



Einsichten eines Wissenschaftsnarren (38)

Politikberatung, bis der Elefant mit dem Rüssel wackelt!

Mathematische Modellierungen sind in diesen Zeiten der Corona-Pandemie sehr gefragt. Allerdings lassen die Vorhersagen aus den entsprechenden Modellen oftmals sehr zu wünschen übrig, weswegen es nachfolgend auch mit deren konkretem Eintreten hapert. Kein Wunder, wenn sie derart auf schlechten oder gar nicht vorhandenen Daten basieren – letztlich also auf bloßen Annahmen.

Das Zeitalter der Universalgelehrten kehrt zurück! Seit etwa einem Jahr eifern Wissenschaftler da Vinci, Leibniz sowie von Humboldt und Co. nach. Virologen äußern sich öffentlich wie auch politischen Entscheidungsträgern gegenüber zur Epidemiologie, Physiker zur Infektionsbiologie, Mathematiker zu viralen Oberflächenproteinen und so weiter. Dabei war es doch bisher die Domäne der Narren, ungestraft Späße zu beliebigen Themen zu machen! Auch deshalb erlaube ich mir heute, mich ungeniert der mathematischen Modellierung in Zeiten der Pandemie zuzuwenden.

»Wie bei Horoskopen passten die vagen Schlussfolgerungen damit zu jedem Verlauf.«

Modellierer sind momentan ja sehr gefragt. Wir lesen ihre Arbeiten in *Nature* und *Science*, man lauscht ihnen bei Markus Lanz und Konsorten, sie beraten Politiker und rechnen für nationale Akademien. Ein Wunder ist das nicht, schließlich versprechen ihre Formeln und Modelle nicht weniger als die Aufklärung komplexer Zusammenhänge. Sie sagen uns, was passieren könnte, wenn wir gewisse Dinge tun oder lassen. Auch erklären sie uns, welche Maßnahmen zur Pandemiebekämpfung wirksam sind – und welche nicht.

Häufig mahnen sie und belegen ihre eigenen Empfehlungen mit konkreten Zahlen.

Genauso wünscht man sich doch Handreichungen aus der Wissenschaft. Die Politik bekommt Argumente für ihre Entscheidungen – und Bürger sehen ein, warum die Schule schließen muss oder das Geschäft die Türe wieder öffnen darf.

Modellierer sind auf vielen Feldern schon länger recht erfolgreich. Ein Paradebeispiel hierfür ist der Wetterbericht. Mit im Mittel etwa siebenzig Prozent Treffsicherheit gelingt es den Meteorologen, das Wetter der nächsten sieben Tagen vorherzusagen. In die Modelle, die auf Supercomputern gerechnet werden, gehen unzählige Messungen ein, die das atmosphärische Geschehen vom Boden bis viele Kilometer in die Höhe abbilden. Ihre Rechnungen berücksichtigen die Temperatur- und Strömungsdynamik der großen Gewässer und sogar die fluktuierenden Bahnen von Mond und Sonne. All dies mit höchster Messgenauigkeit. Möglich wird eine Wettervorhersage mit solcher Treffsicherheit aber nur, weil die meteorologischen Zusammenhänge von verschiedenen Temperaturen und Drücken sowie Wind-, Wasser- und Planetenbewegungen durch internationale wissenschaftliche Kooperationen bereits lange untersucht und mittlerweile recht gut verstanden werden.

Ein anderes schönes Beispiel für erfolgreiche Modellierungen kommt aus der Geophysik. Ausbrüche von Vulkanen lassen sich überraschend gut vorhersagen, wie zuletzt bewiesen beim Fagradalsfjall-Vulkan in Island. Auch diese Vorhersagen beruhen auf einer Vielzahl von exakten seismologischen beziehungsweise Satelliten-Messungen, auf zumindest teilweise verstandenen Mechanismen vulkanischer Aktivität sowie schließlich auf jahrelanger Optimierung der Modelle.

Aber selbst diese Modellierer liegen oft daneben. Dann ärgern wir uns, vor dem Regen nicht gewarnt worden zu sein. Und so mancher Vulkan will trotz eindringlicher Warnungen einfach nicht ausbrechen.

Wie aber steht es angesichts dessen um die Vorhersagekraft und somit um die Nütz-

lichkeit der so allgegenwärtigen Modellierungen in der Pandemie? Leider gibt es mittlerweile eine Menge Hinweise darauf, dass es damit nicht zum Besten steht. Die Modellierer sind offensichtlich so sehr mit dem Generieren neuer Modelle beschäftigt, dass sie kaum dazu kommen, die Güte und das Eintreten ihrer Vorhersagen zu analysieren.

Dies hat man offensichtlich den Journalisten überlassen.

So analysiert etwa ein Artikel in der Tageszeitung *Die Welt* (Literaturzitate wie immer bei <http://dirnagl.com/lj>) die wichtigsten Vorhersagen aus dem Umfeld von Deutschlands prominentester Modelliererin, Viola Priesemann

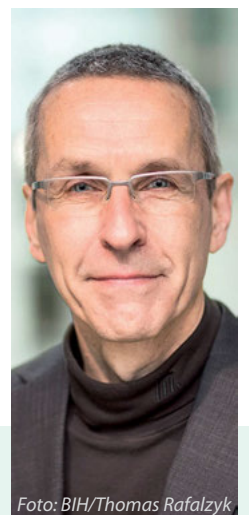


Foto: BIH/Thomas Rafalzyk

Ulrich Dirnagl

leitet die Experimentelle Neurologie an der Berliner Charité und ist Gründungsdirektor des QUEST Center for Transforming Biomedical Research am Berlin Institute of Health. Für seine Kolumne schlüpft er in die Rolle eines „Wissenschaftsnarren“ – um mit Lust und Laune dem Forschungsbetrieb so manche Nase zu drehen.

Sämtliche Folgen der „Einsichten eines Wissenschaftsnarren“ gibt es unter www.laborjournal.de/rubric/narr

(siehe auch LJ 12/2020: 14-17). Dabei zeigt sich zum einen, dass die meisten Schlussfolgerungen aus den Modellrechnungen sehr vage verfasst waren. Wie bei Horoskopen passten sie damit zu jedem Verlauf. Und dort, wo konkrete Zahlen vorhergesagt wurden, sind diese sehr häufig nicht eingetreten. Es sei denn, es handelte sich um Triviales, wie die Vorhersage eines weiteren Anstieges am Anfang eines bereits deutlich sichtbaren Verlaufes.

»Es fehlten Kontrollen, wie wir sie in jeder biomedizinischen Arbeit erwarten würden.«

Sobald es jedoch darum ging, die Wirksamkeit von Pandemiemaßnahmen zu prognostizieren, wurde es richtig problematisch. Nur ein Beispiel hierfür ist die Vorhersage aus der Leopoldina-Stellungnahme vom 8. Dezember letzten Jahres. Dort wurde Folgendes vorausgesagt: „Wenn ab dem 14. Dezember die Maßnahmen streng verschärft werden, dann sinken die Fallzahlen in der Modellrechnung bis Januar auf unter 50 pro 1.000.000 Einwohner.“ Wie wir alle wissen, ist dies trotz erfolgtem hartem Lockdown nicht eingetreten: Die Inzidenzraten stiegen zwar nicht weiter, verharrten aber auf hohem Niveau.

Die zugehörige Modellierung basierte auf dem im Juli 2020 in *Science* veröffentlichten Modell aus dem Max-Planck-Institut für Dynamik und Selbstorganisation in Göttingen – und auf Daten aus dem Frühjahr 2020. Das Modell bezog sich damit auf eine völlig andere Umsetzung und Akzeptanz von Maßnahmen als im Vorhersagezeitraum. Wie vielen solcher Modellierungsstudien fehlten hier aber auch Kontrollen, wie wir sie in jeder biomedizinischen Arbeit erwarten würden. Zum Beispiel hätte man die Güte des Modells durch Anwendung auf anderen Datensätze, zum Beispiel aus einem anderen Land oder über einen anderen – am besten auch längeren – Zeitraum hinweg, überprüfen können.

Moment, Kontrollen beim Modellieren? Ja, das geht – sogar recht einfach. Das Modell, so eine zentrale Aussage des Artikels, würde die Wirksamkeit und damit Notwendigkeit eines harten Lockdowns in Deutschland belegen. Hätten die Autoren ihr Modell aber beispielsweise auch auf Schweden angesetzt, wäre dort ein ganz ähnlich geartetes Absinken der Fallzahlen herausgekommen. Nur dass es dort keinen Lockdown gab! Diese Kontrollrechnung konnte der Neurologe und Physiker Christian Meisel durchführen, da Viola Priesemanns Gruppe ihr Modell inklusive Daten ins Netz stellte (Kudos dafür!). Meisel entwickelt normalerweise Modelle, mit denen sich aus Elektroenzephalogramm-Daten epileptische Anfälle vorhersagen lassen und ist deshalb mit der Technik wohlvertraut.

Ähnliches wie für das Göttinger Modell gilt indes auch für die Modelle des *Imperial College* in London (ICL). Diese hatten großen Einfluss auf die Pandemiemaßnahmen der englischen Regierung. Auch hier lagen die Vorhersagen häufig extrem daneben. Der australische Mathematiker Vincent Chin und andere konnten außerdem zeigen, dass verschiedene publizierte Modelle des ICL zu ganz unterschiedlichen Resultaten kommen, wenn man sie auf die gleichen Länder loslässt. Was die Londoner selbst bezeichnenderweise nicht gemacht hatten.

»Alles hängt davon ab, wie die Maßnahmen in der Bevölkerung tatsächlich umgesetzt werden.«

Ist dies alles überraschend? Deutet es darauf hin, dass die Pandemie-Modellierer ihr Handwerk nicht recht verstehen?

Im Gegensatz zu den Meteorologen basieren ihre Modellierungen auf schlechten oder sogar nicht-vorhandenen Daten, also bloßen Annahmen. Dies gilt sowohl für die Corona-Inzidenzen wie auch viel mehr noch für die Auswirkungen nicht-pharmakologischer

Einfach mal testen!



Foto: Alexander Sidman

LABORJOURNAL
Newsletter

Neuigkeiten
Meinungen
Lustige Zeichnungen
E-Paper
Stellenanzeigen

kommt etwa alle 14 Tage



<https://www.laborjournal.de/aktuell>

Interventionen. Außerdem hängt alles entscheidend davon ab, ob und wie die Maßnahmen in der Bevölkerung dann tatsächlich umgesetzt werden. Bei einer höchst unsicheren Datenlage, wie sie zum Beispiel allein schon durch die sich ständig ändernden Testkapazitäten und -raten, insbesondere am Anfang einer Pandemie, vorkommt, ist es unabdingbar, diese elementare Fehlerbehaftung kritisch zu berücksichtigen.

Datenfehler pflanzen sich fort, das lernt man spätestens im Physik-Praktikum. Und sie tun das umso mehr, wenn sie in komplexe, multiparametrische Modelle und Wachstumsverläufe eingehen. Dazu kommen jede Menge nicht vorhersehbarer Einflussgrößen – wie etwa das Auftreten von Virusmutanten mit veränderter Infektiosität oder Letalität, die Effektivität von Vakzinierungen oder auch unvermeidliche Rückkoppelungs- und Selbstregulierungsmechanismen, weil die Vorhersagen sich ja ihrerseits bereits auf das Verhalten der Bevölkerung auswirken.

»Historisch betrachtet haben Modellierungen von Epidemien keinen guten Track Record.«

In Anbetracht all dessen ist die oft propagierte Pseudogenauigkeit der Modellierungsergebnisse schlichtweg vermessen. Es ist, als würde man mit Kanonen – nämlich komplexen, multiparametrischen Modellierungen – auf Spatzen – also auf grob fehlerbehaftete und nicht-valide Datengrundlagen – schießen.

Ein schönes Beispiel ist hier auch der Rückgang des Autoverkehrs in der Pandemie. In den USA wurden im letzten Jahr rund 13 Prozent weniger Meilen gefahren. Folglich dürfte auch die Zahl der Verkehrstoten abgenommen haben, was demnach einer der wenigen positiven Effekte der Pandemie wäre. Falsch! Die Verkehrstoten haben zugenommen wie seit 1924 nicht mehr, nämlich um 25 Prozent pro gefahrener Meile. Retrospektiv sucht man nun nach Gründen hierfür, wie beispielsweise vermehrter Alkoholkonsum. Doch was hier jetzt viel wichtiger ist: Man hätte diesen überraschenden Effekt wohl kaum vor dessen Bekanntwerden in einem Modell der Gesamtmortalität während der Pandemie berücksichtigen können.

Ein weiterer wichtiger Grund für das Versagen der Modelle ist, dass deren Annahmen ja durch die in der Pandemie angeordneten Maßnahmen modifiziert werden. Dies ist sogar ein erwünschter Effekt, schließlich erheben die Modellierer genau deswegen häufig ihren Zeigefinger. Allerdings wäre das gera-

de so, als wenn sich das Wetter in Abhängigkeit davon ändern würde, ob wir einen Regenschirm aufspannen oder nicht. Dann würde auch der Wetterbericht nicht mehr funktionieren.

Hinzu kommt, dass Modellierungsstudien in der Regel weder Studienprotokolle vorab veröffentlichen noch präregistriert werden – wie dies eigentlich heutzutage für qualitativ hochwertige Studien selbstverständlich sein sollte. Damit ist einem Herumprobieren, „bis es passt“, Tür und Tor geöffnet.

Auch historisch betrachtet haben Modellierungen von Epidemien keinen guten *Track Record*, allerdings erinnert sich heute kaum noch jemand daran. Man denke aber nur mal zurück an die Schweinegrippe oder an die Bovine Spongiforme Enzephalopathie (BSE). Auch damals lagen die prominenten Modellierer, die heute übrigens immer noch ganz vorne mit dabei sind, mit ihren Vorhersagen massiv daneben.

Bei der bereits erwähnten Prädiktion von epileptischen Anfällen – auch hier geht es ja um die Vorhersage zukünftiger Ereignisse aus komplexen Datensätzen – hat man übrigens aus den initialen Fehlern gelernt. Nach einer anfänglichen Euphorie mit darauffolgender kritischer Ernüchterung und Fehleranalyse ist eine etwas demütigere, aber dennoch nicht weniger relevante Wissenschaft entstanden. Mittlerweile gibt es dort rigorose Methoden, mit denen die jeweilige Güte von Vorhersagen geprüft werden kann. Die Pandemie-Modellierer von heute täten gut daran, mal einen Blick hierauf zu werfen.

Vielleicht besteht aber der eigentliche Nutzen der Pandemie-Modellierungen darin, *Worst-Case*-Szenarien wissenschaftlicher erscheinen zu lassen – und damit einschneidende Maßnahmen für die breite Masse einleuchtender und akzeptabler zu machen. Diese also wissenschaftlich zu bebildern. Das ist aber eine gefährliche Strategie: Zum einen, weil Vorhersagen, die danebenliegen, ihre Überzeugungskraft verlieren – zum anderen, weil die Modelle ja behaupten, die Nützlichkeit oder Schädlichkeit bestimmter Maßnahmen und Verhaltensweisen zu „objektivieren“. Wie zum Beispiel Schulschließungen, Ausgangssperren oder Abstandsregeln. Wenn die offensichtlichen und teils schwerwiegenden Limitationen der Modelle nicht erkannt oder berücksichtigt werden, sie aber dennoch die Grundlage für unser Handeln in der Pandemie liefern – dann läuft etwas schief.

Allerdings: Ob und – wenn ja – welchen Einfluss die derzeit sehr medienpräsenten Modellierer überhaupt auf die Politik haben, oder ob sie von dieser nur benutzt werden, um politisch motivierte Entscheidungen zu rechtfertigen, ist unklar. Dafür können die Model-

lierer natürlich erstmal nichts. Allerdings wehren sie sich auch nicht gegen eine solche Instrumentalisierung, sondern genießen die mediale Aufmerksamkeit. Der Narr hatte sich ja bereits in *Laborjournal* 11/2020 (S. 22-24) über das komplette Fehlen einer evidenzbasierten, inklusiven, gründlichen, transparenten und zugänglichen wissenschaftlichen Beratung der Corona-Politik echauffiert. Darin kommt er zu dem Schluss, dass das post-darwinistische Motto „*Survival of the ideas that fit*“ ist.

Modellierung funktioniert in der Pandemie bisher nur dort, wo sie sich auf wenig komplexe sowie teilweise gut verstandene Zusammenhänge verlässt und zudem die Datenlage einigermaßen robust ist. Das ist leider nicht häufig der Fall. Zum Beispiel liefert sie recht verlässliche und nützliche Vorhersagen, wo es um den Zusammenhang von Infektionsinzidenz, Auslastung von Intensivstationen und Todesfällen geht.

»Da die überkomplexen Modelle „overfitted“ werden, modellieren sie mehr Rauschen als Signal.«

Sobald die Modellierer sich aber auf komplexe, kaum oder gar nicht verstandene und zudem noch volatile Zusammenhänge stürzen, sobald die zugrundeliegenden Parameter auf nicht-verlässlichen oder nur geschätzten Daten beruhen, und sobald die Vorhersage Einfluss auf ihr eigenes Ergebnis hat – dann funktioniert es nicht mehr richtig. Die daraus resultierenden überkomplexen Modelle werden, damit überhaupt etwas Plausibles dabei herauskommt, „*overfitted*“ – es wird also mehr Rauschen als Signal modelliert. Eine vertiefte Diskussion der Limitationen und Unsicherheiten solcher Modelle samt deren Aussagen würde dabei genauso stören wie passende Kontrollen – und damit in weniger öffentlicher Aufmerksamkeit resultieren.

John von Neumann, Mathematiker, Physiker und Computer-Pionier, wird mit dem Bonmot zitiert: „Mit vier Parametern kann ich einen Elefanten *fit*ten, und mit fünf ihn mit dem Rüssel wackeln lassen.“ Wenn mit Rüssel-wackelnden Elefanten und dem Gestus mathematisch-physikalischer Autorität Politikberatung gemacht wird, ist das nicht ohne Risiko.

Der Wissenschaftsnarr dankt Christian Meisel und Gerd Antes für anregende Diskussionen. Weiterführende Literatur und Links finden sich wie immer unter: <http://dirnagl.com/lj>.