



Interview

Prof. Dr. Helmut Küchenhoff, Leiter des Statistischen Beratungslabors der Ludwig-Maximilians-Universität München

? Herr Prof. Küchenhoff, wann haben Sie das letzte Mal ein so großes Interesse und gleichzeitig so viele kritische Nachfragen zu statistischen Methoden wie in der Corona-Krise erlebt?

! In dieser intensiven Form eigentlich selten. Aber Kontroversen über statistische Methoden und Ergebnisse haben mich in meiner Arbeit immer begleitet. Besonders wenn es um Interessen geht, wie zum Beispiel beim Thema Wirkung von Luftschadstoffen auf die Gesundheit und in diesem Zusammenhang über die Wirkung von Umweltzonen. Auch die Diskussion über die Kernenergie war von Kontroversen zur Risikoabschätzung und zur Wirkung von Strahlung geprägt. In letzter Zeit habe ich mich im Rahmen der Debatte zum Klimawandel mit Diskussionen zur Bewertung entsprechender Daten beschäftigt. Grundsätzlich gehören Widersprüche zum Wissenschaftsprozess, sie sind sogar ein Motor neuer Erkenntnisse. In der aktuellen Krise sind schnelle Erkenntnisse gefragt, was dazu führt, dass es durch die Vielzahl von Studien und Datenanalysen sehr schwierig wird, den Überblick zu behalten, einzuordnen und die richtigen Schlüsse zu ziehen.

Corona: Statistik mit Verantwortung

? Für den Außenstehenden ist es kaum nachvollziehbar, warum die Daten binnen kürzester Zeit extrem schwanken. Warum ist es so schwer, die konkrete Zahl von neuinfizierten Menschen in einer Region zu bestimmen?

! Es gibt da viele verschiedene Aspekte: a) die Dunkelziffer: Viele Infizierte merken es nicht oder werden nicht getestet. b) die Unzuverlässigkeit von Tests: Manche Infizierte werden nicht positiv getestet. Es gibt auch falsch positive Tests. c) der Meldeverzug: Zwischen Testzeitpunkt und Meldung bei den zuständigen Behörden vergeht eine gewisse Zeit, die zur Laborauswertung, Datenerfassung und Weitergabe benötigt wird. d) die Inkubationszeit und die Zeit der Durchführung des Tests: Es

vergeht auch eine gewisse Zeit zwischen der Ansteckung und der Durchführung eines Tests.

? Da die Datenlage immer einige Tage hinterherhinkt, setzt man das Prinzip des Nowcastings ein, also der Vorhersage des echten heutigen Wertes der Infizierten aus vergangenen verifizierten Daten und einer geeigneten Modellierung der Entwicklung der gesuchten Größen in der Zeit. Wie genau sind die Vorhersagen im Vergleich zu den später tatsächlich registrierten Infektionsfällen?

! In unseren Analysen, die unser StaBLab-Team gemeinsam mit Michael Höhle von der Universität Stockholm durchführt, widmen wir uns den oben genannten Punkten c) und d). Wir nutzen die Tatsache, dass in den dokumentierten Fällen die Frage nach dem Beginn der Symptome gestellt wird. Daher sind wir in der Lage, ein statistisches Modell für die Zeit zwischen Krankheitsbeginn und Meldung aufzustellen. Dieses Modell ist komplex, da sich dieser Verzug im Laufe der Pandemie verändert hat. Er ist abhängig vom Wochentag der Meldung (am Wochenende gibt es weniger Meldungen) und von weiteren Größen. Wir verwenden dafür ein hierarchisch generalisiertes additives Regressionsmodell, mit dem wir in der Lage sind, die Anzahl der Neuerkrankungen zu schätzen, die noch nicht gemeldet sind. Dieses Verfahren funktioniert ähnlich wie das in der Versicherungswirtschaft sehr bekannte Chain-Ladder-Verfahren. Die noch nicht aufgetretenen Schäden entsprechen bei uns den nicht gemeldeten Neuerkrankungen.

Bezüglich der Genauigkeit der Vorhersagen arbeiten wir aktuell intensiv an der Evaluation der Methodik. Wir können basierend auf künstlichen Daten zeigen, dass das Modell die tatsächliche Anzahl von Erkrankten pro Tag adäquat schätzt, auch wenn es Änderungen der Meldeverzugsverteilung über die Zeit hinweg gibt. Inzwischen können wir teilweise auch schon die Daten zu COVID-19-Fällen mit Krankheitsbeginn bis vor circa drei Wochen zur Evaluation nutzen, da wir für diesen Zeitraum keine weiteren gemeldeten Krankheitsbeginne für offiziell registrierte Fälle erwarten. Der Meldeverzug zwischen Krankheitsbeginn und Meldung ist für die meisten gemeldeten Fälle deutlich kleiner.

? Und wo gibt es noch Nachbesserungsbedarf?

! Wir haben unseren Artikel bei einer wissenschaftlichen Zeitschrift eingereicht und wurden von den Gutachtern noch auf kritische Punkte hingewiesen. Insbesondere gibt es Instabilitäten, wenn die Daten aufgrund spezieller Situationen im Meldeprozess Ausreißer haben. Weiter bleibt das Problem der Dunkelziffer. Durch die verstärkte Aktivität beim Testen könnte es bei den gemeldeten Zahlen zu einer Steigerung kommen, obwohl das Infektionsgeschehen in Wirklichkeit zurückgeht. Daher müssen neben unseren Daten auch andere Datenquellen wie Todesfälle oder Krankenhauseinweisungen beobachtet werden.

? Ein tieferer Blick in die Methodik zeigt, dass das RKI und Nowcasting mit sogenannten „95%-Prädiktionsintervallen“ arbeiten. Bitte erläutern Sie das einmal für Statistik-Laien und was das für die Aussagekraft der Ergebnisse bedeutet.

! Zu jeder guten Prognose gehört eine Quantifizierung ihrer Unsicherheit. Hierzu werden üblicherweise Prognoseintervalle benutzt. Ein 95%-Prognoseintervall sollte in 95 Prozent der Fälle den wahren Wert enthalten. Das liefert eine sehr anschauliche Beschreibung der Unsicherheit. Wenn Prognoseintervalle mit statistischen Methoden geschätzt werden, wird vorausgesetzt, dass das zugrunde liegende Modell weiterhin gilt. Daher kann die Unsicherheit noch größer sein, da es ja die Möglichkeit einer Strukturänderung gibt.

? Corona wird leider nicht die letzte Krise bleiben, in der Statistik gefragt ist. Wo sehen Sie bis zur nächsten Krise Bildungs- und Nachbesserungsbedarf bei der Politik, den Medien, der Wissenschaft und auch der Öffentlichkeit?

! Als Erstes wünsche ich mir eine stärkere Einbeziehung der wissenschaftlichen Statistik in den Diskurs. In wissenschaftlichen Beratergremien ist neben den entsprechenden Fachwissenschaften auch die statistische Expertise gefragt. Ein gutes Beispiel hierfür war die von der Bundesregierung eingesetzte Kommission der Leopoldina zu Stickoxiden und Feinstaub.

Wir bringen unseren Studierenden bei, bei Schlüssen aus Daten immer vorsichtig zu sein, die Grenzen zu kennen und zu benennen. Das ist – unabhängig von der aktuellen Krise – auch im Bereich von Big Data und neuerdings auch künstlicher Intelligenz wichtig. Häufig werden die Möglichkeiten des Ergebnisgewinns durch komplexe, rechenintensive Datenanalysen überschätzt. Auch wenn für uns als Statistiker gerade der Erkenntnisgewinn aus Daten das zentrale Geschäftsmodell ist, sind diese Erkenntnisse immer mit der nötigen Portion Skepsis zu sehen. Weiter sollte generell in der Bildung die Fähigkeiten der statistical literacy – also das Verstehen und kritische Bewerten von Daten – stärker in die Schulbildung einbezogen werden. Ein wichtiger Fortschritt aus dem Bereich der Statistik sind die Ideen „open science“, „open data“ und „reproducible research“. Hier ist die Forderung nach der Offenlegung der Daten und Methoden zentral, was erst eine wirklich offene und nachvollziehbare wissenschaftliche Diskussion ermöglicht.

Weiter halte ich interdisziplinäre Gremien von Wissenschaftler*innen, wie es in der Schweiz die Swiss National COVID-19 Science Task Force ist, für sinnvoll. Darüber hinaus sollte das RKI als wichtigstes deutsches Forschungsinstitut seine Statistik- und Methodenabteilung erheblich ausbauen und noch stärker mit den Universitäten vernetzt werden.