

COVID-19: Erwartete Anzahl von Toten und Dunkelziffer in Deutschland

Matthias Kohl

04.04.2020

Hochschule Furtwangen

Institute of Precision Medicine

E-Mail: kohl@hs-furtwangen.de

Einführung

Nicht nur in Italien (Rizzello 2020), sondern weltweit führen wir aktuell einen Krieg gegen COVID-19. Völlig unklar ist noch, wie viele Todesopfer dieser Krieg fordern wird. Auf Basis aktuell vorliegender Zahlen und mit Hilfe einfacher statistischer Mittel wird im Folgenden die erwartete Anzahl von COVID-19 Toten in Deutschland berechnet. Außerdem wird eine Schätzung für die Dunkelziffer ermittelt. Die Berechnungen und graphischen Darstellungen wurden mit der Statistiksoftware R (R Core Team 2020) und den Erweiterungspaketen MKinfer (Kohl 2019), ggplot2 (Wickham 2016), ggsci (Xiao 2018) und plotly (Sievert 2018) durchgeführt. Das Dokument wurde unter Verwendung von knitr (Xie 2015) und R Markdown (Xie, Allaire, und Golemund 2018) erstellt. Eine interaktive html-Version inklusive der zugehörigen R Markdown Dateien findet sich unter <https://github.com/stamats/COVID-19>.

Bevölkerung in Deutschland

Laut dem Statistischen Bundesamt leben aktuell etwas mehr als 83 Mio. Menschen in Deutschland. Für unsere Berechnungen verwenden wir die prognostizierten Zahlen des Statistischen Bundesamtes für 2019 (Simulationsvariante 1) und gehen folglich von einer Bevölkerungszahl von 83.2 Mio. Menschen aus (DESTATIS: Statistisches Bundesamt 2020a).

Die Abbildung 1 zeigt, wie sich die Bevölkerung von Deutschland 2019 auf verschiedene Altersgruppen verteilte.

Wir werden statistische Ungenauigkeiten bei den Bevölkerungszahlen in den folgenden Berechnungen nicht berücksichtigen.

Letalität in China bis 11.02.

Die altersabhängigen Letalitäten – d.h., die Anteile der Verstorbenen unter allen Erkrankten – in China vom 01.01.2020 bis zum 11.02.2020 sind in Abbildung 2 dargestellt (Verity u. a. 2020, Table 1; Wu und McGoogan 2020). In der Arbeit von Verity et al. (2020) wurde zur Berücksichtigung der statistischen Unsicherheit bei den beobachteten Letalitäten unter anderem das 95%-Konfidenzintervall (95%-KI) von Clopper-Pearson (Clopper und Pearson 1934)

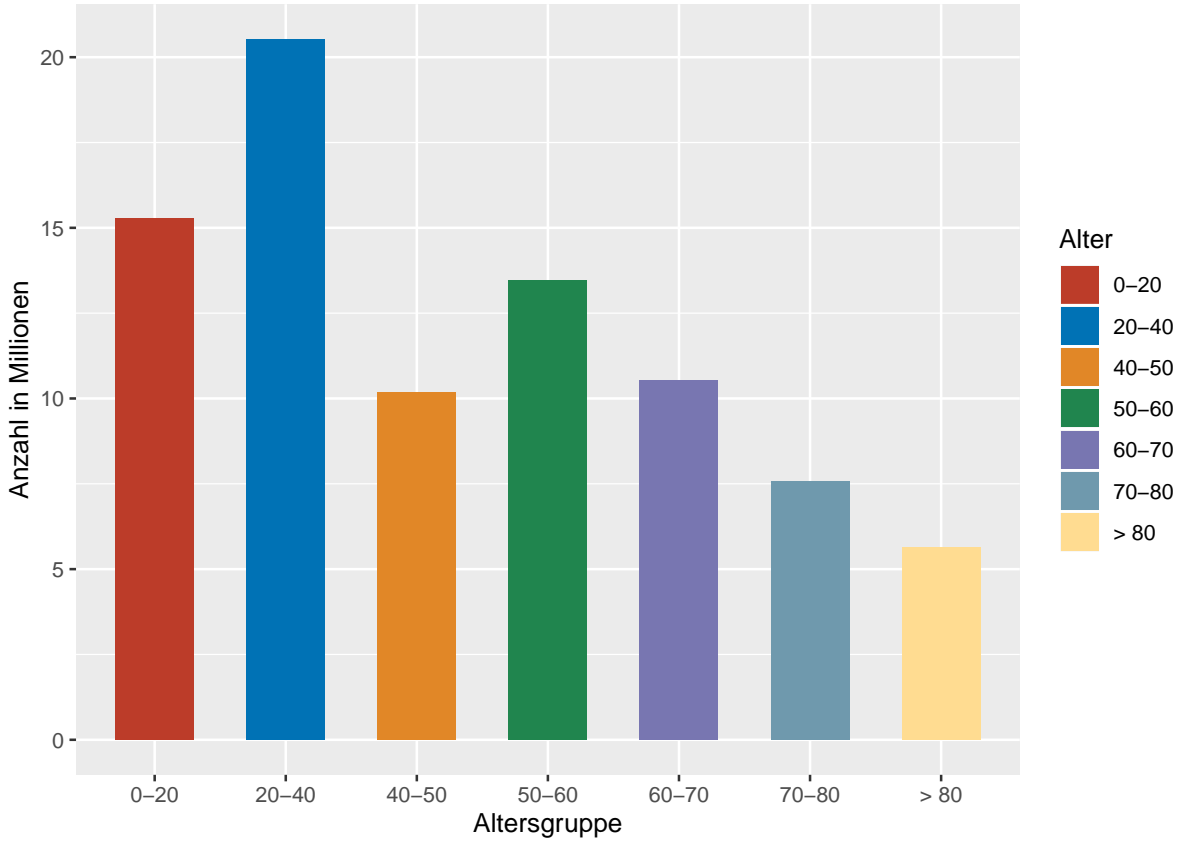


Abbildung 1: Bevölkerung in Deutschland 2019 aufgeteilt in Altersgruppen.

herangezogen. Zur einfacheren Vergleichbarkeit werden wir ebenfalls dieses Konfidenzintervall verwenden, um die statistische Unsicherheit bei den Letalitäten auszudrücken. Wir müssen davon ausgehen, dass es sich hierbei um eine konservative Betrachtung handelt. Die sogenannte Überdeckungswahrscheinlichkeit dieses Intervalls ist gerade bei kleinen Wahrscheinlichkeiten größer als 95%; d.h., das Intervall ist etwas zu lang (Brown, Cai, und DasGupta 2001).

Lässt man das Alter außer Acht, so ergibt sich eine kumulierte Letalität von 2.29 % (95%-KI: 2.15 - 2.43 %). Dies liegt etwas über der Einschätzung des Robert-Koch-Instituts, welches von einer Letalität von 1-2 % ausgeht (Wikipedia 2020). Diese Zahlen liegen auch einer aktuellen Simulationsstudie zur Reduktion von Sozialkontakten und deren Auswirkung auf die Pandemie sowie die Herdenimmunität für Polen und Deutschland zu Grunde (Bock u. a. 2020, Table 1), auf deren Ergebnisse wir später noch etwas eingehen werden.

Erwartete Anzahl von Toten

Wir berechnen die Anzahl der zu erwartenden Toten auf Basis der aktuellen Bevölkerungszahlen, der altersabhängigen Letalitäten aus China und der Annahme einer altersunabhängigen Durchseuchung zwischen 15 % und 85 %. Die Berechnung basiert auf folgender einfacher Formel

$$\text{Erwartete Anzahl Tote} = N_{\text{Altersgruppe}} \times D \times L_{\text{Altersgruppe}}$$

Hiebei steht $N_{\text{Altersgruppe}}$ für die Anzahl der Personen in der jeweiligen Altersgruppe, $L_{\text{Altersgruppe}}$ für die Letalität in dieser Altersgruppe und D für die altersunabhängige Durchseuchungsrate. Folglich entspricht $N_{\text{Altersgruppe}} \times D$ der erwarteten Anzahl von Erkrankten in der jeweiligen Altersgruppe und $N_{\text{Altersgruppe}} \times L_{\text{Altersgruppe}}$ der erwarteten Anzahl von Toten in der jeweiligen

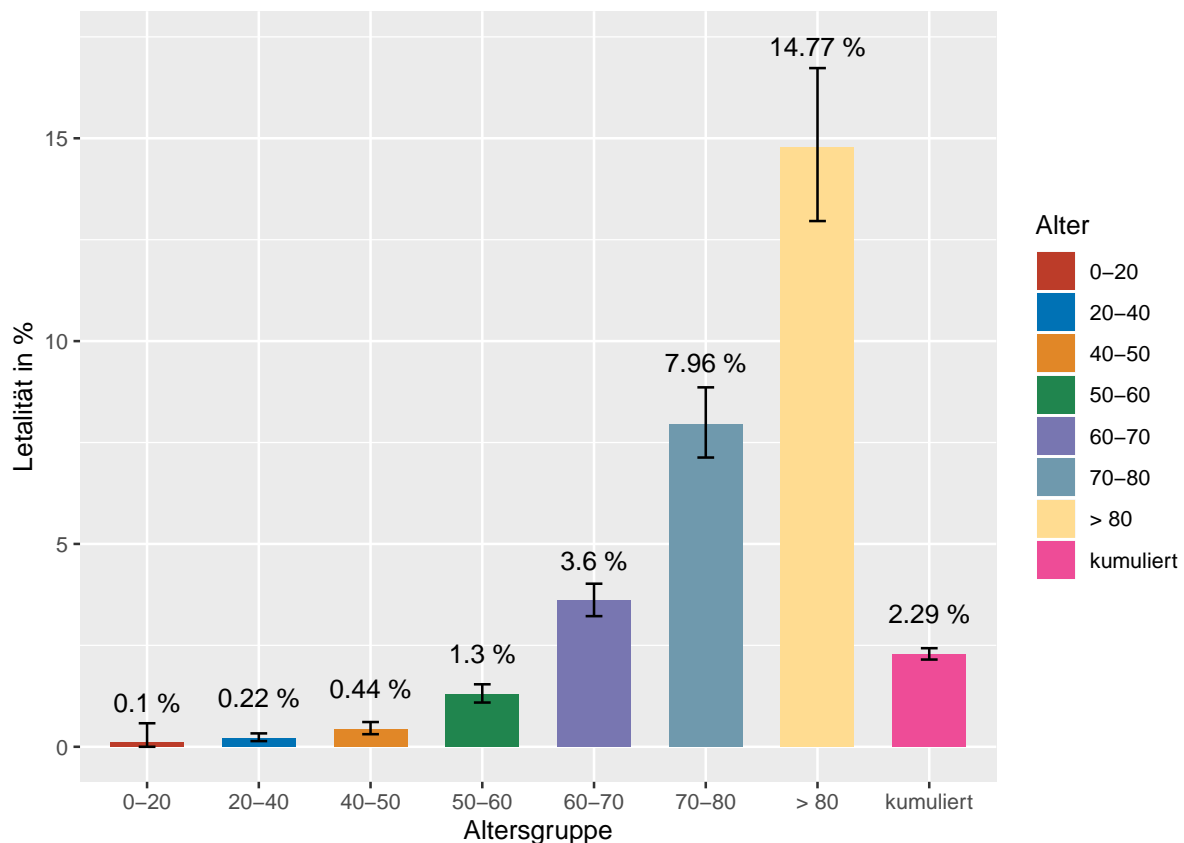


Abbildung 2: Abb. 2: Letalitäten in China vom 01.01.2020 bis zum 11.02.2020 (inkl. 95%-KI).

Altersgruppe bei einer Durchseuchung von 100%. Wir verwenden neben den beobachteten Letalitäten auch die unteren und oberen Grenzen der 95%-Konfidenzintervalle und berechnen hieraus sogenannte (punktweise) 95%-Konfidenzbänder.

Wir sehen anhand des sehr breiten Konfidenzbandes, dass gerade für junge Menschen bis zu einem Alter von 20 Jahren die Vorhersage der erwarteten Anzahl von Toten noch recht schwierig ist. Dies liegt daran, dass für diese Altersgruppe die Letalitätsschätzung noch recht ungenau ist.

Gehen wir davon aus, dass ohne entsprechend strikte Maßnahmen eine Durchseuchung von 60-70% nötig sein wird, um die Pandemie zu stoppen, so ergibt unsere obige Berechnung, dass wir mit ca. 1.26-1.47 Millionen Toten in Deutschland rechnen müssen (Drosten 2020). Zum Vergleich: in Deutschland starben 2018 und 2019 jeweils ca. 950.000 Menschen (DESTATIS: Statistisches Bundesamt 2020b). Gelänge es hingegen durch strikte Maßnahmen eine Durchseuchung von nur 15% oder weniger zu erreichen, was laut einer aktuellen Simulationsstudie durch eine sehr starke Reduktion der Sozialkontakte möglich sein könnte, so könnte dies alleine in Deutschland mehr als 1 Million Menschenleben retten (Bock u. a. 2020).

Dunkelziffer

Wir vergleichen die Ergebnisse unserer Berechnungen mit den aktuellen Zahlen vom 06.04.2020 (Worldometers.info 2020). Es gab bis zu diesem Datum 103.375 bekannte Fälle von COVID-19 in Deutschland. Davon sind 36081 bereits wieder gesund und 1810 verstorben, was 37891 abgeschlossenen und 65484 aktiven Fällen entspricht. Kombinieren wir die Zahlen zu den abgeschlossenen Fällen mit den Letalitäten aus China und nehmen an, dass die Verteilung der

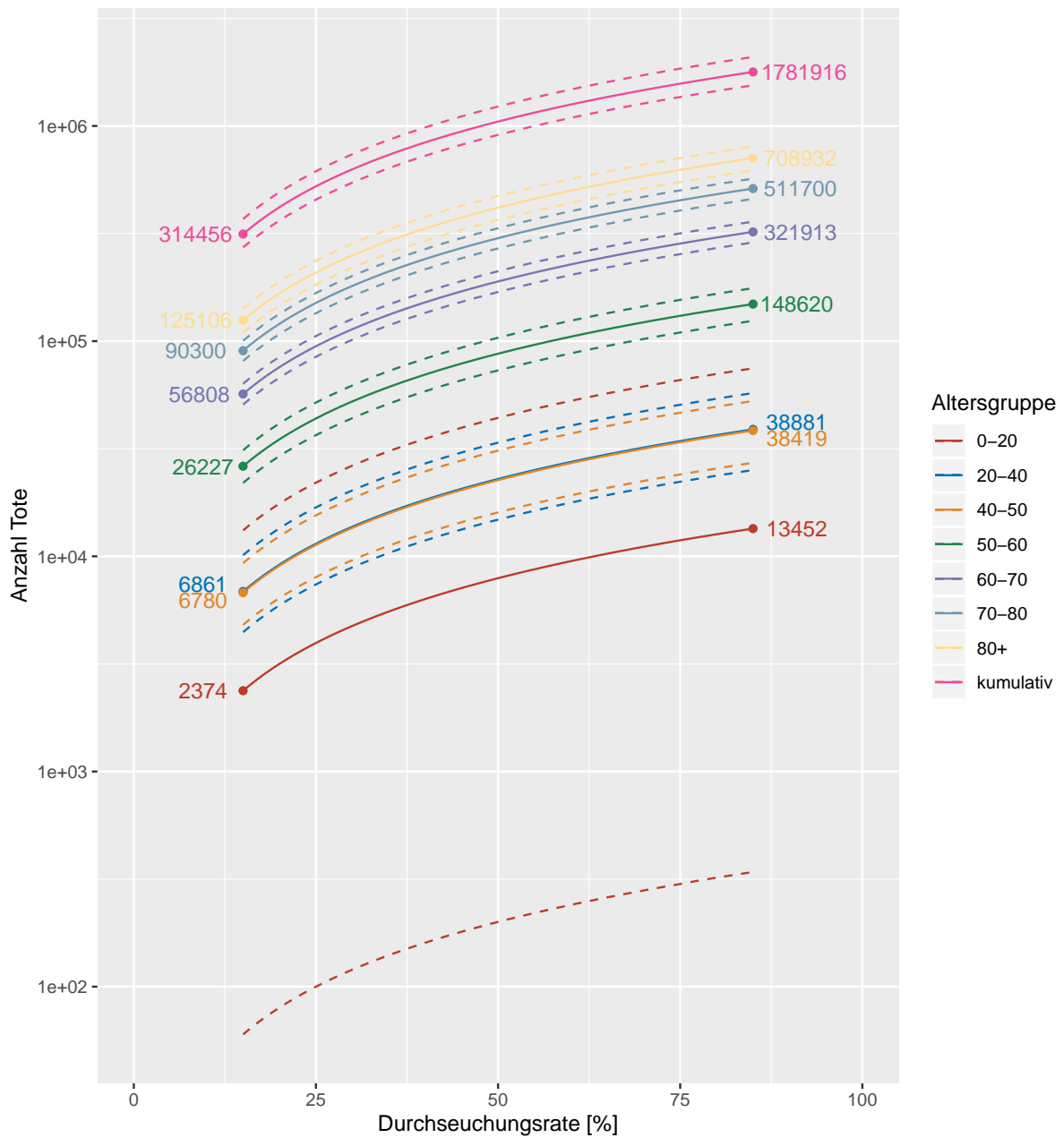


Abbildung 3: Abb. 3: Erwartete Anzahl von Toten in Deutschland in Abhängigkeit von der Durchseuchung (inkl. 95%-Konfidenzbänder).

Fälle auf die Altersgruppen der Verteilung der Altersgruppen in der Bevölkerung entspricht, da hierzu keine genauen Zahlen vorliegen, so können wir mit der oben angegebenen Formel die erwartete Anzahl von Toten je Altersgruppe berechnen.

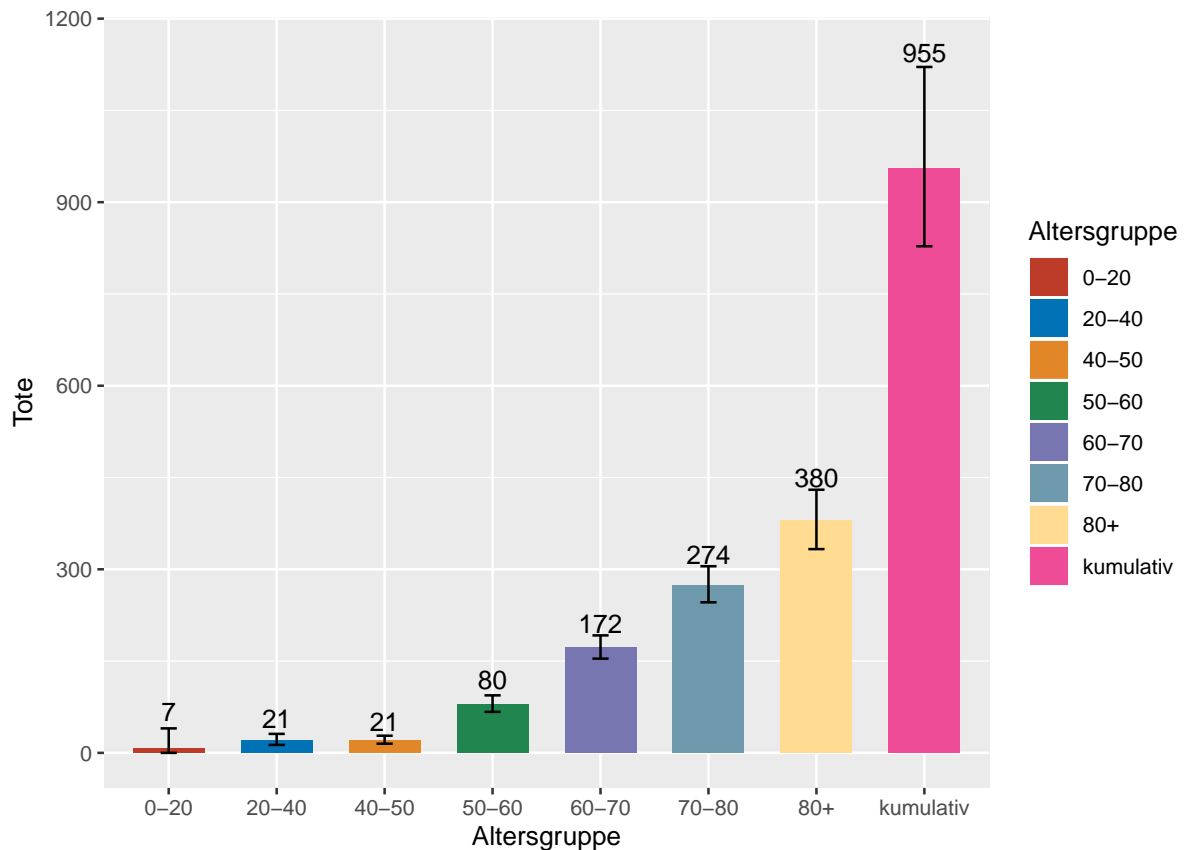


Abbildung 4: Abb. 4: Erwartete Anzahl von Toten pro Altersgruppe in Deutschland zum 04.04.2020 (inkl. 95%-KI).

Demnach sollten sich unter den abgeschlossenen Fällen 955 Tote (95%-KI: 828 - 1121) befinden. Ein Grund für den Unterschied zwischen den beobachteten und den berechneten Zahlen könnte neben der nicht genau bekannten Letalität und Verteilung der abgeschlossenen Fälle auf die Altersgruppen eine entsprechende Dunkelziffer sein. Gehen wir weiter davon aus, dass die Letalitäten von China auch auf Deutschland zutreffen, die Verteilung der abgeschlossenen Fälle auf die Altersgruppen der Verteilung der Altersgruppen in der Bevölkerung entspricht und eine altersunabhängige Durchseuchungsrate vorliegt, so können wir aus der aktuellen Anzahl von Toten die tatsächliche aktuelle Durchseuchungsrate berechnen. Es gilt

$$D_{\text{aktuell}} = \frac{\text{Anzahl Tote aktuell}}{\text{Erwartete Anzahl Tote insgesamt}}$$

Die erwartete Anzahl von Toten insgesamt erhält man, indem man die erwartete Anzahl von Toten für jede Altersgruppe bei 100% Durchseuchung summiert; d.h.,

$$\text{Erwartete Anzahl Tote insgesamt} = \sum_{\text{Altersgruppen}} N_{\text{Altersgruppe}} \times L_{\text{Altersgruppe}}$$

Dies führt auf eine geschätzte Durchseuchung von ca. 863 pro Million (95%-KI: 735 - 996). Sollte dies zutreffen, so ergeben sich nicht 37891, sondern bereits 71835 abgeschlossene COVID-19 Fälle (95%-KI: 61178 - 82855) in Deutschland. Dies würde bedeuten, dass bisher ca. die Hälfte

(95%-KI: 45.7 - 61.2 %) der COVID-19 Fälle in Deutschland auch entdeckt wurden. Dies deckt sich sehr gut mit der Einschätzung des Robert-Koch-Instituts, welches davon ausgeht, dass etwa die Hälfte der Fälle durch Tests erfasst werden (Wikipedia 2020).

Diskussion

Da die genaue Letalität von COVID-19 nicht bekannt ist (Wikipedia 2020), ist die erwartete Anzahl von Toten entsprechend ungewiss. Die obigen Berechnungen legen jedoch nahe, dass durch Maßnahmen, welche die Durchseuchungsrate deutlich reduzieren, alleine in Deutschland hundertausende Menschenleben gerettet werden könnten. Unklar ist jedoch, ob und in welchem Umfang die dafür nötigen strengen Maßnahmen, sei es direkt oder indirekt, ebenfalls Menschenleben kosten könnten.

Es werden jedoch nicht nur sehr viele Menschen durch diese Pandemie ihr Leben verlieren, sondern auch sehr viele Menschen Folgeschäden davontragen. Aktuell geht man davon aus, dass das SARS-CoV-2 Virus neben Lungenschäden auch Schädigungen des Herz-Kreislauf-Systems und des Gehirns verursacht (Zhang u. a. 2020; Madjid u. a. 2020; Li, Bai, und Hashikawa 2020). Die Häufigkeit und Stärke solcher Folgeerkrankungen ist im Moment jedoch noch weitgehend unbekannt. Wir können aber trotzdem davon ausgehen, dass länger andauernde strikte Maßnahmen nicht nur sehr viele Menschen vor dem Tod bewahren, sondern auch sehr viele vor gesundheitlichen Folgeschäden schützen würden. Je nach Umfang und Stärke dieser Schäden könnten diese eine nicht unerhebliche zukünftige Belastung für das Gesundheitssystem darstellen und einen entsprechenden ökonomischen Folgeschaden bedeuten.

Literatur

Bock, Wolfgang, Barbara Adamik, Marek Bawiec, Viktor Bezborodov, Marcin Bodych, Jan Pablo Burgard, Thomas Goetz, u. a. 2020. „Mitigation and herd immunity strategy for COVID-19 is likely to fail“. *medRxiv*. Cold Spring Harbor Laboratory Press. <https://doi.org/10.1101/2020.03.25.20043109>.

Brown, Lawrence D., T. Tony Cai, und Anirban DasGupta. 2001. „Interval Estimation for a Binomial Proportion“. *Statistical Science* 16 (2). Institute of Mathematical Statistics: 101–17. <http://www.jstor.org/stable/2676784>.

Clopper, C. J., und E. S. Pearson. 1934. „The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial“. *Biometrika* 26 (4). [Oxford University Press, Biometrika Trust]: 404–13. <http://www.jstor.org/stable/2331986>.

DESTATIS: Statistisches Bundesamt. 2020a. „14. koordinierte Bevölkerungsvorausberechnung für Deutschland. Variante 1: Moderate Entwicklung bei niedrigem Wanderungssaldo (G2L2W1)“. DESTATIS: Statistisches Bundesamt, letzter Zugriff: 05.04.2020. <https://service.destatis.de/bevoelkerungspyramide/>.

———. 2020b. „Sterbefälle und Lebenserwartung. 2019: Vorläufige Schätzung.“ DESTATIS: Statistisches Bundesamt, letzter Zugriff: 05.04.2020. <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/vorlaeufige-ergebnisse.html>.

Drosten, Christian. 2020. „Chef-Virologe der Charité sicher: 70 Prozent der Deutschen werden Corona bekommen“. *Berliner Zeitung* 01.03.2020, aktualisiert 01.04.2020, letzter Zugriff: 05.04.2020. <https://www.bz-berlin.de/berlin/70-prozent-der-deutschen-werden-corona-bekommen>.

Kohl, Matthias. 2019. *MKinfer: Inferential Statistics*. <http://www.stamats.de>.

- Li, Y. C., W. Z. Bai, und T. Hashikawa. 2020. „The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients“. *J. Med. Virol.*, Februar. <https://doi.org/10.1002/jmv.25728>.
- Madjid, Mohammad, Payam Safavi-Naeini, Scott D. Solomon, und Orly Vardeny. 2020. „Potential Effects of Coronaviruses on the Cardiovascular System: A Review“. *JAMA Cardiology*, März. <https://doi.org/10.1001/jamacardio.2020.1286>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rizzello, Claudio. 2020. „Wir führen Krieg gegen einen Feind, der unsere Gewohnheiten ausnutzt. Interview mit Roberto Burioni“. *Zeit Online* 11.03.2020, letzter Zugriff: 05.04.2020. <https://www.zeit.de/wissen/gesundheit/2020-03/roberto-burioni-italien-virologe-coronavirus>.
- Sievert, Carson. 2018. *plotly for R*. <https://plotly-r.com>.
- Verity, R., L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, u. a. 2020. „Estimates of the severity of coronavirus disease 2019: a model-based analysis“. *Lancet Infect Dis*, März. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7).
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wikipedia. 2020. „COVID-19-Pandemie“. Wikipedia, Die freie Enzyklopädie, letzter Zugriff: 05.04.2020. <https://de.wikipedia.org/wiki/COVID-19-Pandemie>.
- Worldometers.info. 2020. „COVID-19 Coronavirus Pandemic“. *Worldometers.info* 05.04.2020, letzter Zugriff: 05.04.2020. <https://www.worldometers.info/coronavirus/country/germany/>.
- Wu, Zunyou, und Jennifer M. McGoogan. 2020. „Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention“. *JAMA*, Februar. <https://doi.org/10.1001/jama.2020.2648>.
- Xiao, Nan. 2018. *ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'*. <https://CRAN.R-project.org/package=ggsci>.
- Xie, Yihui. 2015. *Dynamic Documents with R and knitr*. 2nd Aufl. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- Xie, Yihui, J.J. Allaire, und Garrett Golemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.
- Zhang, P., J. Li, H. Liu, N. Han, J. Ju, Y. Kou, L. Chen, u. a. 2020. „Long-term bone and lung consequences associated with hospital-acquired severe acute respiratory syndrome: a 15-year follow-up from a prospective cohort study“. *Bone Res* 8: 8. <https://doi.org/10.1038/s41413-020-0084-5>.