

Corona-Prognose. 1. Darstellung der Fallzahlen

Christoph Bandt*

15. April 2020

Zusammenfassung

Für die Beurteilung des Verlaufs der Covid-19 Epidemie stehen uns nur wenige hastig gesammelte Daten zur Verfügung. Aber mit sehr einfacher Mathematik kann man aus diesen Daten viel erkennen. Hier zeigen wir durch Betrachtung wöchentlicher Zahlen von Neuinfizierten, dass wir in Deutschland seit mindestens 10 Tagen über den Berg sind.

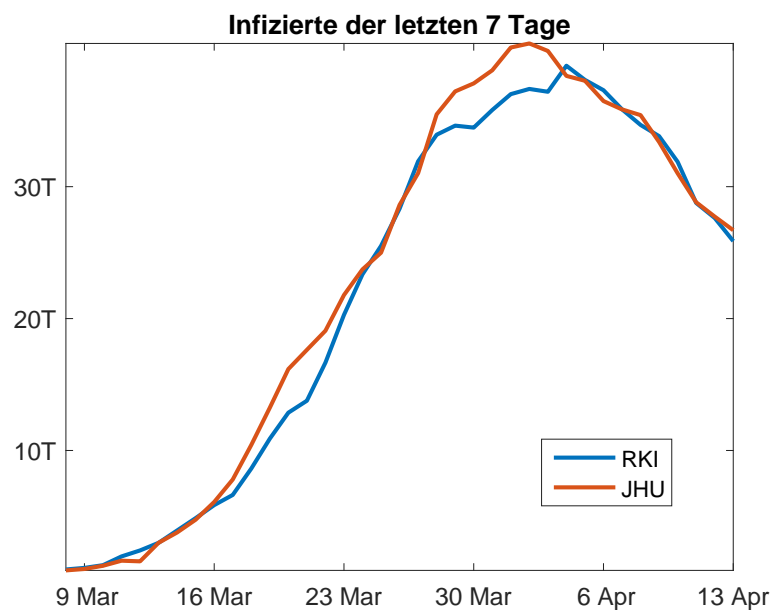


Abbildung 1: Neuinfizierte der letzten 7 Tage in Deutschland, 8. März bis 13. April.

*Institut für Mathematik und Informatik, Universität Greifswald, bandt@uni-greifswald.de

Daten. Wir verwenden nur die Zahlen der laborbestätigten Infizierten. Bei der Johns Hopkins University (JHU) kann man die Zeitreihe im Github-Archiv herunterladen. Beim Robert-Koch-Institut (RKI) habe ich sie aus den täglichen Situationsberichten abgeschrieben. Zum download gibt es dort bisher nur die Zahlen des aktuellen Tages, aufgeschlüsselt nach Bundesland oder Landkreis. Die Daten sind in einer Excel-Datei enthalten, die diesem Text angefügt wird. Damit lassen sich die Rechnungen überprüfen.

Was sehen wir oben in Abbildung 1? Die Kurve zeigt für jeden Tag die Anzahl der in den letzten Tagen neu Infizierten. Diese Zahl ist ein Ausdruck für die Infektionsgefahr in Deutschland. In den Berichten des RKI wird sie als Covid-19 Aktivität bezeichnet, und ihre Verteilung in den deutschen Landkreisen wird als gelb-braun gefärbte Karte dargestellt. Hier geht es nur um die zeitliche Entwicklung. Die Infektionsgefahr ist im März ständig gestiegen. Sie hat laut JHU am 2. April und laut RKI am 4. April ihren Höhepunkt erreicht und verringert sich seitdem von Tag zu Tag. Interessant ist, dass auf der absteigenden Seite die Zahlen von JHU und RKI nahezu übereinstimmen.

Die Summenkurve aller Fälle. In den Medien wird die Entwicklung durch die Kurve aller Infektionen ab Anfang der Epidemie dargestellt. Dies ist die originale Zeitreihe, die bei JHU und RKI angegeben wird. Jeden Tag werden die Neuinfizierten dazugezählt. Es werden immer mehr Fälle. Die Kurve steigt und steigt beängstigend. Diese Kurve ist so definiert, dass sie nicht fallen kann und dass man nicht genau erkennt wann es besser wird. Veränderungen sind optisch in solch einer Summenkurve so gut wie gar nicht zu erkennen.

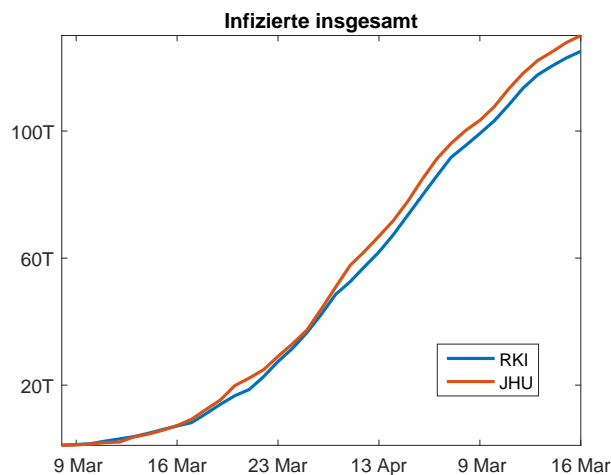


Abbildung 2: Summe aller Infizierten für jeden Tag vom 8. März bis 13. April.

Solche Summen sind auch sonst nicht aufschlussreich, wenn etwa gemeldet wird „88 Menschen waren oder sind im Krankenhaus“. Man möchte doch eher wissen, wie viele heute noch dort liegen müssen.

Am Anfang der Epidemie, in der Zeit des unbegrenzten Wachstums, hatte diese Darstellung noch eine gewisse Berechtigung. Epidemiologische Modelle, die seit etwa hundert Jahren betrachtet werden, liefern Lösungskurven von diesem Typ. Erschreckende Simulationsstudien des unbegrenzten Wachstums, insbesondere eine Studie des Imperial College London vom 16.3.2020, haben die Politiker vieler Länder davon überzeugt, dass entschiedene Sofortmassnahmen nötig sind, um die Pandemie einzudämmen und Überlastung der Gesundheitssysteme zu verhindern.

Nachdem die Einschränkungen des öffentlichen Lebens erfolgt sind, gelten die üblichen Modelle aber nicht mehr. Es gibt für die gegenwärtige Situation keine geeigneten Modelle. Das Problem ist, dass ein Infizierter nicht einfach halb so viele Leute anstecken kann wie sonst, sondern dass er nur ganz bestimmte Leute in seinem Umfeld anstecken kann. Dies ist ein sogenanntes Perkolationsproblem auf zufälligen Netzwerken, was nicht einmal im Modellfall ausreichend geklärt ist. Deswegen müssen wir uns an die vorhandenen Daten halten und das Beste daraus machen.

Für die JHU, die Daten aus aller Welt aus zum Teil einander widersprechenden Quellen sammelt, ist die Summenkurve weiterhin die geeignete Datenstruktur. Wenn man falsch liegt, kann man sich leicht korrigieren, indem man die Kurve am nächsten Tag langsamer oder schneller wachsen lässt. Einmal gemeldete Werte brauchen so nicht geändert zu werden.

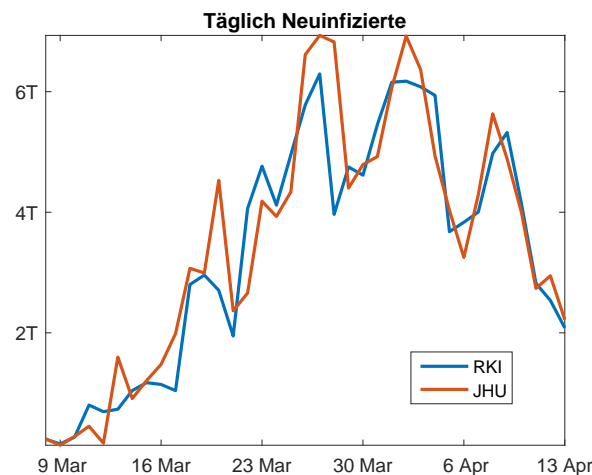


Abbildung 3: Tägliche Neuinfektionen in Deutschland vom 8. März bis 13. April.

Die Kurve der täglichen Neuinfektionen. Warum nehmen wir nicht einfach die Kurve der Neuinfizierten an jedem Tag? Abbildung 3 zeigt den Grund. Diese Kurve schwankt sehr stark. Das liegt nicht am Virus, sondern am Arbeitsrhythmus der Gesundheitsbehörden. Sonntags kommen wenige Fälle dazu, am Mittwoch umso mehr. Auch wer erste Symptome am Samstagabend spürt, zögert gleich zum Arzt zu gehen.

Diese Schwankungen machen eine Voraussage schwer, obwohl sich der Trend von Abb. 1 auch in Abb. 3 andeutet. Heute ist Mittwoch, und eine größere Zahl von Neuinfektionen ist zu erwarten. In der wöchentlichen Summe wirkt sich die aber nicht so stark aus, weil der letzte Mittwoch dafür wegfällt.

Vorteile der wöchentlichen Summen. Durch die Summierung über jeweils eine Woche wird der wöchentliche Rhythmus der Zahlen in Abb. 3 gedämpft. Tatsächlich kommt man dadurch der Realität der Epidemie näher. Im Dashboard des RKI geht man den umgekehrten Weg: jeder Fall wird dem Tag zugeordnet, an dem das Gesundheitsamt die Akte angelegt hat. Dadurch wird der wöchentliche Rhythmus verstärkt. Ausserdem werden auch zurückliegende Zahlen weiterhin täglich verändert. Für die Auswertung stört das mehr als wenn bei den Daten der JHU eventuell mal ein Wert durch Extrapolation zustande kam. Deswegen wurden vom RKI die Daten aus den täglichen Berichten genutzt.

Die Summierung über mehrere Tage ist ein wohlbekanntes Mittel, um generell Unregelmäßigkeiten in Zeitreihen auszugleichen. Man spricht von einer gleitenden Summe, die als Glättungsfilter verwendet wird.

Der wesentliche Vorteil ergibt sich aus den Eigenheiten von Covid-19. Ein Corona-Positiver ist nämlich im Durchschnitt 7 Tage lang infektiös. Ich muss also vor allen positiv Getesteten der letzten Woche auf der Hut sein. Im Einzelfall kann die infektiöse Phase länger dauern, darum soll man zur Sicherheit zwei Wochen in Quarantäne bleiben. Aber bei milden Symptomen kann es auch nur wenige Tage dauern. Für die Statistik sind 7 Tage ein guter Richtwert. Wir wissen so wenig über Covid-19, dass wir das nicht genauer machen können. Das war jetzt etwas vereinfacht dargestellt, aber ich möchte mich kurz fassen.

Eine weiter gehende Überlegung. Die wöchentliche Summe repräsentiert also genau die Gefahr vor der wir uns fürchten. Dies schließt insbesondere die Gefahr durch Infizierte ohne Symptome ein, die unter dem Stichwort Dunkelziffer diskutiert wird. Für die Einschätzung der Gefahr müssen wir diese Dunkelziffer nicht kennen. Es ist anzunehmen, dass symptomlose Infizierte nicht ganz so ansteckend sind wie symptomatische. Aber auch das müssen wir nicht genau wissen. Die Anzahl der Neuinfizierten ist der Desamtschaden, der durch die Ansteckung verursacht wurde. Und das entspricht dem Gefahrenpotential. Kurzfristig brauchen wir die Dunkelziffer nicht. Später ist sie natürlich interessant, weil man wissen möchte welcher Teil der Bevölkerung schon immun ist. Darum werden jetzt Studien mit Antikörpertests begonnen.

Vor den heute gemeldeten Neuinfizierten brauchen wir keine Angst zu haben. Sie sind ja in Quarantäne. Wer heute neu in die Statistik kommt, hat sich im Durchschnitt vor etwa 14 Tagen angesteckt. Bei wem, wissen wir meist nicht. Die Zahl der Neuinfizierten ist damit *ein Maß für die Infektionsgefahr vor 14 Tagen*, und zwar sowohl von Seiten der symptomatischen Infizierten wie auch von Seiten der Infizierten, die selbst keine Probleme haben. Die wöchentliche Summe ist also die beste Schätzung für die wahre Gefahr - nur leider um 14 Tage versetzt. Und sie ist gleichzeitig ein Maß für die Belastung der Krankenhäuser in der zurückliegenden Woche, weil etwa 15 Prozent der Infizierten ins Krankenhaus muss. Es ist nicht zu erwarten, dass der Krankenhausaufenthalt im Durchschnitt genau 7 Tage dauert, da manche Patienten lange zur Genesung brauchen. Ungefähr könnte es jedoch stimmen.

Aus diesen Gründen ist sehr zu empfehlen, die Epidemie in der gegenwärtigen Phase durch gleitende Wochensummen zu illustrieren. Insbesondere in den Medien, weil das leicht zu verstehen ist.

Man sieht ja nicht nur den positiven Trend, sondern auch die noch Besorgnis erregenden absoluten Zahlen. Von 40000 Leuten, die mich in Deutschland im Prinzip anstecken könnten, sind wir jetzt auf 30000 runter. Auch dies sind natürlich Richtwerte. Auf jeden Fall ist die Gefahr nur noch 3/4 von der Gefahr am höchsten Punkt. Aber inzwischen sind ja schon zwei Wochen vergangen!

An dieser Stelle möchte ich schließen und die ganz entscheidende Frage der 14 Tage Verzögerung in einem zweiten Kommentar besprechen. Eine ausführliche Fassung auf Englisch mit den Verläufen einiger europäischer Länder kann man auf dem mathematischen Preprintserver arXiv finden:

<http://arxiv.org/abs/2004.04732>.

Das ist allerdings eine Woche alt und somit nicht mehr aktuell.

Anhang zur Excel-Datei. Die Datei Coronazahlen einfach.xlsx zur Nutzung mit MS Excel oder Libre Office enthält die Summenzahlen aller Infizierten in Zeile 5 und 6. In Zeile 8 und 9 sind daraus die täglichen Neuinfektionen berechnet. Feld C8 ist =C5-B5 . Diese Formel wird mit dem Cursorkreuz dann die ganze Zeile nach rechts gezogen. Also die heutigen Infektionen sind alle Fälle bis heute minus alle Fälle bis gestern.

Die wöchentlichen Zahlen sind in Zeile 11 und 12. Feld I11 ist =I5-B5 . Das wird wieder mit dem Cursorkreuz durch die ganze Zeile gezogen. Also wenn heute Donnerstag ist, sind die Fälle der letzten Woche gleich allen Fällen bis heute minus alle Fälle bis letzten Donnerstag. Das ist die ganze Rechnung, leicht zu überprüfen.

In Excel lassen sich die Abbildungen leicht aktualisieren oder verändern, Libre Office hat damit Schwierigkeiten. Zum Beispiel kann man die y-Achse logarithmisch skalieren.