

ZUSAMMENFASSUNG

PRAXISPROJEKT

Statistik Bachelor, LMU München, Institut für Statistik

ANALYSE DER COVID-19 FALLZAHLEN

Regina Galambos

Lorenz Mihatsch



München, den 08.05.2020

Projektpartner und Betreuer
Dr. André Klima

1 Einleitung

COVID-19 ist eine Atemwegserkrankung, die durch das SARS-CoV-2 Virus verursacht wird und erstmals im Dezember 2019 in Wuhan, China, dokumentiert wurde. Seitdem hat sich das Virus global verbreitet und wurde am 11. März 2020 von der WHO als Pandemie eingestuft. Die vorliegende Analyse widmet sich der Beschreibung der Entwicklung der weltweit dokumentierten Fall- und Todeszahlen weichen dem 21. Januar und dem 27. April 2020.

2 Datengrundlage

Die Auswertung beruht auf drei verschiedenen Datenquellen. Der Datensatz zu den COVID-19 Fall- und Todeszahlen der einzelnen Länder wird täglich von *RamiKrispin* vom *Center of Systems Science and Engineering* der Johns Hopkins University abgerufen und in Form eines R-packages zu Verfügung gestellt [1]. Für die Populationsdaten des Jahres 2018 und die Kontinentzugehörigkeit wurden die Datenbanken der UN und der World Bank genutzt, die mit Hilfe der R-packages *wbstat* heruntergeladen wurden, sowie ein Datensatz von *JohnSnowLabs* [2] verfügbar auf DataHub.io. Informationen zu politischen Maßnahmen einzelner Länder zur Eindämmung der Infektionen wurden über den *Government Response Tracker* Datensatz der *University of Oxford* [3] bezogen.

3 Ergebnisse

Für die Analyse der Fall- und Todeszahlen wurden durchweg die kumulativen Fall- und Todeszahlen pro 100.000 Einwohner verwendet. Weiter sei an dieser Stelle anzumerken, dass es sich “nur” um die aufgezeichneten Fälle und Todesfälle handelt. Damit sind die gegebenen Daten stark von der Aufzeichnungs- und Testpolitik der einzelnen Länder abhängig. Ein direkter Ländervergleich ist somit nur eingeschränkt möglich und man muss davon ausgehen, dass hier tendenziell eher eine untere Schranke der tatsächlich vorliegenden Fall- und Todeszahlen analysiert wird.

3.1 Entwicklung der weltweiten kumulativen Fall- und Todeszahlen

Zum Endpunkt der Analyse (27. April 2020) zeigen sich weltweit etwa 40 bestätigte kumulative Fälle und etwa 3 Todesfälle pro 100.000 Einwohner.

Geographisch scheinen vor allem Westeuropa und die USA besonders schwer betroffen zu sein. In Afrika hingegen sind kaum gemeldete Fälle dokumentiert. Durch eine logarithmische Darstellung sind in der Vergangenheit zwei Infektionswellen erkennbar. Die erste dieser Wellen ging von Beginn der Aufzeichnung bis etwa Mitte Februar und spielte sich insbesondere in China ab, die Zweite dann, durch eine weltweite Verbreitung, ab etwa Mitte März bis Ende April.

3.2 Wachstumsfaktoren und Verdopplungszeit

Unter der Annahme eines exponentiellen Wachstums lassen sich die kumulativen Fall- und Todesfallzahlen über die Berechnung von Wachstumsfaktoren und Verdopplungszeiten direkt miteinander vergleichen. Auch hier lassen sich wieder die zwei Infektionswellen erkennen. Zwischen den Infektionswellen trat sowohl bei den kumulativen Fall- als auch bei den Todesfallzahlen ein lokales Minimum der weltweiten Wachstumsfaktoren (bzw. Maximum der Verdopplungszeit) auf. Dieses Minimum (bzw. Maximum) lag bei den Fallzahlen (min. Wachstumsfaktor 1,01; max. Verdopplungszeit 71 Tage) etwa zwei Wochen vor dem Minimum (bzw. Maximum) der kumulativen Todesfälle (min. Wachstumsfaktor 1,02; max. Verdopplungszeit 31 Tage). Der Gipfel der zweiten Infektionswelle trat sowohl bei den kumulativen Fällen als auch bei den Todesfällen zeitgleich Mitte-Ende März auf. Zum Ende der Analyse sind die Wachstumsfaktoren der weltweiten kumulativen Fälle und Todesfälle rückläufig und die Verdopplungszeiten steigend.

4 Danksagung

Wie möchten uns herzlich bei Dr. André Klima für das Projekt und die Betreuung in diesem Projekt bedanken. Weitere (interaktiven) Grafiken sind in unserer Webapplication unter <https://covid-19-stats-project.herokuapp.com> zu finden. Wir danken weiterhin *RamiKrispin*, der *Oxford University*, den UN und der World Bank für die öffentliche Bereitstellung der Datensätze.

Datenquellen

- [1] RamiKrispin GitHub. <https://github.com/RamiKrispin/coronavirus>. Eingesehen am 07.02.2013.
- [2] JohnSnowLabs. <https://datahub.io/JohnSnowLabs/country-and-continent-codes-list>. Eingesehen am 07.02.2013.
- [3] Government Response Tracker. <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>. Eingesehen am 07.02.2013.

Verwendete R Pakete

- [4] Vincent Arel-Bundock, Nils Enevoldsen, and CJ Yetman. countrycode: An r package to convert country names and country codes. *Journal of Open Source Software*, 3(28):848, 2018.
- [5] Simon Garnier. *viridis: Default Color Maps from 'matplotlib'*, 2018. R package version 0.5.1.
- [6] Garrett Golemund and Hadley Wickham. Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25, 2011.
- [7] Robert J. Hijmans. *raster: Geographic Data Analysis and Modeling*, 2020. R package version 3.0-12.
- [8] Rami Krispin. *coronavirus: The 2019 Novel Coronavirus COVID-19 (2019-nCoV) Dataset*, 2020. R package version 0.2.0.
- [9] Jesse Piburn. *wbstats: Programmatic Access to the World Bank API*. Oak Ridge National Laboratory, Oak Ridge, Tennessee, 2018.
- [10] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.
- [11] William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2019. R package version 1.9.12.
- [12] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

- [13] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2020. R package version 0.8.5.
- [14] Hadley Wickham and Lionel Henry. *tidyr: Tidy Messy Data*, 2020. R package version 1.0.2.
- [15] Achim Zeileis and Gabor Grothendieck. zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, 14(6):1–27, 2005.

Verwendete Python Pakete

- [16] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- [17] Kelsey Jordahl, Joris Van den Bossche, Jacob Wasserman, James McBride, Jeffrey Gerard, Jeff Tratner, Matthew Perry, Carson Farmer, Micah Cochran, Sean Gillies, Matt Bartos, Lucas Culbertson, Nick Eubank, maxalbert, Martin Fleischmann, Geir Arne Hjelle, Dani Arribas-Bel, Christopher Ren, Sergio Rey, Martin Journois, Levi John Wolf, Aleksey Bilogur, Nick Grue, Joshua Wilson, YuichiNotoya, Leah Wasser, Filipe, Chris Holdgraf, Adam Greenhall, and John Trengrove. geopandas/geopandas: v0.4.1, March 2019.
- [18] The pandas development team. pandas-dev/pandas: Pandas (version 0.25.3), February 2020.
- [19] S. van der Walt, S. C. Colbert, and G. Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering*, 13(2):22–30, 2011.
- [20] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.