.

**Frage: Noch zwei Fragen: Die erste Frage betrifft, was Sie am Anfang des Vortrags zum Ursprung des Virus gesagt haben. Da gab es im Grunde immer die Laborhypothese. Und es gab den 400 Seiten starken Bericht der WHO, in dem gesagt wurde, die Laborhypothese sei extrem unwahrscheinlich und müsse verworfen werden. Gibt es neuen Streit über diesen Bericht? Dann gibt es ein paar wissenschaftliche Meldungen, dass in dem Labor in Wuhan unter Sicherheitsstufe-2-Bedingungen Forschung an Coronaviren durchgeführt wurde, die man nur auf einer höheren Sicherheitsstufe hätte machen dürfen. Was Sie gesagt haben, das war mir neu, dass im Grunde dieses Virus nach dem Minenunfall im Labor präsent gewesen ist. Damit gewinnt die Laborhypothese doch wieder an Plausibilität, oder nicht? Werden wir es vielleicht nie mit endgültiger Sicherheit wissen? Da würde ich einfach fragen, ob Sie dem zustimmen können.**

**Antwort:** Dem würde ich zustimmen. Mich macht aber wirklich stutzig, dass das Institut durch die Umbenennung eines Virus offensichtlich versucht hat, die Tatsache zu verschleiern, dass dort ein mit SARS-CoV-2 eng verwandtes Virus bereits 2013 untersucht wurde. Dieser Vorläufer war in Teilen schon sequenziert worden, und die Sequenzen wurden (nach meiner Erinnerung) bereits 2016 veröffentlicht, allerdings unter der alten Identität des Virus. Mit den ersten Proben, die zu Beginn der Pandemie gewonnen wurden, tauchte plötzlich der neue Name auf und das Institut behauptete, 2020 hätten die dortigen Wissenschaftler dieses Virus erstmals in der Hand gehabt. Das stimmt offensichtlich nicht. Vielmehr wurde unter anderem Namen 2020 erstmal ausführlich zum Virus von 2013 publiziert.

**Frage: Meine zweite Frage geht nochmals auf die Datenspende ein. Ich bin sehr für die Datenspende; sie wäre für mich überhaupt kein Problem. Aber darf ich meine Daten spenden, wenn ich nicht genau weiß, für welche Forschung sie dann gebraucht werden? Man muss wahrscheinlich beim ›Informed-Consent‹-Prinzip einfach von der Be-**

dingung ›Consent must be specific‹ abrücken, weil sie nicht erfüllbar ist, wenn man Daten in solche Forschungsprozesse hineingibt. Meine Frage betrifft die Einfachheit der Widerspruchslösung. Eine Widerspruchslösung kann in technischer Hinsicht einfach sein, oder sie kann in psychischer Hinsicht einfach sein. Wir haben in einem anderen Forschungsprojekt mal herausgefunden, dass dies gerade im klinischen Kontext problematisch sein kann: Man ist behandelt worden, die Schwestern waren supernett, die Ärztin war großartig, man ist wieder gesund gemacht worden und soll in diesem klinischen Kontext dann einen Widerspruch einlegen. Das ist einfach psychisch schwer. Man hat irgendwie das Gefühl, das gehört sich nicht. Das ist irgendwie ›unnnett‹ gegenüber dem medizinischen Personal, das einen da so gut behandelt hat. Deswegen genau die Frage: Müsste nicht im Grunde die Einfachheit der Widerspruchslösung neben der technischen Dimension diese psychische Dimension in Rechnung stellen?

**Antwort:** Unbedingt, und deswegen haben wir dafür plädiert, die Entscheidung soweit wie möglich vom Behandlungskontext loszulösen. Zum einen gehört diese Entscheidung nicht in die unmittelbare Aufnahmesituation, weil man nicht vermeiden kann, dass Patienten glauben, ihre Entscheidung hätte einen Einfluss auf ihre weitere Behandlung. Außerdem haben Patienten in dem Moment etwas anderes im Kopf: Sie warten auf eine Behandlung oder eine Diagnose und wollen sich sicherlich nicht mit der Forschungsnutzung ihrer Daten auseinandersetzen. Der zweite Grund ist folgender: Wenn man das Thema tatsächlich für so kompliziert hält, dass man darüber wie in der MII auf zehn Seiten aufklären muss, dann ist es allemal fairer, dass Bürgerinnen und Bürger sich damit in ihrem normalen Leben auseinandersetzen. Zu Hause auf dem Sofa können sie in aller Ruhe überlegen, ob sie es gut finden, dass mit ihren Daten geforscht wird. Wenn nicht, stecken sie im Idealfall beim nächsten Hausarztbesuch oder in der nächsten Apotheke die Versichertenkarte in ein Lesegerät und widersprechen – und dann war es das. Ich glaube, dass man das technisch so niederschwellig hinbekommt. Wenn aber jemand wirklich darüber nachdenkt, was bedeutet die Datenspende an die Forschung denn eigentlich (und da sind wir gefordert, hinreichend Informationen bereitzustellen), dann wird die Mehrheit sagen: »Ja, ich finde es gut, dass mit meinen Daten geforscht wird.« Wir hatten in Veranstaltungen zu diesem Thema immer wieder Patientenvertreter zu Gast, und ich habe nie erlebt, dass ein Patientenvertreter auch nur annähernd gesagt hätte: »Nein, das finde ich nicht gut.« Im Gegenteil, sie betonten stets ihren Wunsch, dass wenigstens ihre Daten etwas Gutes bewirken mit Blick auf die Erforschung ihrer Krankheit und damit der Behandlung zukünftiger Patienten.

**Schlusswort:** Vielen herzlichen Dank für diese sehr klaren Ausführungen. Vielleicht kann man tatsächlich etwas aus der Corona-Erfahrung mitnehmen. Vielleicht betrifft das besonders den Aspekt, den wir heute nochmals vor Augen geführt bekommen haben, nämlich wie wichtig Daten für die Forschung sind. Und vielleicht kommt dadurch ein bisschen Schwung in die Mühe der Datenerhebungen hinein. Dann hätte die Coronapandemie zumindest in dieser Hinsicht doch noch etwas Gutes.

# Autor

**Prof. Dr. Michael Krawczak** 

Institut für Medizinische Informatik und Statistik, Medizinische Fakultät der Christian-Albrechts-Universität zu Kiel, [krawczak@medinfo.uni-kiel.de](mailto:krawczak@medinfo.uni-kiel.de)

Michael Krawczak ist Professor für Medizinische Informatik und Statistik an der Medizinischen Fakultät in Kiel. Er ist Mathematiker und wurde nach dem Diplom 1987 in Entwicklungsbiologie promoviert. Anschließend hat er sich 1994 an der Medizinischen Hochschule Hannover für das Fach Humangenetik habilitiert. Von dort ist er 1996 an die University of Wales in Cardiff gewechselt, wo er als Professor für mathematische Genetik tätig war. Hier hat er eine der größten öffentlichen Datenbanken zu humangenetischen Mutationen mitbegründet und aufgebaut. Seit 2001 ist er Direktor des Instituts für Medizinische Informatik und Statistik in Kiel und hat seine Erfahrung und Expertise beim Aufbau von Daten- und Bioprobenbanken vor Ort vielfältig eingebracht. Er war maßgeblich an der Gründung der Biobank popgen beteiligt, mittlerweile eine der größten populationsbasierten Biobanken in Deutschland. Er hat das PopGen 2.0 Netzwerk verschiedener Biobanken aus der Region maßgeblich mitgestaltet und etabliert, ebenso wie das Kieler Zentrum für klinische Studien. Seit 2006 ist Michael Krawczak Mitglied im Vorstand der Technologie und Methodenplattform für die vernetzte medizinische Forschung e.V., seit 2011 als dessen Vorsitzender.

---

## Open Access

Der Beitrag ist unter der Creative-Commons-Lizenz Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International veröffentlicht. Den Vertragstext finden Sie unter: <https://creativecommons.org/licenses/by-sa/4.0/deed.de>. Bitte beachten Sie, dass einzelne, entsprechend gekennzeichnete Teile des Beitrags von der genannten Lizenz ausgenommen sein bzw. anderen urheberrechtlichen Bedingungen unterliegen können.