

Corona- Statistik für Anfänger

Warum die Corona- Pandemie auf banalen statistischen Fehlern beruht – und was dies mit PISA zu tun hat.

Die PISA- Studie lehrt uns, dass die mathematischen Kenntnisse Deutscher Schüler 2018 deutlich unter den in 2012 gemessenen Werten lagen; lediglich 13% Deutscher Schüler*innen sind in der Lage, komplexe mathematische Probleme zu verstehen und Problemlösungen zu finden – in asiatischen Ländern liegt dieser Prozentsatz bei etwa 44%!

Eine Pandemie ist nicht nur ein medizinisches sondern ein grundlegendes mathematisches Problem, in dem es im einfachsten Fall darum geht eine Kohorte bestehend aus prinzipiell ansteckungsfähigen Personen, infizierten Personen und von der Infektion genesenen Personen im zeitlichen Verlauf zu beschreiben – kein banales Problem.

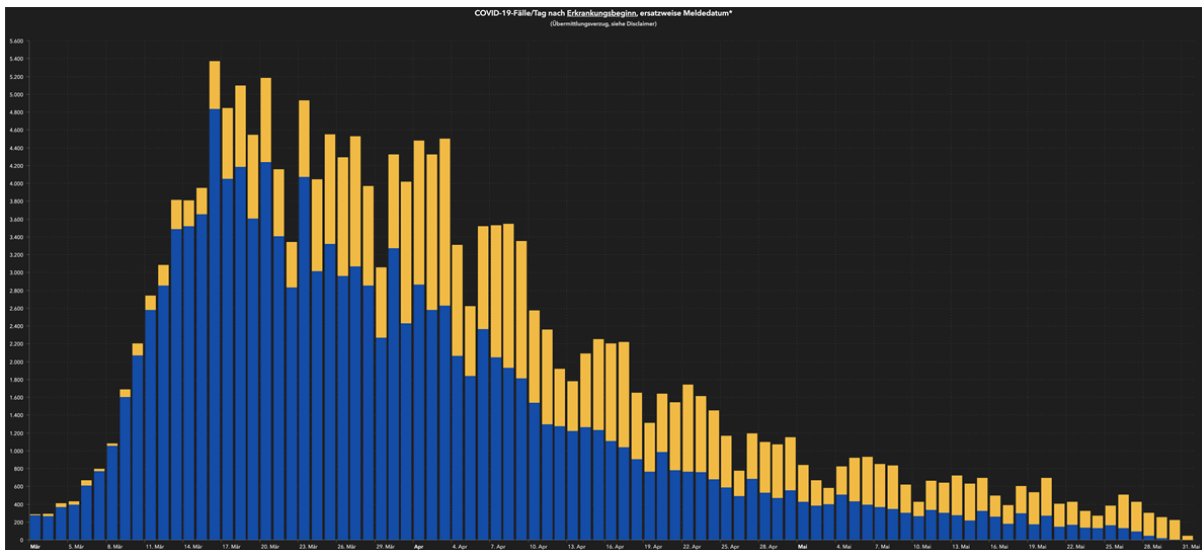
Allem vorangestellt ist jedoch eine einfache beschreibende Statistik, die die Verbreitung einer Infektion darstellen soll. Die Verbreitungsgeschwindigkeit eine Infektionskrankheit einerseits sowie die Schwere der durch sie ausgelösten Erkrankung charakterisieren eine Epidemie und sind Grundlage für die erforderlichen Schutzmaßnahmen.

Die unten abgebildete Darstellung der COVID-19- Epidemie durch das Robert Koch Institut ist ein Beispiel für eine solche Herangehensweise – ein sehr schlechtes Beispiel!

Grundlegendes Problem einer solchen Epidemiebeschreibung ist die Tatsache, dass nicht alle Individuen einer Grundgesamtheit (Kohorte) – z.B. 83.000.000 Deutsche Bürger*innen -täglich auf das Vorhandensein einer Infektionserkrankung getestet werden können, um den Verlauf einer epidemischen Infektionskrankheit zu beschreiben. Stattdessen wird eine die Grundgesamtheit z.B. hinsichtlich Geschlecht, Alter, bestehende Erkrankungen etc. möglichst exakt wiederspiegelnde Stichprobe täglich untersucht. Als Aufgreifkriterium für eine solche Stichprobe hat das RKI das Vorhandensein einer Atemwegserkrankung in Verbindung mit einem stattgehabten Kontakt zu einem SARS-CoV-2- Infizierten festgeschrieben – das ist völlig korrekt. Im weiteren Verlauf hat das RKI diese Kriterien geändert und hat zuletzt gesunde Arbeiter auf Schlachthöfen getestet, die in engen Wohncontainern und unter schlechten Hygienebedingungen einem hohen Infektionsrisiko ausgesetzt sind – das verfälscht die Stichprobe grob, so dass sie nicht mehr als Abbild der Grundgesamtheit angesehen werden kann.

Die Zahlen spiegeln allein aus diesem Grund nicht die Pandemie in der Kohorte (=Grundgesamtheit aller Deutschen) wider – sondern das Infektionsgeschehen dieser Stichprobe! Kardinalfehler der vorliegenden RKI- Statistik ist jedoch ein anderer: Solange davon auszugehen ist, dass die Dunkelziffer der Infektionen in der Kohorte deutlich höher ist als die Anzahl der gefundenen Infektionen/Erkrankungen in der Stichprobe ist der Verlauf der Pandemie abhängig von der Anzahl der durchgeführten Tests und von dem Vorhandensein konkurrierender Krankheitserreger, die eine ähnliche Krankheitssymptomatik verursachen und somit die Eingangskriterien in die Stichprobe beeinflussen.

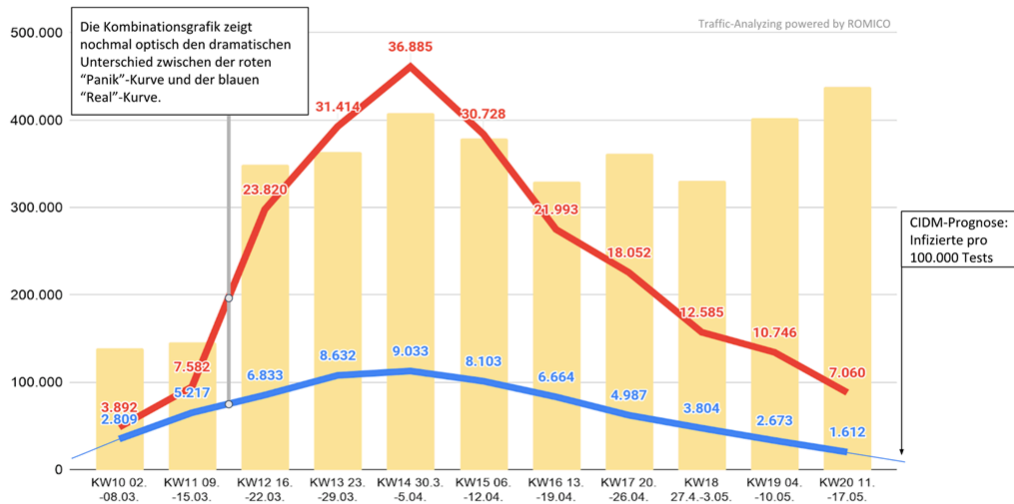
Am Beispiel Deutschland:



Den oben abgebildeten exponentiellen Anstieg der Neuinfektionen in der Stichprobe gab es in der Kohorte nie! Er ist im Wesentlichen durch die Zunahme der durchgeführten Tests erklärt. Am einfachen Beispiel bedeutet dies:

Testet das RKI am Tag T 10.000 Personen, von denen 500 Personen positiv getestet werden (Infektionsrate = 5%) und am Tag T+1 20.000 Personen von denen 1.000 positiv getestet werden (Infektionsrate = 5%) und am Tag T+2 40.000 Personen, von denen 2.000 positiv getestet werden (Infektionsrate = 5%) so verkündet das RKI am Tag T 500 Infizierte, am Tag T+1 1000 Infizierte und am Tag T+2 2.000 Infizierte (ohne Nennung der Testanzahl) und behauptet (implizit – ohne dies ausdrücklich auszusprechen) somit einen exponentiellen Anstieg der Infiziertenzahlen in der Grundgesamtheit aller Deutschen, obwohl im vorliegenden Beispiel keine einzige Neuinfektion hinzugekommen ist – die Infiziertenrate lag an allen Tagen bei 5%! Dieses Vorgehen kann man nur als bewusste Irreführung bezeichnen – denn es ist nicht vorstellbar, dass diese einfachen Zusammenhänge im RKI (vielleicht den Präsidenten des RKI ausgenommen) nicht bekannt sind.

Aktuelle Corona-Statistik Deutschland 6/7: Fälle pro Woche im Verhältnis 3/3
Anzahl Testungen (IST*), Infizierte absolut (IST*), Infizierte pro 100.000 Tests (IST)**



CIDM 2020 *RKI-Lagebericht **RKI-Lagebericht inkl. Glättung für KW10+11, da Labormeldungen nicht ganz vollständig

Berücksichtigt man diese Zusammenhänge und bezieht man die Anzahl der Neuinfektionen auf je 100.000 Tests kommt die oben abgebildete Grafik der Corona- Initiative Deutscher Mittelstand (CIDM) zustande. Die blaue Kennlinie beschreibt den wahren Verlauf der Pandemie in der Kohorte sehr viel wahrheitsgetreuer als die völlig verzerrte obige RKI- Grafik. Von einem exponentiellem Epidemieverlauf ist nichts mehr zu sehen! Berücksichtig man des Weitern, dass im Laufe der Zeit konkurrierende Erreger einer Atemwegsinfektion (Eingangskriterium für Stichprobe) weggefallen sind und somit die Wahrscheinlichkeit eines SARS-CoV-2- Nachweises in der Stichprobe deutlich ansteigt (ohne dass dies eine Zunahme der Infektionen in der Kohorte bedeutet), dann stellt sich das Abflauen der „Pandemie“ in Wahrheit noch drastischer dar.

Um die Fallzahlen in der Stichprobe weiter hoch zu halten und eine schlimme Pandemie in der Kohorte vorzutäuschen hat das RKI zuletzt die Eingangskriterien für die Stichprobe grob geändert und geht inzwischen in Schlachthöfe, um dort gesunde hart arbeitende Menschen auf SARS-CoV-2 zu testen – und siehe da es finden sich asymptomatische Infizierte (die auf engstem Raum in Wohncontainern unter miserablen hygienischen Bedingungen zusammenleben). Das kann man alles tun – man darf nur nicht behaupten, den Verlauf einer Pandemie in der Kohorte abzubilden! Siehe hierzu auch die unten angeführte Arbeit von C. Kuhbandner.

Inzwischen wissen wir außerdem wie es um die Spezifität der RT-PCR (Nachweis von SARS-CoV-2 im Nasen-Rachenabstrich) bestellt ist. Instand e.V. findet in den letzten sog. Ringversuchen zur RT-PCR (siehe die unten angeführte Anlage S. 20) eine Sensitivität von 99% sowie eine Spezifität von 92,4 – 98,6%. Die Sensitivität beschreibt hierbei die Anzahl der vom Test **korrekt gefundenen** Infizierten unter **allen Infizierten** – eine Testsensitivität von 99% besagt, dass der verwendete Test 990 von 1000 nachweislich infizierten Personen korrekt findet – 10 infizierte Personen beschreibt er fälschlicherweise aber als nicht infiziert. Die Testspezifität hingegen prüft den Anteil der **korrekt negativen** Testergebnisse unter den **nicht infizierten** Personen. Eine Testspezifität von 98,6% besagt, dass bei 1000 nicht infizierten Personen der Test in 986 Fällen tatsächlich negativ ausfällt; 14 tatsächlich nicht infizierte Personen erhalten jedoch ein positives Testergebnis!

In der 22. KW 2020 veröffentlichte das RKI 3.152 Neuinfektionen unter 392.437 Tests. Trägt man diese Ergebnisse in eine sog. Vierfeldertafel ein (s.u.), so finden sich unter 392.437 getesteten 3.152 Infizierte und 389.285 nicht Infizierte (re. Spaltensumme). Bei einer Sensitivität von 99% würde der Test 3.120 Infizierte korrekt finden – in 32 Fälle würde die Infektion übersehen.

Andererseits würden bei einer Spezifität von 98,6% unter 389.285 nicht infizierten Personen 5.450 fälschlicherweise als infiziert deklariert! Aus Sicht der getesteten Person bedeutet ein positives Testergebnis somit eine Wahrscheinlichkeit von 36% (Anzahl korrekt positiver Tests/ Anzahl aller positiven Tests) tatsächlich infiziert zu sein (sog. **Positiver prädiktiver Wert**). Eine Münze zu werfen ist genauer!! Der **negative prädiktive Wert** (Anteil der korrekt negativen Befunde an allen negativen Befunden) ist mit 99,9% allerdings sehr hoch.

Dies bedeutet auch, dass bei einer Zunahme der Testhäufigkeit – derzeit liegt die bundesweite Testkapazität bei etwa 1.000.000 Test/Woche - selbst für den Fall, dass es keine einzige echte Neuinfektion mehr gibt immer positive Tests gefunden werden. Werden wöchentlich 1.000.000 Tests durchgeführt so ist auf jeden Fall bei einer Testspezifität von bestenfalls 98,6% mit mindestens 13.956 positiven Testergebnissen zu rechnen – und so ließe sich die Pandemie dauerhaft aufrecht erhalten. Diese Zusammenhänge sollten insbesondere denjenigen Schulleitern bewusst sein, die die Idee verfolgen, Schüler 2x wöchentlich einem Abstrichtest zu unterziehen, um dann das Namensschild testnegativer Schüler mit einem grünen Punkt als „nichtinfiziert“ zu kennzeichnen. Unter den angeblich Infizierten werden bei dem derzeitigen Infektionsgeschehen 2 von 3 positiv getesteten Schülern in Wahrheit nicht infiziert sein!

	RT- PCR pos.	RT- PCR neg.	Summen
aktiv Infizierte	3.120	32	3.152
nicht Infizierte	5.450	383.835	389.285
Summe	8.570	383.867	392.437

Grün hinterlegt sind die korrekt positiven oder korrekt negativen Tests – rot hinterlegt sind die falsch positiven oder falsch negativen Tests. Es errechnet sich ein positiver Vorhersagewert von 36% ($3.120/8570 = 0,364$) bzw. ein negativer Vorhersagewert von 99,9% ($383.835/383.867 = 0,999$).

Fr. Priesemann und Mitarbeiter vom Max Planck Institut schießen m.E. aber den Vogel ab! Sie entwickeln ein mathematisches Pandemiemodell basierend auf dem SIR (Susceptible-infected-Recovered)- Modell der Epidemiologen sowie Bayes'scher Statistik und behaupten, den Verlauf der Pandemie damit beschreiben zu können, Wendepunkte des Pandemieverlaufs zu sehen und den jeweiligen weiteren Verlauf der Pandemie ab einem der identifizierten Wendepunkte vorhersagen zu können. Tatsächlich finden sich 3 Wendepunkte, die jeweils den Maßnahmen zur Bekämpfung der Pandemie durch unsere unfehlbare Regierung zugeordnet werden (mild distancing, strong distancing, contact ban). Das Modell belegt Fr. Priesemann zufolge, dass der „contact ban“ erforderlich war (grüne unterbrochene Linie), um die Pandemie abflauen zu lassen – und sie gibt inzwischen ganzseitige Interviews in der FAZ hierzu.

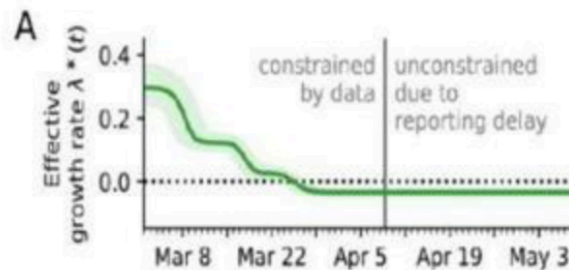
Die zugrunde liegenden Rohdaten sind natürlich die Daten des RKI ohne Berücksichtigung der obigen Anmerkungen zu den Testzahlen, zu konkurrierenden Infektionsursachen bzw. Änderung der Testkriterien. Was sie allenfalls beschreibt ist der Verlauf der RKI- Stichprobe – nicht jedoch den Verlauf der Pandemie in der zugrundeliegenden Kohorte! Ich habe Sie angeschrieben und vorgeschlagen, doch anstelle der RKI- Rohdaten, die oben aufgeführten CDIM- Daten oder die Kuhbandner- Daten zu nutzen – ich wettete sie würde es nicht tun!

Inzwischen hat die Gruppe um Fr. Priesemann tatsächlich auf die nicht nur von mir geäußerte Kritik reagiert und korrigierte Eingangsdaten verwendet. Ich entnehme einer Mitteilung von Prof. Stefan Homburg, dass nach Korrektur der Eingangsdaten, das mathematische Pandemiemodell jetzt völlig anders aussieht. Jetzt findet sich am 5.3.20 ein deutlicher Knick im Verlauf der Pandemieausbreitung – ab dem 9.3. findet sich keine relevante Änderung der Wachstumsrate mehr, die am 16.3.2020 gegen 0 tendiert und dann negativ wird! Fr. Priesemann, die als konsequente Befürworterin der behördlichen Seuchenpolitik alle Maßnahmen gutgeheißen hat korrigiert ihre Thesen somit selbst – so geht Wissenschaft – Chapeau! - Fr. Priesemann.

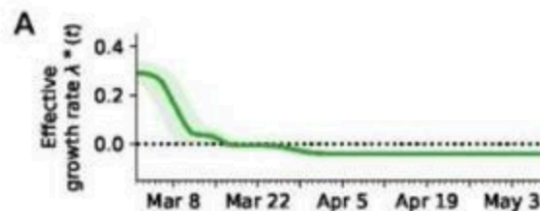
Prof. Dr. Stefan Homburg

8. Juni 2020

1. Dehning et al. (2020) haben in der renommierten Fachzeitschrift *Science* einen Artikel mit folgendem Hauptergebnis veröffentlicht (S. 4): „Das volle Ausmaß der Staatseingriffe war notwendig, um ein exponentielles Wachstum zu stoppen.“
2. Begründung in ihrer Abbildung 3A: Das Viruswachstum sinkt in drei Stufen 9. März (Absage Großveranstaltungen), 16. März (Schulschließungen), 23. März (Lockdown).



3. Dieses Ergebnis haben zahlreiche Wissenschaftler kritisiert. Dehning et al. haben u. a. die Kritik ungeeigneter Daten aufgegriffen und präsentieren das Ergebnis nunmehr in folgender Form (Technical Report, Abbildung 17A):



4. Ab dem 5. März sinkt die Virusausbreitung scharf. Am 9. März ist kein Knick im Verlauf zu erkennen. Am 16. März ist die Wachstumsrate schon Null und lange vor dem Lockdown vom 23. März negativ. Das neue Ergebnis entspricht exakt dem, was ich knapp zwei Monate zuvor anhand der R-Zahl gezeigt hatte: Der Lockdown war überflüssig. Neu ist, dass die Absage von Großveranstaltungen und die Schulschließungen ebenfalls überflüssig waren.

Quellen:

science.sciencemag.org/content/early/2020/05/14/science.abb9789/tab-e-letters

github.com/Priesemann-Group/covid19_inference_forecast/blob/master/technical_notes.md

Ursprüngliches Schaubild:

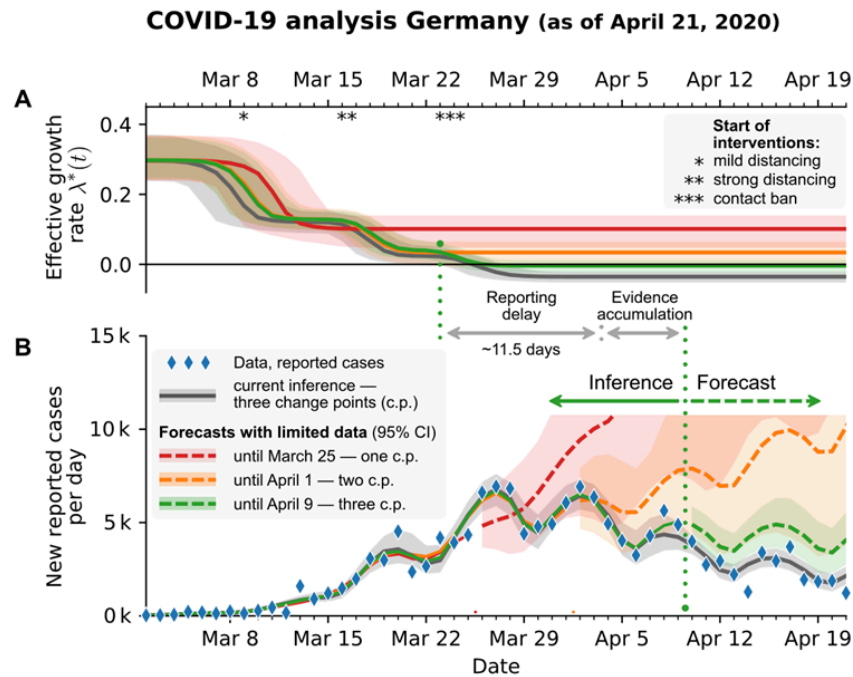


FIG. 0. Bayesian analysis of the German COVID-19 data (blue diamonds) reveals three change points that match the timing of publicly announced interventions. **A:** The inferred effective growth rate (difference between spreading and recovery rate, $\lambda^* = \lambda - \mu$) for an SIR model with weekly reporting modulation and reporting delay that includes scenarios with one, two or three change points (red, orange, green; fitted to case reports until March 25, April 1 and April 9, respectively). The timing of the inferred change points in growth rate is consistent with the timing of German governmental interventions (depicted as *, **, and ***). **B:** Comparing inferred models with the actual new reported cases per day reveals the effectiveness of governmental interventions. After the first two interventions, the number of new cases still grew exponentially (red, orange); only after the third intervention did the number of new cases start to saturate (green) or even to decline (gray, data until April 21). This illustrates that the future development strongly depends on our distancing behavior. Note the delay between a change point and the observation of changes in the number of new cases of almost two weeks — a combination of reporting delay and a minimal period of evidence accumulation.

Mittlerweile sagt sogar unser Topvirologe Drosten, dass der Pandemieverlauf wohl hauptsächlich durch „Superspreader“ oder „Superspreading Events“ gekennzeichnet ist (und damit natürlich durch Bayes'sche Normalverteilungsstatistik nicht beschrieben werden kann), so dass allein das Verbot von Großveranstaltungen einen Einfluss auf den Pandemieverlauf haben könnte. Dies wiederum ist spätestens seit dem 15.4.2020 nach Veröffentlichung des R-Wert-Verlaufs im Tagesbericht des RKI bekannt – interessante Anmerkung: aus dem Tagesbericht vom 15.4.20 ist diese Darstellung verschwunden – sie findet sich aber versteckt im epidemiologischen Bulletin 17 2020 – honi soit qui mal y pense!

Zum Reproduktionswert:

Erstmals veröffentlicht das RKI im Lagebericht vom 15.4.2020 den bisherigen Verlauf des R- Wertes und markiert die bekannten Maßnahmen zur Infektionsbekämpfung: Verbot von Großveranstaltungen, die Schließung von Kindertagesstätten und Schulen etc. sowie das generelle Kontaktverbot ab dem 23.3.2020. Auf die Bedeutung des R- Wertes sei an dieser Stelle nicht explizit eingegangen – wohl aber auf seine Berechnung durch das RKI.

Zunächst einmal beschreibt das RKI mit dem R- Wert den Verlauf der COVID-19- Erkrankung anhand der Neuerkrankungen (nicht Neuinfektionen!) in zwei unmittelbar aufeinanderfolgenden 4-Tageszeiträumen. Das RKI bildet hierbei die Summe der Neuerkrankungen in den ersten 4 Tagen sowie die Summe der Neuerkrankungen im 2. Viertageszeitraum. Die Summe der Neuerkrankungen im 2. Viertageszeitraum wird dann durch die Summe der Neuerkrankungen im 1. Viertageszeitraum geteilt um am 8. Tag dieser Rechnungsperiode eingetragen

Der so berechnete R- Wert beschreibt also das Krankheitsgeschehen vier Tage zuvor, worauf das RKI ausdrücklich hinweist. Die Maßnahmen zur Verhinderung einer Infektionsausbreitung richten sich aber nicht gegen das Krankheitsgeschehen – dies können sie auf keinen Fall beeinflussen – sondern gegen das Infektionsgeschehen! Aufgrund der Inkubationszeit von etwa 7 Tagen findet sich das einem bestimmten R- Wert zugehörige Infektionsgeschehen weitere 7 Tage früher. So beschreibt der R- Wert vom 23.3. das Infektionsgeschehen 7 Tage früher am 16.3.2020 (genau genommen müssen aus den oben angeführten berechnungstechnischen Gründen weitere 4 Tage abgezogen werden: 12.3.2020). Für die Grafik bedeutet das eine Verschiebung der R- Kurve um 7 – 11 Tage nach links!

RKI- Grafik des R- Verlaufs vom 15.4.2020 im Original

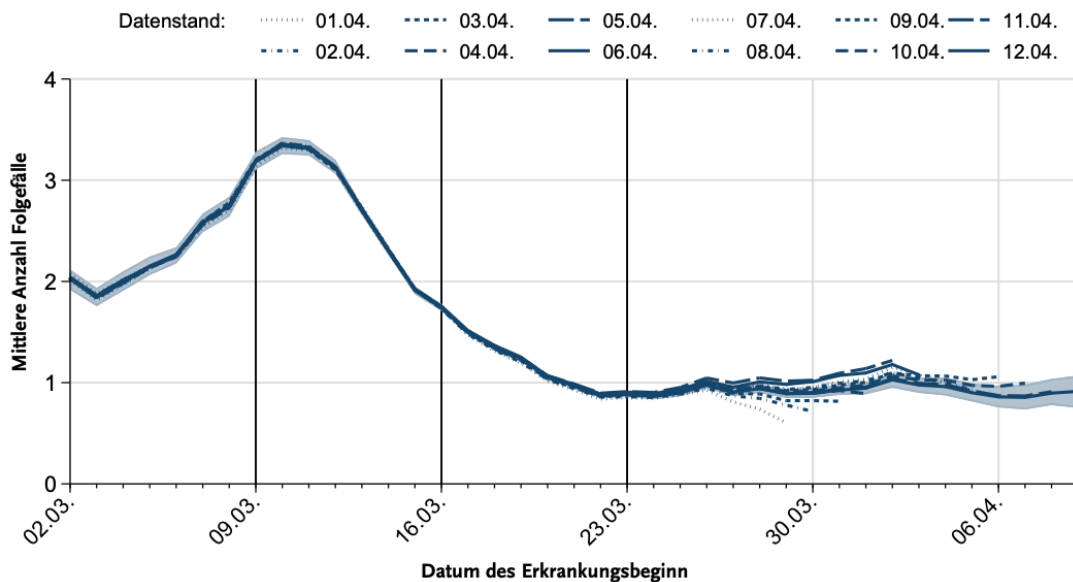
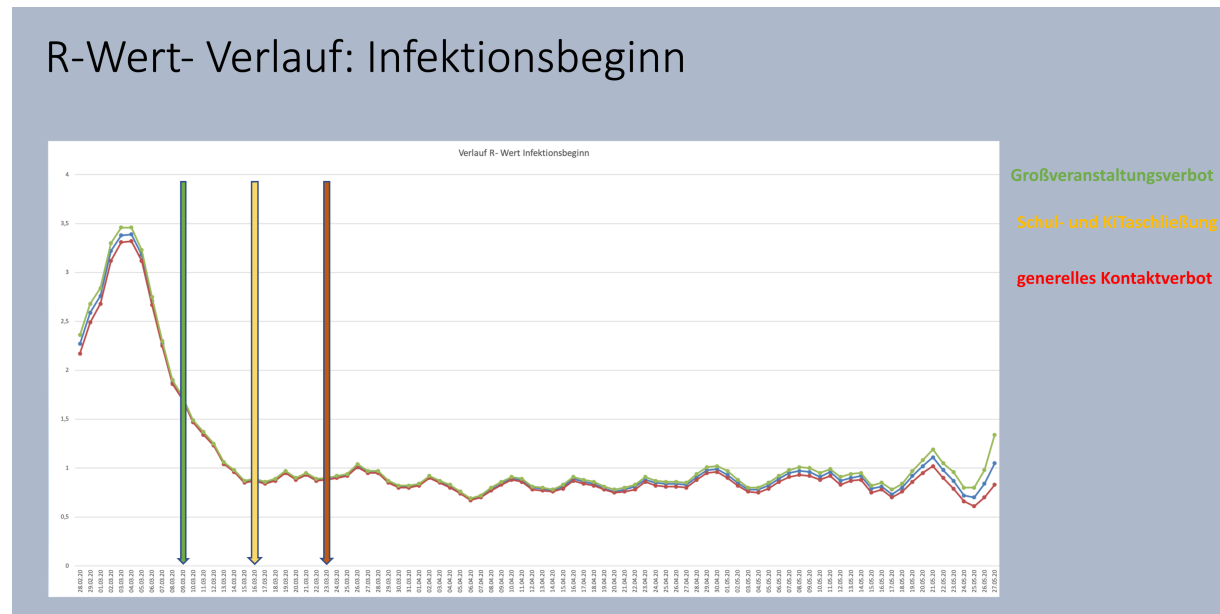
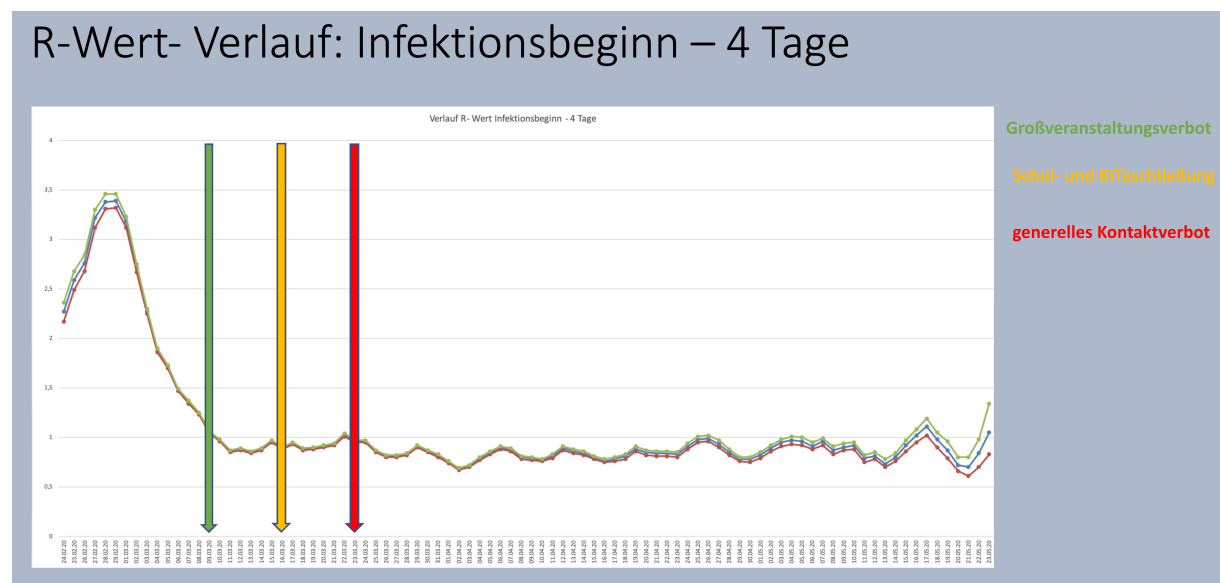


Abb. 5 | Vergleich der Schätzung der effektiven Reproduktionszahl R für eine angenommene Generationszeit von 4 Tagen zu unterschiedlichen Datenständen. Die schwarz hervorgehobenen vertikalen Linien kennzeichnen den Start der in Tab. 1 (S. 15) genannten Maßnahmen am 9. März, 16. März und 23. März 2020.

R-Wert-Verlauf basierend auf den öffentlich zugänglichen R-Werttabellen des RKI korrigiert um die Inkubationszeit von 7 Tagen (Datum des Infektionsbeginnes)

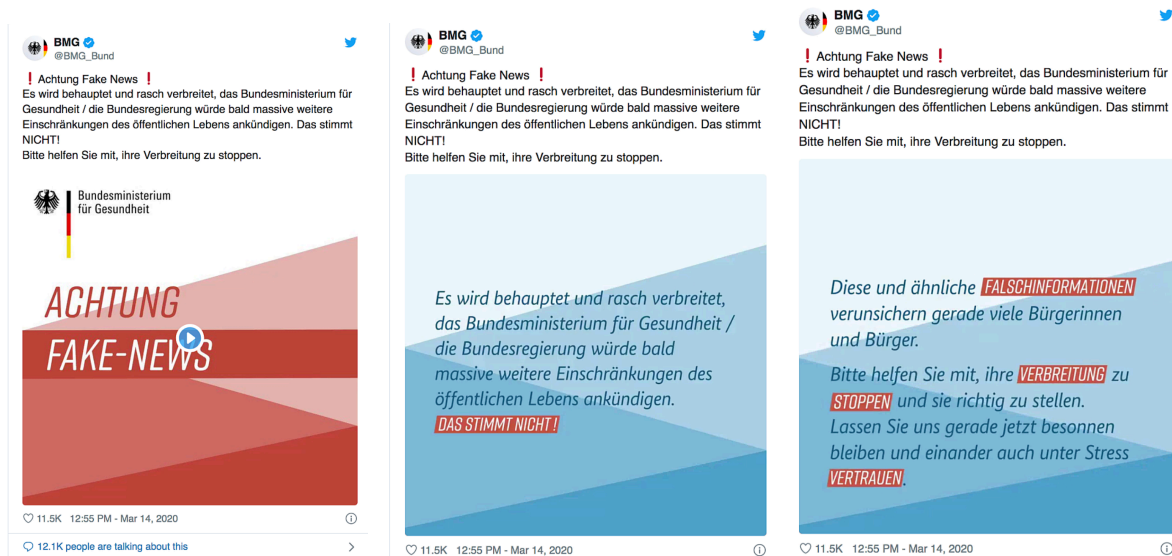


R-Wert-Verlauf basierend auf den öffentlich zugänglichen R-Werttabellen des RKI korrigiert um die Inkubationszeit von 11 Tagen (Datum des Infektionsbeginnes – 4 Tage)



Es ist auf den ersten Blick ersichtlich, dass es bereits am 09.03. zu einem drastischen Rückgang des Infektionsgeschehens gekommen war – am 16.03. und 23.03. war die Pandemie bereits beherrscht – der R- Wert lag dauerhaft unter 1. Einmal abgesehen von der Tatsache, dass alle Neuerkrankungszahlen, die in die R- Wertberechnung eingehen wie oben ausführlich dargelegt NICHT den wahren Pandemieverlauf widerspiegeln sondern allenfalls die RKI- Stichprobe beschreiben, kann nicht einmal für die RKI- Stichprobe belegt werden, dass die massiven Grundrechtseingriffe den Pandemieverlauf in irgendeiner Weise beeinflusst hätten. Das gerne geäußerte Argument, in Wahrheit hätten Betriebe und Vereine bereits in vorausseilendem Gehorsam vor dem 23.3.2020 die dann erst beschlossenen Kontaktbeschränkungen vorweggenommen greift

nicht – sie hätten dies dann mindestens 2 Wochen zuvor tun müssen! Um dies ausdrücklich zu verhindern verbreitet das Bundesgesundheitsministerium mit Datum vom 14.3.2020 noch ein Kurzvideo, in dem vor **FAKE-NEWS** gewarnt wird, wonach behauptet würde die Bundesregierung plane weitere massive Einschränkungen des öffentlichen Lebens – **DIES STIMME NICHT!**



Die obige Stellungnahme bezieht sich allein auf die handwerklichen statistischen Fehler des RKI und bestreitet keinesfalls das Vorhandensein einer durch SARS-CoV-2 verursachten Infektionskrankheit. Diese verläuft in der weit überwiegenden Zahl der Fälle sehr milde und führt in einer niedrigen Anzahl von Fällen ganz überwiegend bei sehr altern und mehrfach vorerkrankten Menschen zum Tod. Die Sterberate liegt hierbei wie inzwischen weltweit beschrieben zwischen etwa 0,05% und 0,4% - im Gegensatz zur jährlich wiederkehrenden Influenza, die bei ähnlicher Sterberate (trotz vorhandener Impfung!) auch regelmäßig Kleinkinder tötet gilt dies für die SARS-CoV-2- Infektion nicht!

Diese Infektion hatte zu keinem Zeitpunkt das Potential zu einer bedrohlichen Pandemie zu deren Schutz eine Grundrechtseinschränkung gerechtfertigt war!

Literaturhinweise:

The Scenario of a Pandemic Spread of the Coronavirus SARS-CoV-2 is Based on a Statistical Fallacy –
 Autor: Christof Kuhbandner, Universität Regensburg, Institut für Humanwissenschaften

Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions
 Priesemann et al, Science 15 May 2020

Instand Gesellschaft zur Förderung der Qualitätssicherung in medizinischen Laboratorien e.V. –
 Kommentar zum Extra- Ringversuch 340, Veröffentlichung vom 02.05.2020

Epidemiologisches Bulletin des Robert Koch Institutes 17/ 2020 vom 23.4.2020