

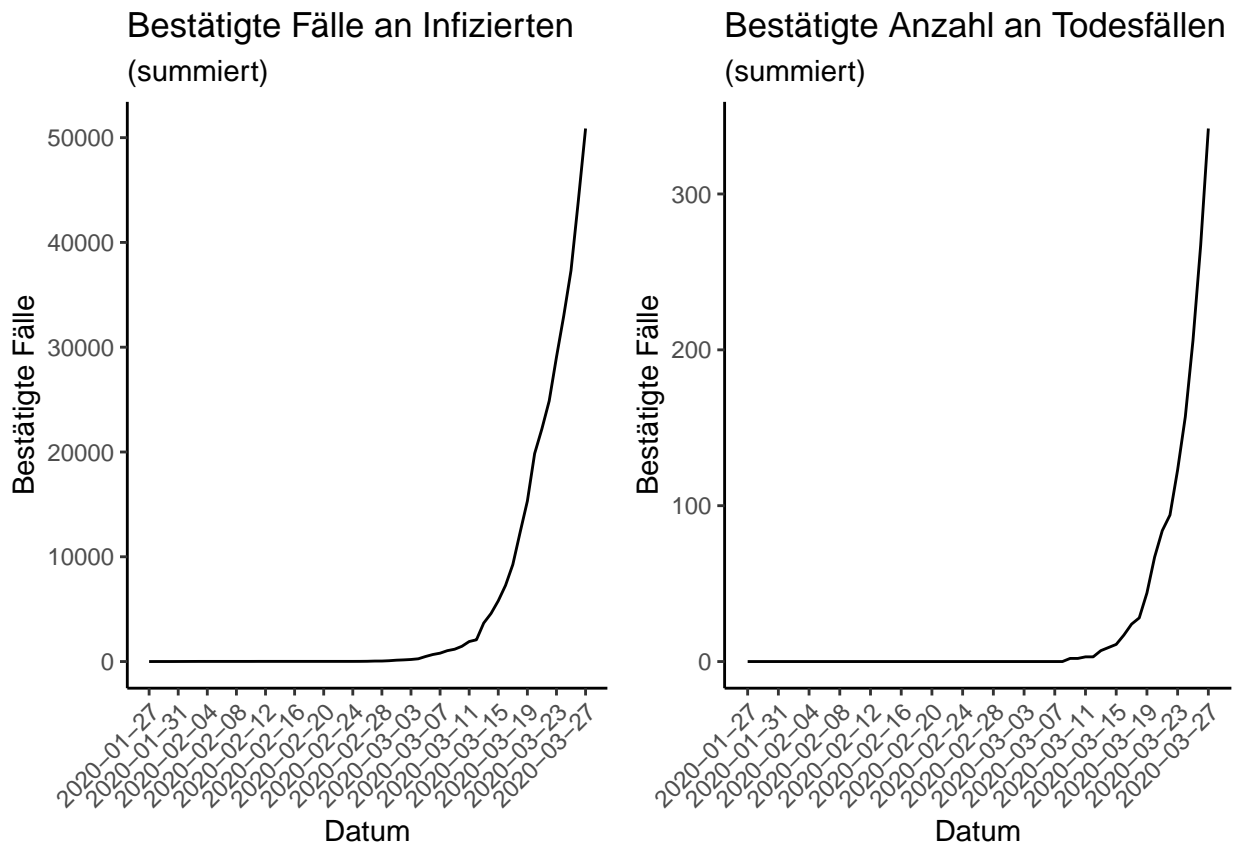
Was man auf Basis der vorliegenden Daten zur Ausbreitung von COVID-19 in Deutschland sagen kann – und was nicht

von Peter Pütz

28.03.2020

Worum geht es?

Zur Beurteilung der Ausbreitung und der Gefährlichkeit von COVID-19 in Deutschland werden in der Öffentlichkeit aktuell vor allem die Entwicklungen von zwei Zahlen angeschaut und diskutiert: Die Anzahl der bestätigten Infizierten (genauer: der positiv Getesteten) und die der Todesfälle. Auch das Robert-Koch-Institut veröffentlicht diese Zahlen und präsentiert sie den Medien. So sehen diese Kurven aktuell für Deutschland aus¹:



Was sagen uns diese Kurven - und was nicht? Können wir am Verlauf der Kurve der bestätigten Infizierten erkennen, ob sich die Ausbreitung des Virus beschleunigt oder verlangsamt, wie häufig behauptet wird? Ich nehme die Antwort vorweg: Nicht wirklich. Ist die Anzahl der Todesfälle verlässlicheres Maß? Hier lautet meine Antwort: Jein. Wie ich auf diese beiden Antworten komme, versuche ich im Folgenden in einfachen Worten zu erklären.

¹basierend auf den Zahlen des Johns Hopkins University Center for Systems Science and Engineering (CSSE), s. hier: <https://github.com/datasets/covid-19/blob/master/data/countries-aggregated.csv>.

Die bestätigten Fälle an Infizierten

Fangen wir an mit der ersten Kurve, die die bestätigten Fälle an Infizierten beschreibt. Es ist vielen bekannt, dass die tatsächliche Anzahl an Infizierten viel höher ist, vor allem deshalb, da viele Menschen keine oder nur schwache Symptome aufweisen und oft nicht getestet werden, wenn sie keinen Kontakt zu einer nachweisbar infizierten Person hatten oder sich nicht in einem Risikogebiet aufgehalten haben. Die verbreiteten noch geringen Zahlen gerade in einzelnen Städten und Landkreisen verleiten evtl. zu einer Unterschätzung der Gefahr, sich anzustecken. Leider wird trotzdem oft von der "Anzahl an Infizierten" gesprochen, was schlichtweg falsch ist.

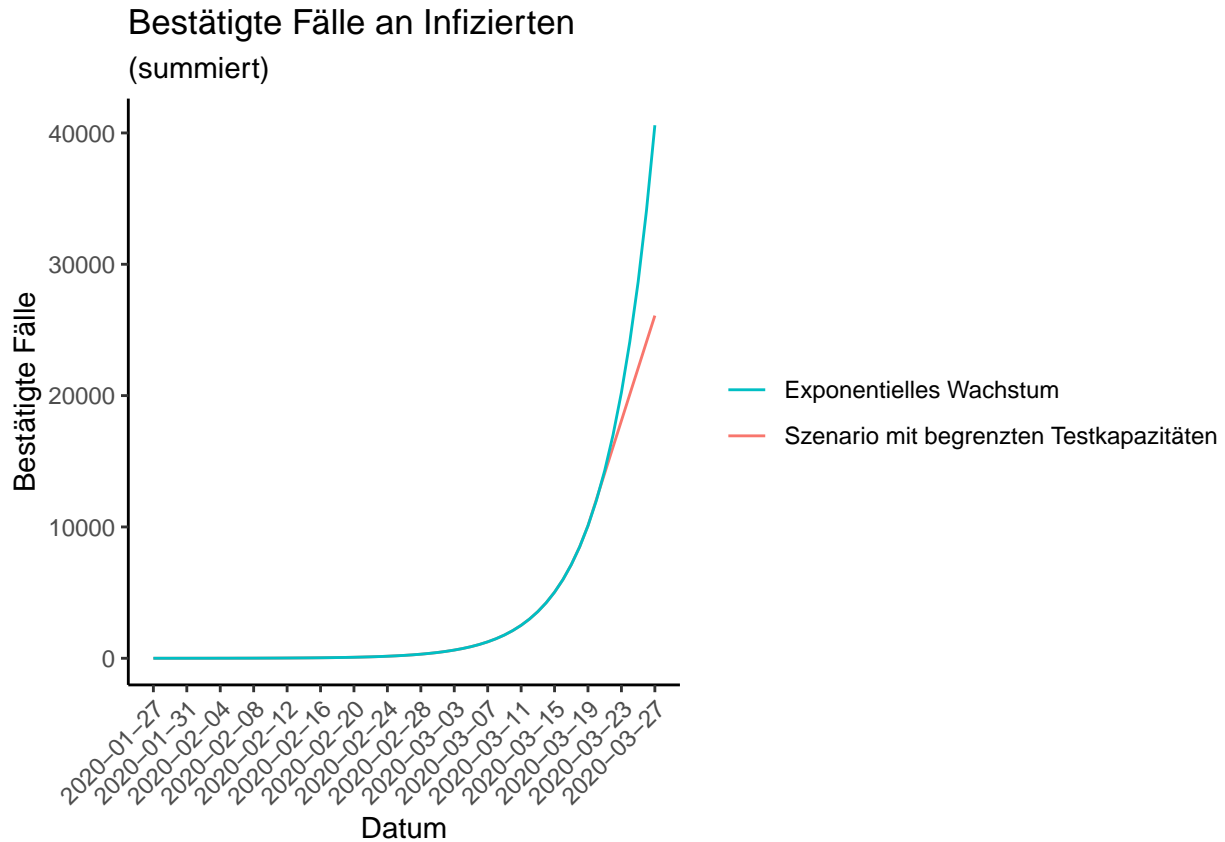
Hier soll es jedoch nicht um die tatsächliche Anzahl an Infizierten gehen, sondern um die Entwicklung der bestätigten Fälle und was man daraus lesen kann. Die Ausbreitung des Virus könnte an der vorliegenden Kurve beurteilt werden, solange z.B. alle Leute mit Symptomen getestet werden, da in diesem Fall angenommen werden kann, dass der Anteil an Infizierten mit Symptomen an allen Infizierten über die Zeit in etwa gleich bleiben sollte. Anders gesagt: Wenn die ganze Zeit ein bestimmter (konstanter!) Prozentsatz der tatsächlich Infizierten erkannt wird, dann geht auch die Anzahl an tatsächlich Neuinfizierten zurück, wenn die Anzahl an (durch Tests) neu erkannten Fällen zurück geht - die Kurve also nicht mehr so stark ansteigt ("abflacht"). Viele hoffen auf ein "Abflachen der Kurve" der bestätigten Infizierten, damit das Gesundheitssystem nicht überlastet wird, weniger Leute an COVID-19 sterben und Zeit erkaufte wird. Die Kurve wird zur Bewertung herangezogen, ob politische Maßnahmen und Verhaltensänderungen der Menschen zu einer Verlangsamung der Ausbreitung oder sogar Eindämmung des Virus geführt haben. Das Problem ist, dass uns diese Kurve allein nur sehr wenig über genau diese Dinge sagen kann. Dafür bräuchte es nämlich zusätzlich Informationen über die Anzahl an täglich durchgeführten COVID-19-Tests und Informationen darüber, nach welchen Kriterien getestet wird und wie viele Menschen diesen Kriterien entsprechen. Diese Punkte hängen eng zusammen: Wie schon oben beschrieben, wäre die Kurve aussagekräftig für die Ausbreitung des Virus, wenn über den gleichen Zeitraum alle Menschen mit Symptomen getestet werden - also alle Menschen mit Symptomen erfasst werden und auch getestet werden können. Es kann sogar noch allgemeiner gesagt werden: Die Kurve hätte diese Aussagekraft ebenso, wenn die Kriterien, nach denen Menschen getestet werden, über die Zeit gleich bleiben und alle Menschen, die diese Kriterien erfüllen, auch getestet werden. Die Begründung dafür bleibt gleich: Solange bei konstanten Testkriterien der Anteil an getesteten Menschen an allen Infizierten gleich bleibt, können mit Hilfe der Kurve valide Aussagen über die Ausbreitung des Virus getätigt werden. In Deutschland sollen nach Empfehlungen des Robert-Koch-Instituts im Moment (seit dem 24.3.2020) alle Menschen getestet werden, die akute respiratorische Symptome aufweisen oder für die es klinische oder radiologische Hinweise auf eine virale Pneumonie gibt.² Diese Kriterien haben sich bereits über die Zeit geändert und es ist nicht klar, inwiefern den Empfehlungen gefolgt wird bzw. werden kann. Denn ganz entscheidend sind dafür die Anzahl an Tests, die durchgeführt werden (können).

Warum die Anzahl der durchgeführten Tests so wichtig ist

Eine konstante Anzahl an täglich durchführbaren Tests führt bei einer stetigen Zunahme an nach den Kriterien zu testenden Menschen irgendwann dazu, dass nicht mehr alle getestet werden können oder sich die Kriterien ändern müssen. Die Folge ist, dass die Kurve "abflacht", die Anzahl der registrierten Neuinfizierten also abnimmt. Das lässt sich an einem ausgedachten Zahlenbeispiel verdeutlichen: Nehmen wir an, dass jeden Tag 10.000 Tests auf COVID-19 in Deutschland durchgeführt werden können. Weiterhin nehmen wir an, dass es heute in Deutschland genau 10.000 Menschen gibt, die die Kriterien zum Testen erfüllen. Jeder dieser Menschen kann also heute getestet werden und wir nehmen an, dass das auch passiert. Angenommen, bei 5.000 der 10.000 Tests wird eine Infektion mit COVID-19 festgestellt. Diese 5.000 gehen in die Statistik (die Kurve) ein. Nehmen wir weiter an, dass das Virus sich weiter ausbreitet, sodass es in drei Tagen schon 20.000 Menschen gibt, die die Testkriterien erfüllen. Es können aber weiterhin nur 10.000 Tests durchgeführt werden, sodass aus diesen 20.000 mehr oder weniger zufällig 10.000 Menschen getestet werden. Nehmen wir wieder an, dass die Hälfte der 20.000 infiziert sind und auch die Hälfte der 10.000 Menschen, die getestet werden.

²Stand 28.3.2020: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Massnahmen_Verdachtsfall_Infografik_DINA3.pdf?__blob=publicationFile.

Dann wird das Robert-Koch-Institut wieder 5.000 neue Infizierte vermelden und einige werden schließen: Wir haben es geschafft, das exponentielle Wachstum zu verlangsamen, das Virus breitet sich langsamer aus. Dass das aber völlig falsch ist, liegt wohl auf der Hand. Ein Beispielszenario, in dem die Testkapazitäten, Testkriterien und die wahre Infektionsrate unter den die Testkriterien erfüllenden Leuten konstant bleiben und die Anzahl der die Testkriterien erfüllenden Menschen exponentiell steigt, weil sich das Virus ungehemmt ausbreitet, ist noch einmal in folgender Graphik zu sehen:



Es wurde hier angenommen, dass sich in dem simulierten Szenario zunächst alles so verhält wie eine angenommene wahre Verbreitung mit einem exponentiellen Wachstum, bei der alle Menschen, die die Testkriterien erfüllen, auch getestet werden können (blaue Kurve). In unserem Szenario mit endlichen Testkapazitäten (rote Kurve) übersteigt ab dem 20. März die Anzahl an Menschen, die die Testkriterien erfüllen, die vorhandenen Testkapazitäten. Diese reichen nur noch dazu aus, 2000 Infektionen pro Tag nachzuweisen. Man sieht, dass die Anzahl der registrierten Fälle in diesem Fall konstant bleibt, das Wachstum sich also verlangsamt. Das Beispiel mag extrem sein, aber die Grundproblematik sollte deutlich geworden sein. In der Realität ist es momentan so, dass die Testkapazitäten in der letzten Zeit deutlich zugenommen haben.³ Das kann aber dazu geführt haben, dass mehr Leute positiv getestet wurden und so in die Statistik eingehen, nur weil mehr getestet wurde. Fazit: Die Kurve alleine sagt nichts über die Ausbreitung des Virus aus. Es ist unbekannt, wie viele Tests täglich in Deutschland durchgeführt wurden und werden. Im Gegensatz zu anderen Ländern, z.B. Südkorea, werden diese Daten nicht vom Robert-Koch-Institut bereit gestellt, obwohl sie verfügbar sind.

Es gibt natürlich noch andere Probleme mit der Statistik im Falle der Kurve der bestätigten Fälle, z.B. das Verhalten von potenziell Infizierten mit entsprechenden Symptomen. Es ist unwahrscheinlich, dass dieses über die Zeit konstant bleibt, diese Infizierten also mit gleicher Wahrscheinlichkeit ihr Recht auf den Test wahrnehmen. Auch andere Probleme wie der zeitliche Verzerrung der Auswertung in unterschiedlichen

³https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-03-26-de.pdf?__blob=publicationFile&fbclid=IwAR3v8hZBxj7mK2bo2GoW2_QgWMy9xq9GIHSwK-lVnE3UmpaDOXpAVOamz_o.

Testzentren und die Zusammentragung der Daten aus den einzelnen Gesundheitsämtern können einige Tücken bereithalten. Die mangelnde Aussagekraft der Kurve über die Ausbreitung des Virus wird aber vor allem dadurch ausgelöst, dass wir sehr wenig über die Anzahl der täglich durchgeführten und der den Testkriterien entsprechenden benötigten Tests wissen.

Die Anzahl der Todesfälle

Die einzigen öffentlich verfügbaren Daten, die für die Ausbreitung von COVID-19 in Deutschland bisher und wohl auch in naher Zukunft einigermaßen aussagekräftig sein könnten, ist die Anzahl der Todesfälle im Zusammenhang mit COVID-19, sofern diese über die Zeit einheitlich bestimmt werden.⁴ Es ist davon auszugehen, dass sich der Prozentsatz der Menschen, die im Zusammenhang mit einer COVID-19-Erkrankung sterben, an allen Infizierten in etwa gleich bleibt.⁵ Die Interpretation der Kurve ändert sich erst, wenn sich die medizinischen Kapazitäten derart verknappen, dass die intensivmedizinische Betreuung leidet und prozentual mehr an COVID-19 erkrankte Leute sterben, da sie nicht mehr (ausreichend) behandelt werden können. Ab diesem Zeitpunkt überschätzt die Kurve die Ausbreitung der Pandemie. Ansonsten sind die Nachteile der Kurve, um die Ausbreitung abzuschätzen, folgende:

- Es gibt bisher in Deutschland nur geringe Fallzahlen, sodass man vor allem auf regionaler Ebene noch nicht viel sagen kann, wie stark sich das Virus verbreitet. Der Zufallsfaktor in den Daten ist noch zu hoch, um klare Trends und Änderungen der Trends zu erkennen.
- Die Zahlen sind nicht tagesaktuell. Es dauert in der Regel einige Tage vom positiven Testergebnis bis zum Tod, eine Schätzung der durchschnittlichen Dauer beträgt ca. 13 Tage.⁶ Deshalb können die Ausbreitung und die Eindämmung des Virus erst verzögert festgestellt werden. Daraus folgt übrigens auch, dass die Todesraten (case fatality ratio) nicht der Quotient aus Anzahl aller bisher Gestorbenen und Anzahl aller bisher Infizierten ist (ein weiterer oft verbreiteter Irrtum), sondern in Wirklichkeit höher sind, die Zeitverzögerung mit eingerechnet werden muss.⁷ Auch bei der Schätzung der Todesraten kommt noch das anfangs erwähnte Problem der hohen Dunkelziffer an Infizierten hinzu.

Zusammenfassung

Die in der Corona-Krise so oft zitierten und angeschauten beiden Kurven haben eine andere Aussagekraft, als häufig angenommen und nahegelegt wird. Insbesondere die Ausbreitung des Virus allein an der Anzahl der bestätigten Infizierten festzumachen, ist gefährlich. Dazu ist die Anzahl an Todesfällen prinzipiell geeigneter, aber sie bildet Entwicklungen nur recht verzögert ab. Grundsätzlich fehlt es in Deutschland an relevanten Daten, um die Entwicklung des Virus treffend und zeitnah beurteilen zu können. Dafür am geeignetsten wären regelmäßige Zufallsstichproben in der Gesamtbevölkerung, die auch vernünftige Aussagen über die Todesrate der Corona-Infizierten zulassen würden.

Hinweis: Dieser Beitrag ist eine rein statistische Betrachtung von in der Öffentlichkeit viel diskutierten Zahlen im Zusammenhang mit COVID-19. Diese Betrachtung dient nicht dazu die bisher vorgenommenen Maßnahmen der Bundesregierung zu beurteilen und reicht dafür auch in keinster Weise aus.

Für Fragen und Anmerkungen: ppuetz@uni-goettingen.de.

⁴Dies wäre zum Beispiel der Fall, wenn alle Menschen, die mit für COVID-19 typischen Symptomen im Krankenhaus sterben, im Laufe ihres Aufenthalts auf COVID-19 getestet werden und somit in die Statistik eingehen.

⁵Es ist nicht immer eindeutig zu erkennen, was schlussendlich den Tod verursacht. Ein Mensch kann an COVID-19 erkrankt sein und an einer anderen Krankheit sterben oder schon eine tödliche Krankheit haben und COVID-19 kommt noch hinzu. Dies ist zur Einschätzung der Ausbreitung des Virus irrelevant, zur Berechnung der Sterberate hingegen schon.

⁶Linton NM, Kobayashi T, Yang Y et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *Journal of Clinical Medicine* 2020;9:538.

⁷Das wird ganz gut in dieser Studie erklärt: https://cmmid.github.io/topics/covid19/severity/global_cfr_estimates.html.